INTERNATIONAL JOURNAL OF ADVANCES IN PHARMACY, BIOLOGY AND CHEMISTRY

Research Article

Molecular Modeling Studies of RNA Polymerase II Inhibitors as Potential Anticancer Agents

Ankita Agarwal*, Sarvesh Paliwal, Ruchi Mishra and Prerana

Department of Pharmacy, Banasthali University, Tonk-304022, Rajasthan, India.

ABSTRACT

Global physicochemical descriptor based QSAR models were developed using multiple linear regression (MLR) and partial least squares (PLS) for a set of 44 molecules as Derivatives of Oncrasin-1synthesized against cancer. Leave out one row method is used to validate the developed model. The MLR and PLS generated excellent models with good predictive ability and all the statistical values, such as r, r^2, r^2_{cv}, r^2 (test set), F and S values were 0.88, 0.78, 0.77, 0.77, 30.75 and 0.40 for MLR and r^2_{cv}, r^2 (test set) and statistical significance value were 0.77, 0.77 and 0.92 for PLS respectively, were satisfactory. The results obtained from this study provides insights regarding role of Bond dipole moment (subst. 2), Balaban topological index (whole molecule), First Atom E- state index (subst. 1) and VAMP heat of formation (Whole molecule) in determining the RNA polymerase II inhibitory activity. The results clearly reveal that the anti-cancer activity.

Keywords: QSAR, TSAR, MLR, PLS, Derivatives of Oncrasin-1, RNA polymerase II inhibitor.

INTRODUCTION

To optimize antitumor activity of oncracin-1, a small molecule RNA polymerase II inhibitor, oncracin-1 analogue for their cytotoxic activity against normal human epithelial cells and k-ras mutant tumor cells. Oncrasin-1 inhibits phosphorylation of the CTD of RNA polymerase II; oncrasin-1 contains an indole structure similar to indole-3-carbinol and a benzyl ring attached to the N atom of the indole. Indole-3 carbinol is a natural constituent of cruciferous vegetables that has been tested for cancer prevention and therapy.[1]

Oncrasin-1, that was identified through synthetic lethality screening on isogenic human ovarian epithelial cells with or without oncogenic Ras expression.[2] Molecular characterization revealed that oncrasin-1 induced apoptosis in a subset of cancer cell lines associated with induction of coaggregation of protein kinase Cι and RNA splicing factors in the nucleus and suppression of phosphorylation of the C-terminal domain (CTD) of the largest subunit of eukaryotic RNA polymerase II.[3] CTD phosphorylation is known to be essential for efficient transcription elongation and recruitment of mRNA processing factors, including the capping enzyme and splicing factors required for efficient processing of RNA transcripts.[4,5] Inhibiting CTD phosphorylation will ultimately disrupt RNA polymerase II function and promote cell death.

We decided to model experimentally determined IC50 values from the computationally derived molecular descriptors and to analyze the influence of descriptors on the biological activity, as well as to offer guidance for the design of some new compounds. We preferred simple and less error prone global descriptor based multivariate approach, as it is more beneficial compared to conformational and structural alignment-based methods.

MATERIAL AND METHODS

The structures and anti-diabetic activities of 55 RNA polymerase II inhibitors [1] (Table 1) were sketched using Chem Draw software and were imported on TSAR (Version 3.3; Accelrys Inc, oxford, England) software. The generated 3D models of all derivatives created were cleaned up and subjected to charge calculation and energy minimization.

More than 300 molecular descriptors were calculated for all the compounds under consideration. TSAR affords the calculation of the following descriptors: atomic attributes (like molecular properties, dipole moment and verloop steric parameters), atomic

indices (like shape, connectivity and topological indices) and Vamp electrostatic properties (total energy, HOMO, LUMO, heat of formation, etc).[6] To reduce data redundancy, pairwise correlation analysis was carried out.[7] Among the highly intercorrelated descriptors the one that had high correlation with biological activity was kept and other was discarded. This process was repeated number of times and finally five descriptors were retrieved that were highly correlated with biological activity and were not having intercorrelation among each other.

