QSAR Studies on Thiosemicarbazone Derivatives as an Anticancer Agent

Niharika Gokhale[1*] and Sanjay Jain[2]

[1], Department of Pharmaceutical Chemistry, Suresh Gyan Vihar University, Jaipur, RJ, India-302004

[2,] Smriti College of Pharmaceutical Education, Indore, MP, India-452010

Research Article

Please cite this paper as Niharika Gokhale[1*] and Sanjay Jain[2]. QSAR Studies on Thiosemicarbazone Derivatives as an Anticancer Agent. IJPTP, 2015, 6(2), 2111-219.

Corresponding Author:

Niharika Gokhale

Department of Pharmaceutical Chemistry,
Suresh Gyan Vihar University,
Jaipur, RJ, India-302004
Phone number: +919826980809

Abstract

QSAR studies of a set of thiosemicarbazone derivatives have been performed in present investigation, showing inhibitory activity against ribonucleotide reductase, with multiple linear regressions (MLR) analysis and Support vector machine (SVM). Four relevant descriptors calculated are identified as LOGP values. Their relationship with biological activity IC_{50} have provided structural insights in interpretation of results. Internal validations of QSAR models have been achieved using R^2_{CV} (LOO), PRESS, SDEP and Y-Scrambling.

Keywords: *QSAR, Thiosemicarbazone, inhibitory activity, support vector machine, biological activity*

Introduction

Cancer is one of the most important leading causes of death worldwide. Cancer is the disease of remarkable significance today. After cardiovascular disease it is a second leading cause of death and it will turn out to be primary cause of death in coming years.[1] Each year, the American Cancer Society estimates the numbers of new cancer cases and deaths expected in the United States in the current year and compiles the most recent data on cancer incidence, mortality, and survival based on incidence data from the National Cancer Institute and the Centers for Disease Control and Prevention, and mortality data from the National Center for Health Statistics. This year's report estimates there will be 1,665,540 new cancer cases and 585,720 cancer deaths in the United States in 2014.[2] Cancer is basically a disorder at cellular level. When the proteins implicated in the cell division is not able to coerce the sequence from one cell cycle stage to next, causes disruption of normal cell division.[3] Cells escape from normal control of replication and migration and multiply uncontrollably and invade nearest cells.[2,3] Thiosemicarbazones belong to a class of compounds that occupy a wide range of biological activities.[4] Thiosemicarbazones are the biologically important scaffold is known to possess anticancer activity.[5] SAR studies showed that a large number of Thiosemicarbazones with heterocyclic compounds and N N S ligand system proved to be an important feature for carcinogenic potency.[6] Also it is having strong correlation between enzyme RNR and rate of tumor growth.[6,7] Thiosemicarbazones exhibits its potential by inhibiting ribonucleotide reductase. Ribonucleotide reductase catalyzes the reduction of ribonucleotides . It acts as a machine for the synthesis of deoxyribonucleotides. [8]

QSAR is a tool that includes statistics to identify and predict the biological response of compounds as a function of their structure.[9] Multiple linear regression analysis is the most simple approach helps to identify the linear relationship among molecular structures and biological activities. For Nonlinear relationship an accurate, robust and fast statistical tool is SVM.[9-11] Efficiency of SVM in identifying non linear relationship is appreciable.[12] The computational tool which has been extensively used in the present investigation is Sarchitect Designer 2.5.0 (Strand Life Science).[12, 13] Sarchitect, a sophisticated tool with simple user interface, provides latest machine learning techniques like MLR, SVM, ANN, PLS *etc.* for regression and

classification. Its heuristic accuracy, ro-bustness and simple interface makes it a tool of choice to achieve QSAR models.[13-14]

Material and Method

Dataset

Data set has always been significant factor for the finest results and their interpretation. In this QSAR thiosemicarbazones as an anticancer agents has been adopted. There by a total of twenty thiosemicarbazone derivatives along with their biological activity are taken. 2D Structure of the molecules were generated and optimized in Marvin Sketch 5.1.5 developed by ChemAxon freeware software.[15] The structure and IC_{50} values of anticancer activity on MCF-7 of these analoges are give in (**Table 1**).

