

2D QSAR Study on Saponins of *Pulsatilla koreana* as an Anticancer Agent

Bhushan A Baviskar, Sharada L Deore, Sagar Alone

Government College of Pharmacy, Amravati-444 604, Maharashtra, INDIA.

ABSTRACT

In the present communication, total seventeen saponin molecules were collected and subjected to quantitative structure-activity relationship (QSAR) analyses. 2D QSAR model is developed by using multiple linear regression method against 4 different cell lines. QSAR model was generated by using training set of 11 and test set of 6 molecules having correlation coefficient (r^2), significant cross validated correlation coefficient (q^2) and F-test (For statistical significance) is as given below (A-549: r^2 - 0.9281, q^2 - 0.8691, F-test- 51.6079), (SK-OV-3: r^2 - 0.9554, q^2 - 0.9184, F-test- 85.7357), (SK-MEL-2: r^2 - 0.9160, q^2 - 0.8285, F-test- 43.6084), (HCT15: r^2 - 0.9203, q^2 - 0.8357, F-test- 46.1887). In this QSAR study Alignment independent descriptors such as T_2_C_7, T_O_O_5 and physicochemical descriptors

like Chain path count such as 6 chain count and Chi chain such as Chi 6 chain were most responsible descriptors for anticancer activity.

Key words: QSAR, *Pulsatilla koreana*, Cytotoxic activity, Multiple linear regressions.

Correspondence:

Bhushan A Baviskar

Government College of Pharmacy, Amravati-444 604, Maharashtra, INDIA.

E-mail: baviskarbhushan@gmail.com; sharudeore_2@yahoo.com

DOI: 10.5530/pc.2019.1.2.

INTRODUCTION

Cancer is a one of the dared disease characterized by progressive, abnormal and uncontrolled propagation of tissues. Currently, cancer is most leading cause of death of human being in world.^{1,2} A variety of anticancer agents have been developed over the last decades. Still, there is no doubt that further research into the design of new anticancer compounds with low toxicity and higher selectivity is needed.³⁻⁵ Natural products or herb extracts have been emerged as a novel approach to control a variety of diseases, including cancers. They are the basis for the traditional medicine philosophies and practices in China, India and tribal peoples.⁶ The *Pulsatilla koreana* (PKE) belongs to the family Ranunculaceae, in Korea is an important herbal medicine that has been used to treat amoebic dysentery, malaria⁷ and also demonstrate antifungal, antibiotic properties,⁸⁻⁹ anti-inflammatory, anti-acne activities.¹⁰ In addition, several studies have been reported that the saponins, mainly the oleanane and lupane-type saponins in PKE have anticancer effects in human melanoma, colon, hepatocellular and lung carcinoma.¹¹⁻¹⁵

Presently, computational methods and molecular modeling are the inseparable parts in drug design and discovery and no one can talk about drug design without having a bit comprehension about computational chemistry.¹⁶⁻¹⁸ Amongst the various computational technique exist in drug design, Quantitative structure-activity relationships (QSARs) have been found the major popularity.¹⁹ QSAR, as one of the most significant application in chemometrics, gives information which is helpful for design of new molecule acts on specific target.²⁰⁻²² QSAR have been developed since 1963, when Hansch and coworker developed quantitative relationships between activity and the octanol-water partition coefficient.²³ It is assumed that the sum of substituent effects on the steric, electronic and hydrophobic interaction of the compounds with their receptor determines the biological activity. Another criterion in constructing the QSAR models is finding one or more molecular descriptors that represent variation in the structural property of the molecules by a number.²⁴ These include both 2D (two dimensional) and 3D (three-dimensional) QSAR methods. The major differences of these methods can be analyzed from two viewpoints: (1) the structural parameters that are used to characterize molecular identities and (2) the mathematical procedure that is employed to obtain the quantitative relationship between a biological activity and the structural parameters.