To develop QSAR models, stepwise MLR analysis with leave-one-out (LOO)[8] cross-validation was applied to the training set. The molecules of the series were divided randomly into training set (41 molecules) and test set (14 molecules) with descriptors retrived by data reduction. Training set was used to build linear models so that an accurate relationship could be found between structure and biological activity.[9] The test set of eight molecules was not used to develop the regression model but served to check the predictive power of the developed model. In addition to MLR, partial least squares (PLS)[10] analysis was also performed to check the predictive ability and robustness of the developed model.

RESULT AND DISCUSSION

Multiple linear regression (MLR) and partial least squares (PLS) were used to derive the QSAR equations. The statistically significant model was constructed from the training set by using 4 parameters.

In order to improve the predictivity of the model, 7 potential outliers namely 26, 39, 45, 53, 54, 58 and 59 (which exhibited high residual value and were two far away from regression line) were identified and deleted.

The final regression equation obtained from MLR analysis (final 34 molecules in training set) after deleting the outliers is represented as equation 1

$$Y = 0.179 \times X1 - 4.649 \times X2 - 0.406 \times X3 + 0.005 \times X4 + 8.048 \quad \text{–Equation 1}$$

Where, X1 is Bond dipole moment (subst. 2), and X2 is Balaban topological index (whole molecule), X3 is First Atom E- state index (subst. 1) X4 is VAMP heat of formation (Whole molecule).

This best model was selected on the basis of various statistical parameters such as coefficient of determination (r^2), predictive power of model (r^2_{cv})standard deviation (SD), sequential Fisher test (F) and test for statistical significance (t). The value of r^2 should always be greater than 0.6 (a good model should have an $r^2 > 0.9$) and the value of r^2_{cv} could fall into three categories [11]:

- $r^2_{cv} > 0.6$: The model is fairly good.
- $0.4 < r^2_{cv} < 0.6$: The model is questionable.
- $r^2_{cv} < 0.4$: The model is poor.

r = 0.88, r^2 = 0.78, r^2_{cv} = 0.77, F = 30.7, S = 0.40, predictive r^2 for test set = 0.77

PLS analysis was also performed on the same data set to check the soundness of the MLR model. The resulted r^2_{cv} value of 0.77 clearly demonstrates the high predictive ability of the developed PLS model (equation 2).

$$Y = 0.179 \times X1 - 4.649 \times X2 - 0.406 \times X3 + 0.005 \times X4 + 8.048 \quad \text{–Equation 2}$$

Where, X1 is Bond dipole moment (subst. 2), and X2 is Balaban topological index (whole molecule), X3 is First Atom E- state index (subst. 1) X4 is VAMP heat of formation (Whole molecule).

n = 14, Statistical significance = 0.92, r^2_{cv} = 0.77, Predictive r^2 for test set = 0.77

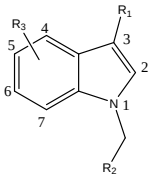
Since for a well defined problem, both MLR and PLS should generate comparable results [12], the r^2_{cv} values of MLR and the PLS models were evaluated and it was found that both the models have comparable r^2_{cv} value of 0.77 and 0.77 for MLR and PLS respectively. The predictive ability of the model was also validated using the external test set of 8 compounds in context of minimum difference between the actual and predicted biological activity values of MLR and PLS analysis for training and test which is shown in table 2, 3 and their respective plots are depicted in figure 1, 2, 3, 4 and 5.