Descriptor Calculation

Molecular descriptors are numerical values to evaluate and establish the structural activity relationship. Structures are drawn in chemsketch which then converted into their SMILES (Simplified molecular line entry specification).[16] Structure representations were used to calculate descriptors using E-dragon software. 1647 descriptors belonging to various classes were calculated.

Model building

The data set consist of twenty thiosemicarbazone derivatives are considered as a training set for the present study. Once descriptors were generated, descriptor-screening methods were used to select the most relevant descriptors to establish the models.1647 descriptors were imported into Sarchitect designer 2.5.0. (Strand Life Science). Pruning of descriptors were performed by considering the parameters (standard deviation less than equal to 0, and missing values greater than equal to 1) to remove constant and missing set of descriptors. Remaining set of molecular descriptors were further evaluated for Pearson correlation coefficient calculation followed by redundancy check removal of highly correlated descriptors. Forward selection wrapper algorithm used to obtain the best statistically significant combination of descriptors with descriptor target size of four.

SVM and MLR Model Building

QSAR models provides valuable tool to predict the activity through statistical modeling. SVM is a machine learning method developed by Vapnik and coworkers.[17] SVM is the machine learning method.[18] Which is the powerful technique for classification and regression. Recent researches have approved the efficient role of SVM in QSAR studies as an accurate, robust and fast statistical tool in identifying on linear relationship. Literature avails sufficient theoretical aspects of SVM.[19]

MLR QSAR MODELS

We carried out through a comparative SVM an MLR towards QSAR of anticancer agents. Descriptors have been chosen with the target size four. MLR results have been discussed using trivariable and tetravariable models (eq.1 and 2) with appreciable sets of parameters showing goodness of predictions**.** MLR models and their statistical fitness have been shown in (**Table 2**).

Model-3 Trivariable

pIC50 = 6.4063215 + 0.38642403 (MACCS151)-35.30811 (X5A_v1.0) + 0.045728557(mlogpPRX_v1.0) Eq. (1)

N=20 R^2= 0.9144288 R^2A= 0.8983842 S.E. = 0.39827764 F= 56.99296

Model-4 Tetravariable

pIC50 = 6.367 + 0.3532(MACCS151) - 38.8177(X5A) + 0.0955(GATS5e) +0.0541(mlogpPRX) Eq. (2)

N=20 R^2=0.9608775 R^2A=0.9504449 S.E. = 0.27828595 F= 92.10292

N is the number of compounds used, S.E. is standard error, R^2 is multiple regression coefficient, R^2A is adjusted R^2 and F is Fisher's statistics.

Stepwise forward selection reveals gradual increase in R^2 values with subsequent increase in R^2A and standard error minimization.

SVM QSAR MODELS

Dataset was processed in SVM using Linear, Polynomial and Gaussian kernel functions Gaussian kernel function yielded appreciable results in terms of model fitness

and predictive powers of training set. Gaussian kernel functions proved that thiosemicarbazones posses nonlinear relationship with biological activity $IC_{50.}$ SVM models and their statistical fitness have been provided in (Table 3).

MLR AND SVM MODEL VALIDATION

Correlation of observed and predicted activity, R^2CV (Cross Validated R^2), Y-scrambling or Y-randomization, Q^2 and PRESS have been used as internal validation parameters. R^2CV for MLR and SVM models show successive increase in magnitudes on moving down to the models to propose stability (Table **3**). Tetravariable models of MLR and SVM produced R^2CV respectively 0.92863 and 0.89234. Correlation among observed and predicted IC_{50} values have been presented in (**Table 4) and fig (1 & 2).**Y scrambling tetra variable MLR and SVM model is presented as a validation event in **Fig (3 & 4)** to avoid by chance correlation and modeling. R^2CV, PRESS and Q^2 values found appreciable (**Table 5**). R^2CV for MLR and SVM show successive increase in magnitudes on moving down to multivariate model to propose stability.