The present communication deals with 2D QSAR analysis of 17 saponins molecules with natural resource in which eight lupane-type and nine oleanane-type by using multiple linear regression method against 4 different cell lines for anticancer activity.

MATERIALS AND METHODS

2D QSAR Methodology

Molecular structure generation and Data set:

A dataset of 17 saponins molecules isolated from the *Pulsatilla koreana* were taken from the literature and used for QSAR analysis.¹¹ The 2D-QSAR models were generated for this series using multiple linear regression (MLR) methods against four human solid tumor cell lines (A-549, SK-OV-3, SK-MEL-2 and HCT-15), those models which come out with promising results are discussed here. The structures and their inhibitory activities (ED_{50} values) are listed in Table 1. All the molecular modeling and statistical analysis were carried out by using Vlife MDS 4.0 software. The structures of the compounds were built using molecular sketching facilities offer in the modeling environment of Vlife engine on ASUS laptop with a Dual core processor and Windows 7 operating system. Energy minimization and batch optimization was performed using Merck Molecular force field. All the molecules were initially optimized and then used for the calculation of descriptors and further QSAR study.²⁵⁻²⁶

Selection of training and test set

The dataset of 17 molecules was divided into the training set (11 compounds) and test (6 compounds) for generating multiple linear regressions (MLR) 2D QSAR models.²⁷ Therefore, the care was taken in such a way that biological activities of all compounds in test lie within the maximum and minimum value range of biological activities of training set of compounds. The Uni-Column Statistics of test and training sets further reflected the correct selection of test and training sets. A Uni-Column statistics for training set and test set were generated to check correctness of selection criteria for trainings and test set molecules Table 1.

The maximum and minimum value in training and test set were compared in a way that:

Table 1: Uni-column statistics of the training and test sets for QSAR models.

Cell line	Training/Test set	Average	Max	Min	Std. Dev	Sum
A-549	Training set	94.1109	300.000	4.2400	114.9137	1035.2200
	Test set	131.5417	300.000	2.5600	138.64.97	789.2500
SK-OV-3	Training set	88.6536	300.0000	3.9500	112.7344	975.1900
	Test set	127.3850	300.0000	2.3100	139.3976	764.3100
SK-MEL-2	Training set	94.0000	300.0000	3.0400	116.5289	1034.0000
	Test set	136.1183	300.0000	1.5700	139.8473	816.7100
HCT-15	Training set	96.4845	300.0000	5.5000	114.9111	1061.3300
	Test set	138.8817	300.0000	8.3600	138.1815	833.2900

Table 2: Experimental and Predicted activity of *Pulsatilla koreana* saponin analogues.

Sr.No.	Saponins	ED ₅₀ (µM)							
		A-549		Sk-OV-3		SK-MEL-2		HCT15	
		Expt.	Pred.	Expt.	Pred.	Expt.	Pred.	Expt.	Pred.
1	PK01	>300.0	341.798	>300.0	333.826	>300.0	340.972	>300.0	340.723
2	PK02	>300.0	310.389	>300.0	313.994	>300.0	294.399	>300.0	294.808
3	PK03	>300.0	317.875	>300.0	308.737	>300.0	311.381	>300.0	311.289
4	PK04	>300.0	286.467	>300.0	288.905	>300.0	372.206	>300.0	272.733
5	PK05	38.14	80.1154	36.27	73.2989	37.44	119.877	40.06	122.038
6	PK06	11.25	17.2984	13.27	20.1392	3.04	11.9355	11.86	15.4923
7	PK07	145.62	104.038	120.38	93.2087	167.82	127.275	174.34	129.396
8	PK08	13.27	17.2984	11.41	14.9598	12.16	19.3334	13.58	22.8506
9	PK09	135.28	104.038	114.32	93.2087	165.54	112.479	171.29	114.68
10	PK10	2.56	17.2984	2.31	14.9598	1.57	4.53761	8.36	8.13401
11	PK11	43.64	80.1154	38.90	68.1195	39.67	90.2853	49.04	92.6047
12	PK12	13.49	-6.624	13.71	-10.1294	14.12	-17.6561	14.17	-13.9409
13	PK13	155.32	104.038	124.12	93.2087	145.68	127.275	138.73	129.396
14	PK14	4.24	17.2984	3.95	14.9598	3.47	19.3334	5.50	22.8506
15	PK15	41.35	80.1154	38.91	68.1195	40.02	75.4896	40.86	77.8882
16	PK16	9.58	-6.624	11.39	-10.1294	10.37	-32.4518	12.74	-28.6574
17	PK17	10.73	41.2208	10.66	40.049	9.81	48.925	14.09	52.2837