Analysis of Descriptors

Bond dipole moment (subst.2) is an electronic descriptor. It is positively correlated with the biological activity. So with increase in Bond dipole moment of molecule there will be increase in biological activity. Balaban topological index (whole molecule) is negatively correlated with the biological activity. So with decrease in Balaban topological index of molecule there will be increase in biological activity. First atom E-state index (subst. 1) is negatively correlated with the biological activity. So with decrease in First atom E-state index of molecule there will be increase in biological activity. VAMP

heat of formation (whole molecule) is positively correlated with the biological activity. So with increase in VAMP heat of formation of molecule there will be increase in biological activity.

Table 1: Series of Oncrasin-1 derivatives along with their biological activities



Comp. Name	R_1	R_2	R_3	IC_{50} values
1	CHO	2'-F Ph	-	0.1
2	CHO	3'-F Ph	-	0.031
3	CHO	3'-Cl Ph	-	0.037
4	CHO	3'-Br Ph	-	0.039
5	CHO	4'-Br Ph	-	0.031
6	CHO	3'-CF_3 Ph	-	0.14
7	CHO	4'-CF_3 Ph	-	1.58
8	CHO	3'-Me Ph	-	0.039
9	CHO	4'-Me Ph	-	0.063
10	CHO	3'-NO_2 Ph	-	0.039
11	CHO	4'-NO_2 Ph	-	0.50
12	CHO	2',4'-2Cl Ph	-	0.1
13	CHO	3',4'-2Br Ph	-	0.063
14	CHO	Ph	-	0.045
15	CHO	4'-Cl$(CH_2)_2$OPh	-	0.045
16	CHO	3'-Cl Ph	5-F	0.37
17	CHO	4'-Cl Ph	5-F	0.29
18	CHO	4'-Cl Ph	6-F	0.12
19	CHO	4'-Cl Ph	5-Cl	0.4
20	CHO	4'-Cl Ph	6-Br	0.09
22	CHO	$CH(CH_3)_2$	-	1.99
26	$COCH_3$	Ph	-	10
30	CH_2OH	Ph	-	0.112
31	CH_2OH	3'-F Ph	-	0.04

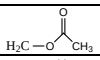
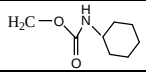
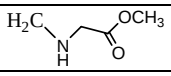
32	CH_2OH	4'-F Ph	-	0.118
33	CH_2OH	2'-Cl Ph	-	0.10
34	CH_2OH	3'-Cl Ph	-	0.019
35	CH_2OH	4'-Cl Ph	-	0.016
36	CH_2OH	3'-Br Ph	-	0.028
37	CH_2OH	4'-Br Ph	-	0.031
38	CH_2OH	3'-I Ph	-	0.1
39	CH_2OH	3'-CF_3 Ph	-	0.039
40	CH_2OH	4'-CF_3 Ph	-	1.25
41	CH_2OH	3'-Me Ph	-	0.031
42	CH_2OH	4'-Me Ph	-	0.031
43	CH_2OH	3'-NO_2 Ph	-	0.039
44	CH_2OH	4'-NO_2 Ph	-	0.079
45	CH_2OH	4'-OMe Ph	-	0.079
46	CH_2OH	2',4'-2Cl Ph	-	0.079
47	CH_2OH	3',4'-2Cl Ph	-	0.031
48	CH_2OH	3',4'-2Br Ph	-	0.05
49	CH_2OH	H	5-F	16.9
50	CH_2OH	3'-F Ph	5-Cl	0.07
51	CH_2OH	4'-Cl Ph	5-F	0.03
52	CH_2OH	4'-Cl Ph	6-F	0.45
53	CH_2OH	4'-Cl Ph	5-Cl	5.75
54	CH_2OH	4'-Cl Ph	5-OMe	1.51
55	CH_2OH	4'-Me Ph	5-F	0.01
58	$C(OH)HCH_3$	4'-Cl Ph	-	25.1
59	CH_2OCH_3	4'-Cl Ph	-	1.99
60		4'-Cl Ph	-	0.039
61		4'-Cl Ph	-	0.02
64	$COOCH_3$	4'-Cl Ph	-	31.6
65	$COOCH_3$	4'-Br Ph	-	5.00
67		4'-Cl Ph	-	0.01