Results and Findings

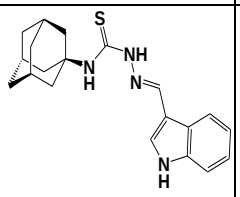
INTERPRETATION OF STATISTICAL MODELS

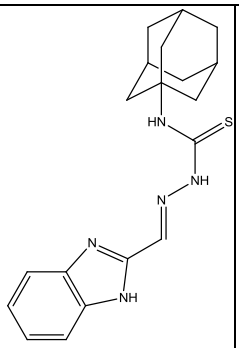
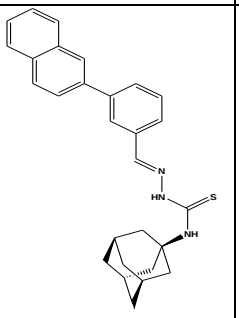
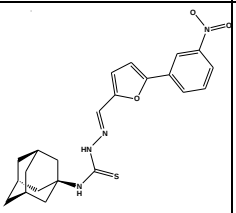
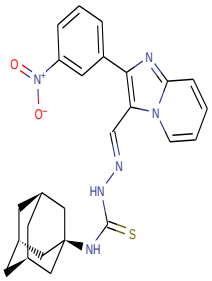
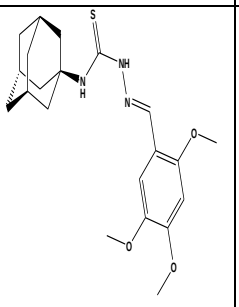
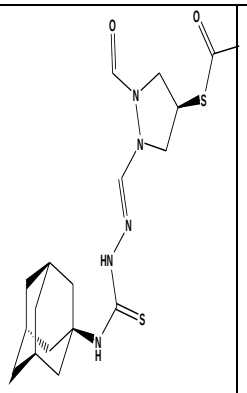
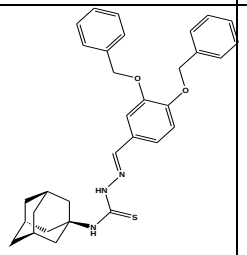
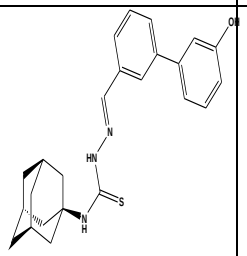
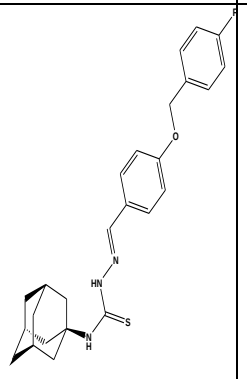
In the present investigation we produce some precise discussion in terms of molecular structure and IC_{50} relationship. Chosen descriptors have been correlated with biological activity.