Expt. - Experimental activity; Pred. - Predicted activity

1. The maximum value of test set should be less than or equal to maximum value of training set.
2. The minimum value of test set should be higher than or equal to minimum value of training set.

This observation showed that test set was interpolative and derived within the minimum–maximum range of training set. The mean and standard deviation of ED₅₀ values of sets of training and test provide insights to the relative difference of mean and point density distribution (Along mean) of the two sets (www.Vlifesciences.com).²⁸

QSAR models were generated by a training set of 11 molecules for each model. And a test set of 6 molecules with uniformly distributed biological activities. The structures of all the compounds along with their actual and predicted biological activities are presented in Table 2 and used as dependent variables in 2D QSAR calculations.

Descriptor calculation

The basis of energy minimization is that the drug binds to effectors/receptors in the most stable form, i.e., the minimum energy form. For

2D-QSAR study requires the calculation of molecular descriptors. A large number of theoretical 2D descriptors such as individual, Mol. Wt., Volume, XlogP, smr; retention index (chi), atomic valence connective index (chiv), path count, chi chain, chiv chain, chain path count, cluster, path cluster, kappa, physiochemical such as Estate Numbers, Estate contributions, Polar Surface Area, Element Count, Dipole moment, Hydrophobicity XlogpA, Hydrophobicity SlogpA; distanced based topological, alignment-independent and atom-type have been computed for these geometrically optimized structures from the chemical structures of the compounds referred to above with a view to develop structure–activity.²⁹

Regression analysis

The 2D-QSAR models were generated with stepwise multiple linear regression (MLR) analyses using forward–backward technique set as 0.0 kcal/mol cut-off, applying auto scaling. It relates the dependent variable y (biological activity) to a number of independent variables xi (molecular descriptor) by using linear equations. MLR is the traditional

and standard approach for multivariate data analysis.

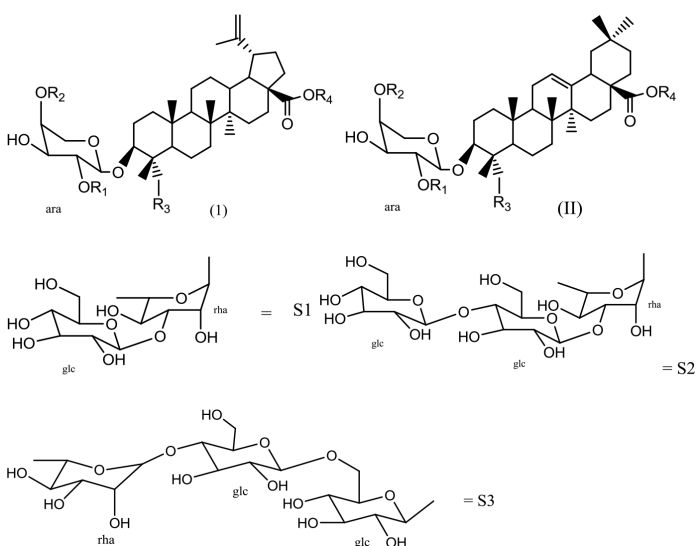
$$Y = b_1 \times x_1 + b_2 \times x_2 + b_3 \times x_3 + c$$

Where Y is dependent variable, 'b's are regression coefficients for corresponding 'x's independent variable, 'c' is a regression constant or intercept.^{30,31}

The program computes the best model on the basis of statistical parameters including optimum number of components (N), square of correlation coefficient (r²), cross-validated correlation coefficient (q²), r² for external test set (pred_r²), F-test (Fischer's value) for statistical significance and standard error of estimation (r² se).