Table 2: Actual and predicted activity of Training set obtained by MLR and PLS analysis

Comp. Name	Actual value	Predicted value	
		MLR	PLS
1	1	1.025	1.025
4	1.408	1.177	1.177
5	1.508	1.068	1.068
6	0.853	0.340	0.340
7	-0.198	0.225	0.225
10	1.408	1.013	1.013
11	0.301	0.796	0.796
12	1	0.886	0.886
14	1.346	1.267	1.267
16	0.431	0.861	0.861
19	0.397	0.890	0.890
20	1.045	0.923	0.923
22	-0.298	-0.747	-0.747
30	0.950	1.513	1.513
31	1.397	1.177	1.177
32	0.928	1.089	1.089
33	1	1.447	1.447
36	1.552	1.427	1.427
37	1.508	1.358	1.358
38	1	1.519	1.519
41	1.508	1.490	1.490
42	1.508	1.554	1.554
43	1.408	1.293	1.293
44	1.102	1.044	1.044
48	1.301	1.231	1.231
49	-1.228	-0.403	-0.403
50	1.154	1.084	1.084
51	1.522	1.012	1.012
55	2	1.328	1.328
60	1.408	0.904	0.904
61	1.698	2.096	2.096
64	-1.499	-1.123	-1.123
65	-0.698	-1.002	-1.002
67	1	0.955	0.955

Table 3: Actual and predicted activity of Test set obtained by MLR and PLS analysis

Comp. Name	Actual value	Predicted value	
		MLR	PLS
2	1.508	1.630	1.630
3	1.431	1.087	1.087
8	1.408	1.197	1.197
9	1.200	1.262	1.262
13	1.200	1.039	1.039
15	1.346	1.149	1.149
17	0.537	0.761	0.761
18	0.920	0.911	0.911
34	1.721	1.811	1.811
35	1.795	1.638	1.638
40	-0.096	0.503	0.503
46	1.102	1.137	1.137
47	1.508	1.423	1.423
52	0.346	0.960	0.960

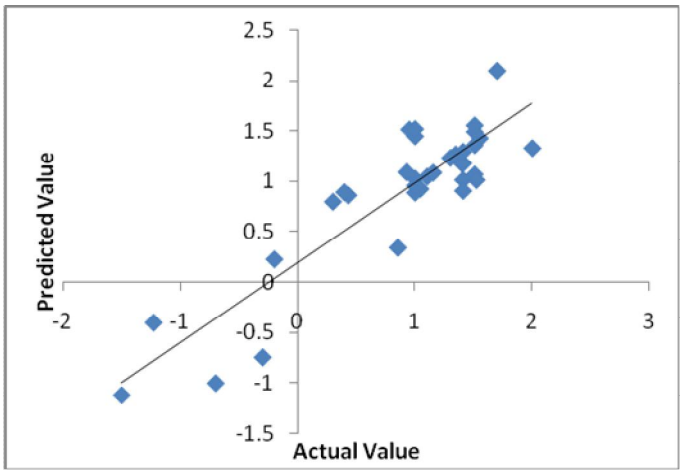


Fig. 1: Actual vs. predicted activity for the training set of compounds derived from MLR analysis

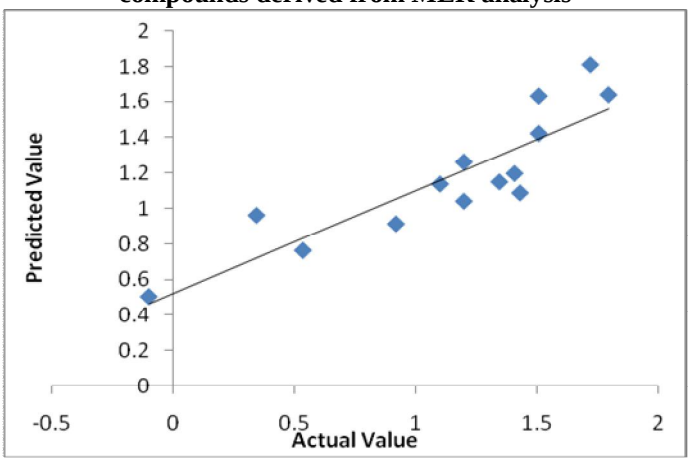


Fig. 2: Actual vs. predicted activity for the test set of compounds derived from MLR analysis