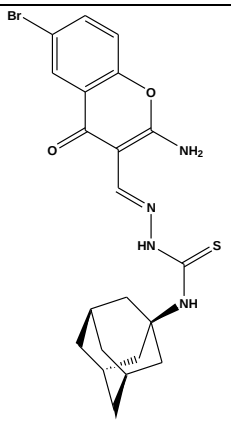
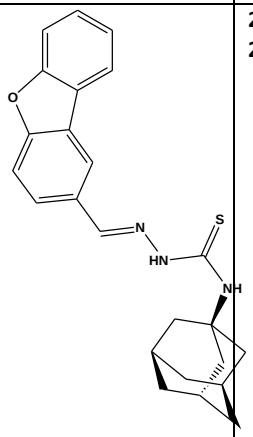
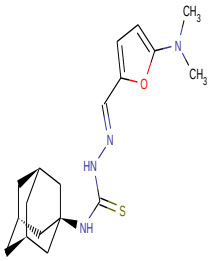
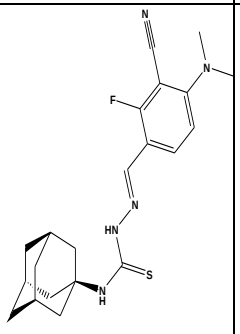
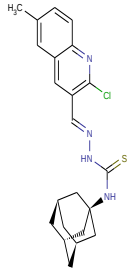
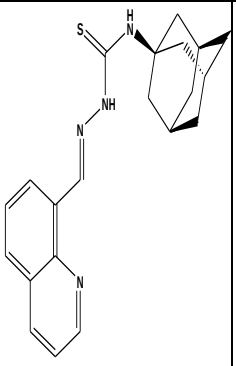
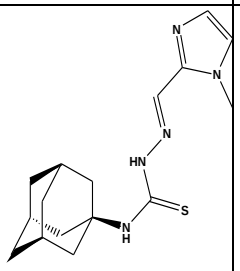
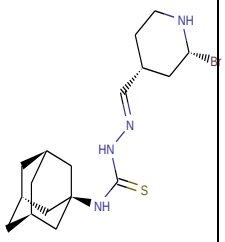
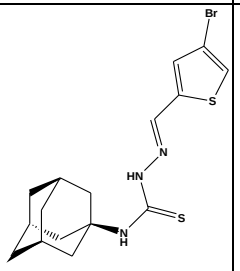
MACCS151 (Molecular ACCess system) is the descriptor that describe the molecule on the basis of an algorithm like 2D (two dimensional) fingerprint. Which is employed for similarity searching in databases and wide range of activities including prediction of adsorption, distribution, metabolism, excretion and toxicity properties.[20] The positive value shows the increment in activity. X_5A average connectivity index chi-5.Molecular connectivity indices is denoted by chi. It is a Kier-Hall connectivity indices, which is calculated based on the graph representation of the molecule (hydrogen-depleted molecular graph). A path of length represents the connection of two atoms with the bond. It relates to the valence electrons and number of atoms in the molecule. It has a negative effect on the model. Its negative coefficient shows that decrease in heteroatom and fifth order valence connectivity increases the activity.

GATS5e corresponds to Geary autocorrelation lag 5, one of the 2D autocorrelations, and is weighted by atomic Sanderson electronegativities. It reveals that electronegativity of the substituent strongly affect the biological response. GATs5e represent the importance of electronegativity. The positive value of GATS5e (weighted by atomic Sanderson electronegativities) will be in favour of activity.

[MlogpPRX] Moriguchi based lipophilicity descriptor (proximity effect). Proximity effect of N & O. PRX, a proximity correction factor for nitrogen and oxygen atoms that are close to one another. Lipophilicity is an important aspect of transport and distribution of drugs in biological systems. Descriptors that relate to the biological properties of a compound such as toxicity descriptors which include an important solubility index called Log P, a partition coefficient between octanol and water. An algorithm to calculate Log P was introduced using a regression analysis of biological activities to establish a QSAR. The more hydrophobic a compound, the easier is its uptake into (e.g. human) fat. The method uses an equation with 13 parameters determined through regression analysis. The parameters are based on counting atoms, bonds, and fragments or functional groups. In the discovery setting 'the rule of 5' predicts that poor absorption or permeation is more likely when there are more than 5 H-bond donors, 10 H-bond acceptors, the molecular weight (MWT) is greater than 500 and the calculated Log P (C LogP) is greater than 5 (or MlogP > 4.15).

s. no	Compound code	compound name	Structure	IC_{50} (µM /ml)
1.	TSC-1	1-(adamantan-1-yl)-3-[(E)-(1H-indol-3-ylmethylidene)amino]thiourea		1.463

2	TSC-2	1-(adamantan-1-yl)-3-[(E)-(1H-1,3-benzodiazol-2-ylmethylidene)amino]thiourea		1.317
3	TSC-3	3-(adamantan-1-yl)-1-[(E)-{[3-(naphthalen-2-yl)phenyl]methylidene}amino]thiourea		3.499
4	TSC-4	1-yl)-3-[(E)-({5-[3-(hydroxynitroso)phenyl]furan-2-yl}methylidene)amino]thiourea		3.203
5	TSC-5	1-(adamantan-1-yl)-3-[(E)-{[2-(3-nitrophenyl)imidazo[1,2-a]pyridin-3-yl]methylidene}amino]thiourea		1.337
6	TSC-6	3-(adamantan-1-yl)-1-[(E)-[(2,4,5-trimethoxyphenyl)methylidene]amino]thiourea		4.532
7	TSC-7	3-[(E)-{[(4S)-4-(acetylsulfanyl)-2-formylpyrazolidin-1-yl]methylidene}amino]-1-(adamantan-1-yl)thiourea		2.354
8	TSC-8	3-(adamantan-1-yl)-1-[(E)-{[3,4-bis(benzyloxy)phenyl]methylidene}amino]thiourea		5.629
9	TSC-9	3-(adamantan-1-yl)-1-[(E)-{[3-(3-hydroxyphenyl)phenyl]methylidene}amino]thiourea		5.889
10	TSC-10	3-(adamantan-1-yl)-1-[(E)-({4-[(4-fluorophenyl)methoxy]phenyl}methylidene)amino]thiourea		6.673