From the statistical viewpoint, the ratio of the number of samples (N) to the number of variables used (M) should be N/M ≥ 5.

Structures of Isolated Pulsatilla saponins



Sr. No	Saponin	Aglycogen	R1	R2	R3	R4
1	PK01	(I)	Rha	Glc	OH	S3
2	PK02	(II)	Rha	Glc	OH	S3
3	PK03	(I)	Rha	H	OH	S3
4	PK04	(II)	Rha	H	OH	S3
5	PK05	(I)	Rha	Glc	OH	H
6	PK06	(II)	Rha	Glc	OH	H
7	PK07	(I)	S1	H	OH	H
8	PK08	(II)	S1	H	OH	H
9	PK09	(I)	Rha	GLC	H	H
10	PK 10	(II)	Rha	Glc	H	H
11	PK 11	(I)	Rha	H	OH	H
12	PK 12	(II)	rha	H	OH	H
13	PK 13	(I)	S1	H	H	H
14	PK 14	(II)	S1	H	H	H
15	PK 15	(I)	H	Glc	OH	H
16	PK 16	(II)	H	Glc	OH	H
17	PK 17	(II)	S2	H	OH	H

Lupane-type saponins expressed as (I) and oleanane-type saponins as (II).

Validation of QSAR model

The generated QSAR model was validated for predictive ability inside the model by using cross validation (leave-one-out—LOO) for q² and external validation, which is more robust alternative method by dividing the data into training set and test set.³² For calculating q², each molecule in the training set was eliminated once and the activity of the eliminated molecule was predicted by using the model developed by the remaining molecules. The q² was calculated using the equation Eq. 2 which describes the internal stability of a model.

$$q^2 = 1 - \frac{\sum [y_i(\text{Act}) - y_i(\text{Pred})]^2}{\sum [y_i - y(\text{mean})]^2}$$

Where *y_i Act* and *y_i Pred* are the actual and predicted activity of the *i*th molecule in the training set, respectively and *y mean* is the average activity of all molecules in the training set.

The predictive ability of the selected model was also confirmed of test set compounds which is also denoted with pred_r². The pred_r² value is calculated as follows

$$\text{pred}_r^2 = 1 - \frac{\sum [y_i(\text{Act}) - y_i(\text{Pred})]^2}{\sum [y_i - y(\text{mean})]^2}$$

Where *y_i* and *y_i* are the actual and predicted activity of the *I* th molecule in the training set, respectively and *y mean* is the average activity of all molecules in the training set. The statistical significance of selected 2D-QSAR model was further supported by the 'fitness plot' obtained, this is a plot of observed versus predicted activity of training and test set compounds and provides an idea about how well the model was trained and how well it predicts the activity of the external test set (Figure 1). The contribution chart for the significant model is presented in Figure 2, which gives the percentage contribution of the descriptors used in deriving the model.

RESULTS AND DISCUSSION

The aim of our study was to developed 2D-QSAR model for a series of analogs of *Pulsatilla koreana* saponin by using MLR method. 2D-QSAR equations were selected by optimizing the statistical results generated along with variation of the descriptors in this model. The fitness/pattern plots were also generated for evaluating the dependence of the biological activity on various different types of the descriptors. The frequency of use of a particular descriptor in the population of equations indicated the relevant contributions of the descriptors.