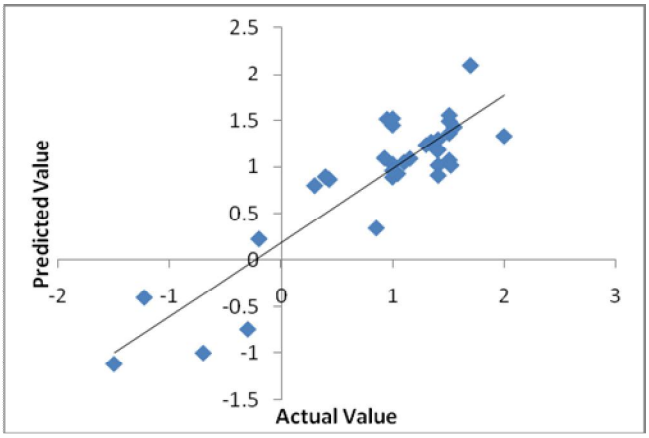


Fig. 3: Actual vs. predicted activity for the training set of compounds derived from PLS analysis

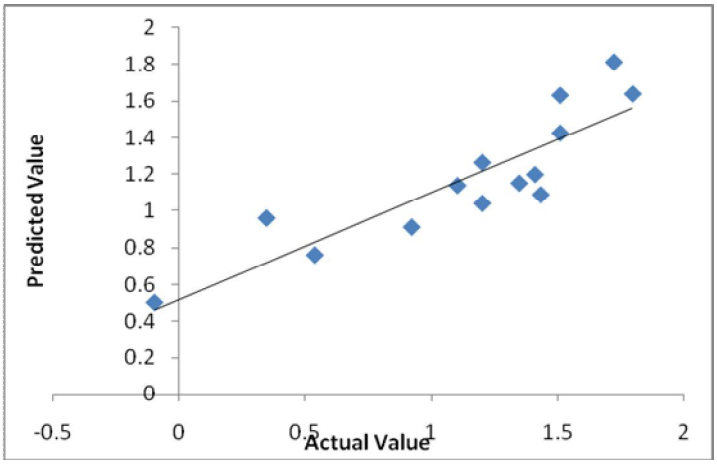


Fig. 4: Actual vs. predicted activity for the test set of compounds derived from PLS analysis

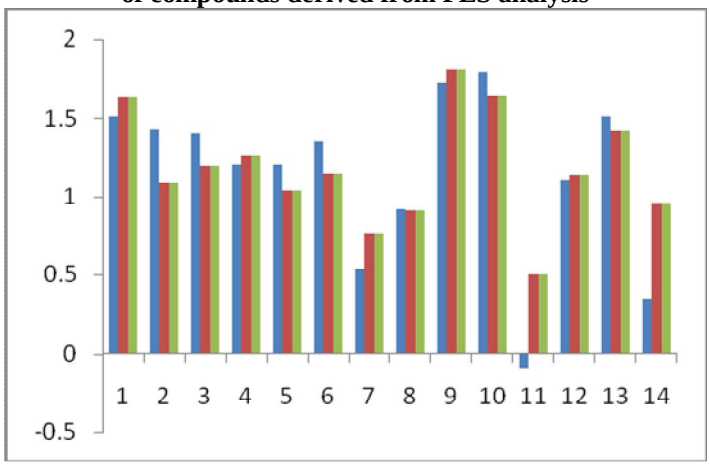


Fig. 5: Graph showing comparison between actual (blue) and predicted activity for the test set of compounds derived from MLR (red) and PLS (green) analysis