11	TSC-11	3-(adamantan-1-yl)-1-[(E)-[(2-amino-6-bromo-4-oxo-4H-chromen-3-yl)methylidene]amino]thiourea		1.289
12	TSC-12	1-(adamantan-1-yl)-3-[(E)-{8-oxatricyclo[7.4.0.0{2,7}]trideca-1(13),2,4,6,9,11-hexaen-4-ylmethylidene}amino]thiourea		2.562
13	TSC-13	1-(adamantan-1-yl)-3-[(E)-{[5-(dimethylamino)furan-2-yl]methylidene}amino]thiourea		3.452
14	TSC-14	3-(adamantan-1-yl)-1-[(E)-{[3-cyano-4-(dimethylamino)-2-fluorophenyl]methylidene}amino]thiourea		4.452
15	TSC-15	3-(adamantan-1-yl)-1-[(E)-[(2-chloro-6-methylquinolin-3-yl)methylidene]amino]thiourea		2.889
16	TSC-16	3-(adamantan-1-yl)-1-[(E)-(quinolin-8-ylmethylidene)amino]thiourea		2.684
17	TSC-17	1-(adamantan-1-yl)-3-[(E)-[(1-methyl-1H-imidazol-2-yl)methylidene]amino]thiourea		3.826
18	TSC-18	3-(adamantan-1-yl)-1-[(E)-{[(2R,4R)-2-bromopiperidin-4-yl]methylidene}amino]thiourea		1.262
19	TSC-19	1-(adamantan-1-yl)-3-[(E)-[(4-bromothiophen-2-yl)methylidene]amino]thiourea		3.471

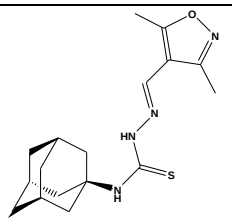
20	**TSC-20**	**1-(adamantan-1-yl)-3-[(E)-[(3,5-dimethyl-1,2-oxazol-4-yl)methylidene]amino]thiourea**		**4.137**
Standard drug	**Doxorubicin**			**3.710**

Table 2. Multiple linear regression analysis and their statistical parameters

Conclusion

Models of MLR and SVM presented significant statistics towards model fitness and predictability. SVM identifies nonlinear relationship among structures of thiosemicarbazones. It concludes that any structural changes would leads to nonlinear change in IC_{50}. SVM facilitated by Gaussian kernel function produced efficient QSAR models as compare to other kernels and MLR. R^2CV using N-Fold method, Y-scrambling validation further conferred QSAR model validation events. The descriptors identified in MLR analysis highlighted the role of atomic properties, increase in atomwise electronegativity distribution ensures the ligand – receptor interaction. Overall, the results suggested the robustness of the developed QSAR models in predicting the anti cancer activity. Such knowledge serves as a general guideline for future structural modifications of substituted

Model	No of Descriptors	R Square Train Metric	Max Absolute Error Train Metric	Mean Absolute Error Train Metric	R Square Validation Metric	Max Absolute Error Validation Metric	Mean Absolute Error Validation Metric
One-descriptor model	1	0.562305	0.4223752	0.10513754	0.51107216	0.44635534	0.11239352
Two-descriptor model	2	0.8106641	0.200562	0.0789042	0.7375884	0.2619686	0.09326775
Three-descriptor model	3	0.9144288	0.11810207	0.052671336	0.8757063	0.15166903	0.06567001
Four-descriptor model	4	0.9608775	0.09767771	0.03607998	0.9286385	0.13904715	0.049648356