The best regression equation obtained is represented in Equation:

- For (A-549): ED₅₀ = + 55.3309(± 6.5096) T_{2_C_7} + 23.9224(± 8.9012) 6ChainCount - 616.7280
- For (SK-OV-3): ED₅₀ = + 53.2371(± 0.5713) T_{2_C_7} + 553.1007(± 130.2520) chi6chain - 612.2515
- For (SK-MEL-2): ED₅₀ = + 53.9707(± 7.2274) T_{2_C_7} + 7.3979(± 2.5587) T_{O_O_5} - 493.8088
- For (HCT15): ED₅₀ = + 53.2728(± 6.9414) T_{2_C_7} + 7.3583(± 2.4574) T_{O_O_5} - 484.2732

The results of statistical parameters are summarized in Table 3. From the QSAR studies it was found that the Alignment Independent category in which T_{2_C_7} descriptor is most powerful descriptor responsible for anticancer activity to all 4 types of cell lines at range of 70% - 80% A-549, SK-OV-3, SK-MEL-2 and HCT15. This descriptor signifies No. of double bounded atoms (i.e.- Any double bonded atoms, T-2) separated from carbon atom by 7 bonds in a molecule. T_{O_O_5} descriptor is responsible for anticancer activity in SK-MEL-2 and HCT15 cancer cell line at range 20%-30% and signifies No. of oxygen atoms (Single double

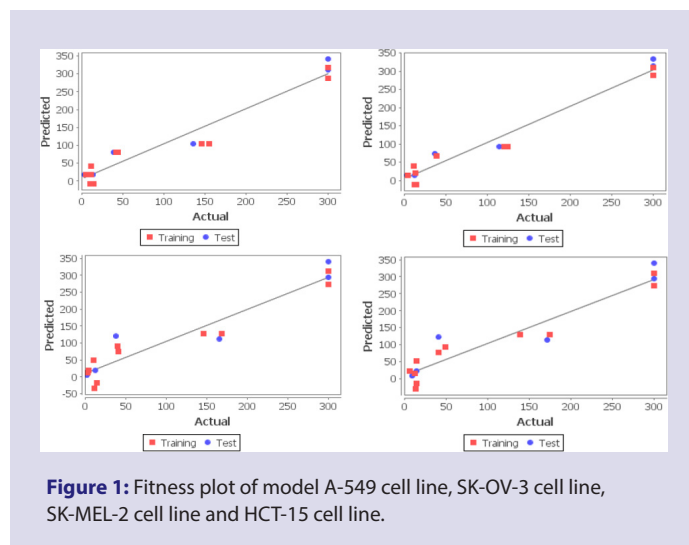
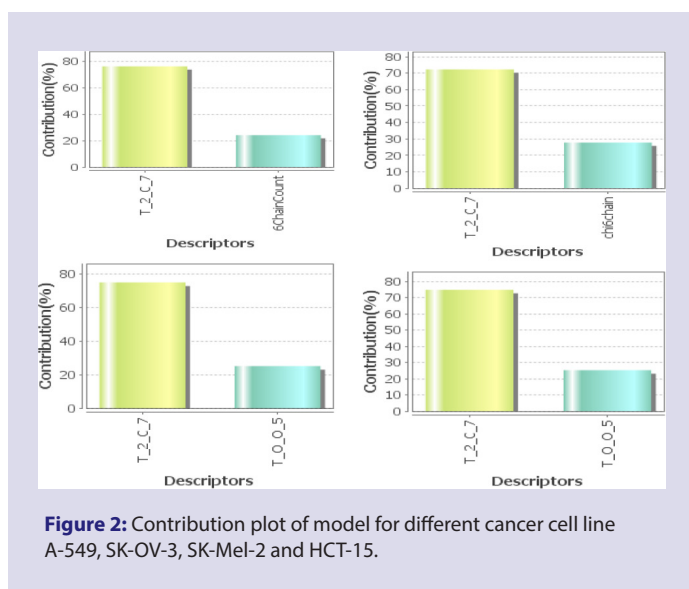
Table 3: Statistical results of 2D QSAR equation generated by MLR method.