CONCLUSION

On the basis of present study, it can be concluded that the physicochemical descriptors have sufficient reliability to relate the biological activity of Derivatives of Oncrasin-1synthesized with their structural features. A highly predictive QSAR model has been obtained using the MLR. It was validated using external test set of 14 compounds and PLS to all molecules. Based on the thorough analysis of the experimental data of the target property, this computational approach will help to add up new molecules as anti diabetic agents and to combat with existing problems of this class. The findings of present study will certainly aid in the design of more potent anti cancer agents with improved activity and reduced mechanism based side effects of traditional anti cancer agents.

ACKNOWLEDGEMENT

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

REFERENCES

1. Shuhong Wu, Li Wang, Wei Guo, Xiaoying Liu and Jinsong Liu. Analogues and Derivatives of Oncrasin-1, a Novel Inhibitor of the C-Terminal Domain of RNA polymerase II and Their Antitumor Activities. Journal of medicinal chemistry. 2011;28(54):2668-79.
2. Guo W, Wu S, Liu J and Fang B. Identification of a small molecule with synthetic lethality for K-ras and protein kinase C iota. Cancer Res. 2008;68:7403–7408.

3. Guo W, Wu S, Wang L, Wang R and Wei L. Interruption of RNA processing machinery by a small compound1-[(4-chlorophenyl) methyl]-1H-indole-3-carboxaldehyde (oncrasin-1). Mol Cancer Ther. 2009;8:441–448.
4. Archambault J, Chambers RS, Kobor MS, Cartier M and Bolotin D. An essential component of a C-terminal domain phosphatase that interacts with transcription factor IIF in Saccharomyces cerevisiae. Proc Nat Acad Sci U.S.A. 1997;94(26):14300–14305.
5. Misteli T and Spector DL. RNA polymerase II targets premRNA splicing factors to transcription sites in vivo. Mol Cell. 1999;3:697–705.
6. Agarwal A, Mishra R and Paliwal S. A QSAR Study Investigating the Potential Anti-Leishmanial Activity of Cationic 2-Phenylbenzofurans. In Chemistry of Phytopotentials: Health, Energy and Environmental Perspectives. Edited by Khemani LD, Srivastava MM, Srivastava S. Springer-Verlag Berlin Heidelberg, New York. 2012:137-141.
7. Mishra R, Agarwal A and S Paliwal. QSAR Analysis of Anti-Toxoplasma Agents. In Chemistry of Phytopotentials: Health, Energy and Environmental Perspectives. Edited by Khemani LD, Srivastava MM, Srivastava S. Springer-Verlag Berlin Heidelberg, New York. 2012:131-135.
8. JinCan C, Li Q, Yong S, LanMei C and KangCheng Z. A QSAR study and molecular design of benzothiazole derivatives as potent anticancer agents. Sci China Ser B-Chem. 2008;51:111-119.
9. Golbraikh A, Shen M, Xiao Z, Xiao YD and Tropsha A. Rational selection of training and test sets for the development of validated QSAR models. Journal of Computer-Aided Molecular Design. 2003;17:241–253.
10. Yu Y, Su R, Wang L, Qi W and He Z. Comparative QSAR modeling of antitumor activity of ARC-111 analogues using stepwise MLR, PLS, and ANN techniques. Medicinal Chemistry Research. 2009;DOI 10.1007/s00044-009-9266-9.
11. Prasad RY, Kumar RP, Smiles JD and Babu AP. QSAR studies on chalcone derivatives as antibacterial agents against Bacillus pumilis. Arkivoc. 2008;11:266-276.
12. Cramer RD, Partial Least Squares (PLS): its strengths and limitations. Perspect Drug Discov Des. 1993;1:269–278.