Table 3. Support Vector Machine analysis and their statistical parameters

Model	No of Descriptors	R Square Train Metric	Max Absolute Error Train Metric	Mean Absolute Error Train Metric	R Square Validation Metric	Max Absolute Error Validation Metric	Mean Absolute Error Validation Metric
One-descriptor model	1	0.56156576	0.41577864	0.10430813	0.54169714	0.4275999	0.110353544
Two-descriptor model	2	0.83377403	0.20637369	0.06381283	0.80836326	0.2105031	0.07575829
Three-descriptor model	3	0.9320515	0.14187479	0.036532737	0.87795883	0.15791893	0.06757712
Four-descriptor model	4	0.9697373	0.1031518	0.021075297	0.89234996	0.18221664	0.056169488

thiosemicarbazones as therapeutic agents against cancer with potentially higher potency and less toxicity.

Table 4. Comparison of Observed and Predicted IC_{50} Values for tetra variable Model Using MLR and SVM

Molecule No.	Observed Activity IC50	MLR Tetra variable Model.		SVM Tetra variable Model	
		Predicted log IC50(nm)	Standard Error(pIC50)	Predicted log IC50(nm)	Standard Error(pIC50)
Molecule 1	5.835	5.8369694	0.037409503	5.835985	1
Molecule 2	5.88	5.8987913	0.03738053	5.8785267	1
Molecule 3	5.456	5.3855143	0.021775797	5.4498973	1
Molecule 4	5.494	5.504449	0.039567765	5.435588	1
Molecule 5	5.874	5.822976	0.05054345	5.8723636	1
Molecule 6	5.344	5.3920107	0.02178918	5.3459625	1
Molecule 7	5.628	5.625646	0.044379216	5.6338286	1
Molecule 8	5.25	5.191642	0.03986553	5.245896	1
Molecule 9	5.23	5.3400445	0.023216149	5.2255697	1
Molecule 10	5.176	5.2389107	0.033397097	5.18155	1
Molecule 11	5.89	5.9873424	0.039947327	5.8846445	1
Molecule 12	5.591	5.5947742	0.041592542	5.596765	1
Molecule 13	5.462	5.4341087	0.018637065	5.450952	1
Molecule 14	5.351	5.4504094	0.024300097	5.466835	1
Molecule 15	5.539	5.5351443	0.033346504	5.5395346	1
Molecule 16	5.571	5.4534445	0.024534578	5.484404	1
Molecule 17	5.417	5.4058266	0.019691233	5.4536986	1
Molecule 18	5.899	5.780898	0.039053217	5.8978524	1
Molecule 19	5.46	5.3940883	0.02180537	5.465711	1
Molecule 20	5.383	5.457014	0.021918012	5.4301295	1

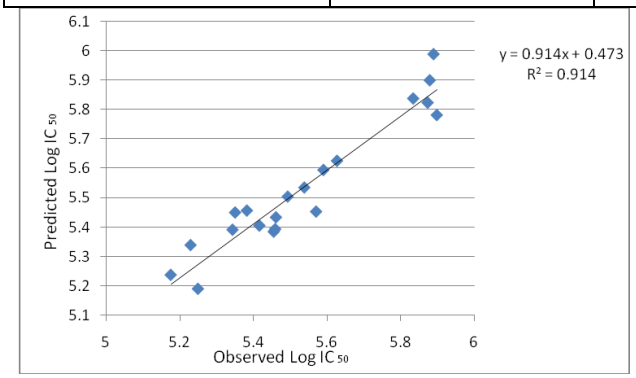


Fig.(1). Graphical Correlation of Observed (Log IC_{50}) Vs. Predicted (Log IC_{50}) MLR

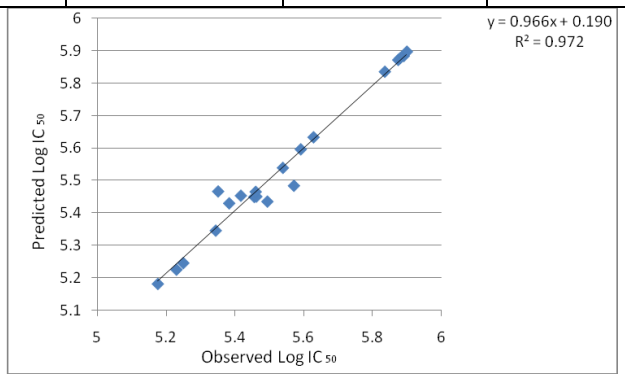


Fig. (2). Graphical Correlation of Observed (Log IC_{50}) Vs. Predicted (Log IC_{50}) SVM

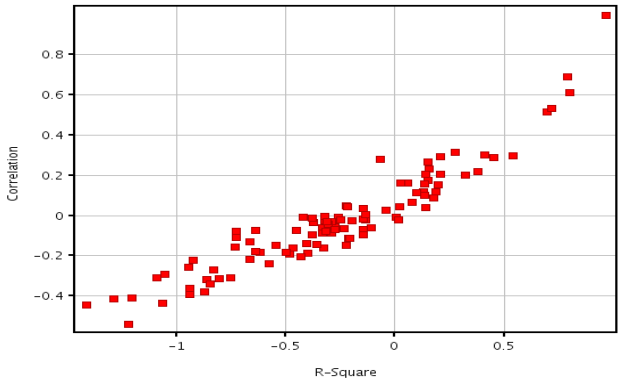


Fig. (3). Y-Scrambling for tetravariable MLR MODEL.

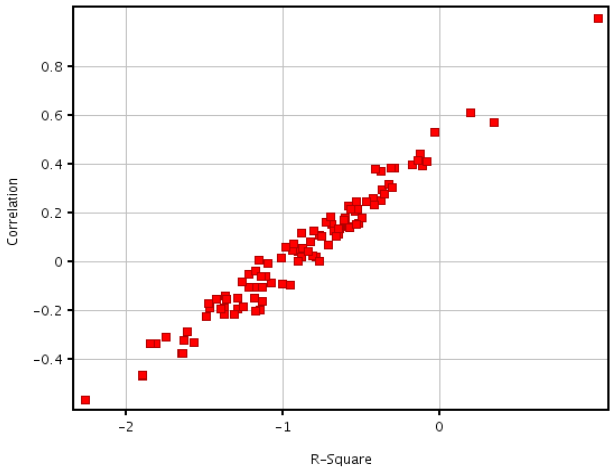


Fig. (4). Y-Scrambling for tetravariable MLR MODEL

References

1. Hurmath Unnissa S, Gayam Krishna Reddy, Aravazhi T,Synthesis and *in Vitro* Anti Tumor Activity of Some Novel 2, 3-Disubstituted Quinazolin 4(3H)-one Derivatives, Journal of Applied Pharmaceutical Science Vol. 3 (10), pp. 136-140, October, 2013

2. Rebecca L. Siegel, MPH ; Kimberly D. Miller, MPH ; Ahmedin Jemal, DVM, PhD, Cancer Statistics, CA CANCER J CLIN 2015;65:5–29,

3. Hanahan D, Weinberg A. The Hallmarks of Cancer. Cell 2000; 100:57-70.

4. Kalinowski DS, Quach P, Richardson DR ,Thiosemicarbazones: the new wave in cancer treatment. Future Med Chem. 2009 Sep;1(6):1143-51.

5. Umasankar Kulandaivelu, Valisakka Gari Padmini ,Kyatham Suneetha , annu Vincent Vidyasagar, Tadikonda Rama Rao, Arijit Basu, Venkatesan Jayaprakash, Synthesis, antimicrobial and anticancer activity of new thiosemicarbazone derivatives. Archiv der Pharmazie ,Volume 344, Issue 2, 84–90, February 2011.

6. Hamilton Mitsugu Ishiki, Antonia T. do Amaral, Three-Dimensional Quantitative Structure-Activity Relationship Study of Antitumor 2-Formylpyridine Thiosemicarbazones Derivatives as Inhibitors of Ribonucleotide Reductase. QSAR & Combinatorial Science, Volume 28, Issue 11-12, pages 1334–1345, December 2009.

7. Narayana S.H.N. Moorthy* , Nuno M.F.S.A. Cerqueira, Maria J. Ramos and Pedro A. Fernandes, Development of Ribonucleotide Reductase Inhibitor: A Review on Structure Activity Relationships, Mini-Reviews in Medicinal Chemistry, 2013,13(13):1862-72.

8. . Cerqueira NMFSA, Pereira S, Fernandes PA, Ramos MJ , Overview of ribonucleotide reductase inhibitors: an appealing target in anti-tumour therapy. Curr Med Chem ,2005 12(11):1283–1294.

9. Aksyonova, T.I., Volkovich, V.V., Tetko, I.V Robust polynomial neural networks in quantative- structure activity relationship studies. Syst Anal Model Simul, 2003. 43: 1331-1339.

10. Chao-Bin Xue[a, b], Li Zhang[b], Wan-Chun Luo[a,], Xian-Ye Xie[c], Lin Jiang[c], Ting Xiao[a].3D-QSAR and molecular docking studies of benzaldehyde thiosemicarbazone, benzaldehyde, benzoic acid, and their derivatives as phenoloxidase inhibitors ., Bioorganic & Medicinal Chemistry, 2007 Volume 15, Issue 5, Pages 2006–2015.

11. Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995; 20:273-97.

12. Niu B , Lu WC, Yang SS, Cai YD, Li GZ. Support vector machine for SAR/QSAR of phenethyl-amines. Acta Pharmacol Sin. 2007, 28(7):1075-86.

13. A Combinatorial Approach to the Variable Selection in Multiple Linear Regression: Analysis of Selwood *et al.* Data Set – A Case Study. QSAR & Combinatorial Science. Volume 22, Issue 6, pages 583–595, August 2003.

14 http://www.strandls.com/sarchitect/freetrial.php.

15. http://www.chemaxon.com/free-software/

16. http://www.vcclab.org/lab/pclient/

17. Pavlidis, P.; Wapinski, I.; Noble, WS; Support vector machine classification on the web. *Bioinformatics,* 2004, *20,* 586-587.

18. Vapnik, V.; The Support Vector method of function estimation. *U.S. Patent 5,* 950, 146, 1999.

19. Renu Vyas, S.S. Tambe, B.D. Kulkarni, Applications of Support Vector Machines as a Robust tool in High Throughput Virtual Screening. , 2012, International Journal for Computational Biology (IJCB) Vol.1, No.1, pp. 43-55

20. T. Scior , J.L. Medina-Franco , Q.-T. Do , K. Martínez-Mayorga , J. A. Yunes Rojas and P. Bernard. How to Recognize and Workaround Pitfalls in QSAR Studies: A Critical Review, *Current Medicinal Chemistry,* 2009, *16,* 4297-4313.

21. Polanski, J. ; Bak, A. ; Gieleciak, R. and Magdziarz, T. Modeling Robust QSAR. *J. Chem. Inf. Model.,* 2006, *46,* 2310-2318.

22. Esther Vicente , Pablo R Duchowicz , Eduardo Castro , Antonio Monge , QSAR analysis for quinoxaline-2-carboxylate 1,4-di-N-oxides as anti-mycobacterial

agents,journal of molecular graphics and modeling,2009,28,28-36.

23. Mohammad Goodarzi a, Richard Jensenb, Yvan Vander Heydena, QSRR modeling for diverse drugs using different feature selection methods coupled with linear and nonlinear regressions. Journal of Chromatography, Volume 910, 2012, Pages 84–94.
24. Javier García, , Pablo R. Duchowicz, María F. Rozas , José A. Carama, María V. Miríficoa, Francisco M. Fernández, Eduardo A. Castroa. A comparative QSAR on 1,2,5-thiadiazolidin-3-one 1,1-dioxide compounds as selective inhibitors of human serine proteinases, Journal of Molecular Graphics and Modelling 31 (2011) 10–19.

AUTHORS' CONTRIBUTIONS

Authors contributed equally to all aspects of the study.

PEER REVIEW

Not commissioned; externally peer reviewed.

CONFLICTS OF INTEREST

The authors declare that they have no competing interests.