Statistics	A-549	SK-OV-3	SK-MEL-2	HCT-15
N	11	11	11	11
DOF	8	8	8	8
r ²	0.9281	0.9554	0.9160	0.9203
q ²	0.8691	0.9187	0.8285	0.8357
F test	51.6079	85.7357	43.6084	46.1887
r ² se	34.4579	26.6109	37.7639	36.2697
q ² se	46.4909	35.9315	53.9485	52.0702
pred_r ²	0.9538	0.9686	0.8950	0.8899
pred_r ² se	31.0692	25.8051	47.7110	48.3659

N - Number of molecules, K - Number of descriptors in a model, DOF - Degree of freedom (higher is better), r² - Coefficient of determination (>0.7), q² - Cross-validated r (>0.5), pred_r² - r for external test set (>0.5), F-test - F-test for statistical significance of the model (higher is better, for same set of descriptors and

or triple bonded) separated from other oxygen by 5 bond distance in a molecule. 6 chain count descriptor signify total numbers of six membered rings in a compound and chi6chain descriptors signifies a retention index for six membered rings, responsible for anticancer activity for A-549, SK-OV-3 respectively. Increase in six membered rings in the *Pulsatilla koreana* saponin analogues increases anticancer activity. It was found that all descriptors have positive correlation with the activity. Increase in all descriptor will help to increase the anticancer activity of *Pulsatilla koreana* saponins analogue. Existence of free carboxylic group is at C-28 position is most responsible for cytotoxic activity and hydroxyl group at C-23 had a negative effect on the cytotoxic activity. It is due to electron donating effect of two loan pair towards C-3 of aglycon. The results obtained from this 2D-QSAR study are in agreement with the observed SAR of *Pulsatilla koreana* saponin studied. There is difference in cytotoxic between oleanane and lupane saponin generally cytotoxicity of lupane type saponin were much weaker than those of oleanane type saponin that means oleanane type *Pulsatilla koreana* saponin analogues having more anticancer activity than that of lupane type.

Statistical significance of these models was further supported by 'fitness plot' (Figure 1) obtained for each model; this is a plot of experimental versus predicted activity of training and test set compounds and provides

**Figure 1:** Fitness plot of model A-549 cell line, SK-OV-3 cell line, SK-MEL-2 cell line and HCT-15 cell line.**Figure 2:** Contribution plot of model for different cancer cell line A-549, SK-OV-3, SK-Mel-2 and HCT-15.

an idea about how fit the model was trained and how well it predicts activity of external test set. Contribution charts for all the significant models are presented in Figure 2, which gives percentage contribution of descriptors used in deriving the QSAR models.

CONCLUSION

In order to determine anticancer activities of various *Pulsatilla koreana* saponins may be treated statistically to find out the structural characteristics which are essential for high activity. 2D-QSAR has been carried out on *Pulsatilla koreana* saponins analogue. It was found that all descriptors have positive correlation with the activity and all descriptor will helpful to increase the anticancer activity of *Pulsatilla koreana* saponins analogue. Hence, the model proposed in this work is useful and can be employed to design new saponin derivative of *Pulsatilla koreana* as most active anticancer agent.

ACKNOWLEDGMENT

Authors are thankful to entire team of Vlife MDS for their constant support.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- Varmus H. The new era in cancer research. *Science*. 2006;312(5777):1162-5. doi:10.1126/science.1126758
- Kingston DG, Newman DJ. Taxoids: cancer-fighting compounds from nature. *Curr Opin Drug Disc Dev*. 2007;10(2):130-44.
- Cook SJ, Wakelam M. The search for new targets for cancer chemotherapeutics. *Curr Opin Pharmacol*. 2005;5(4):341-2.
- Fan QL, Zou WY, Song LH, Wei W. Synergistic antitumor activity of TRAIL combined with chemotherapeutic agents in A549 cell lines *in vitro* and *in vivo*. *Cancer Chemother Pharmacol*. 2005;55(2):189-96.
- Inagawa H, Nishizawa T, Honda T, Nakamoto T, Takagi K, Soma G. Mechanisms by which chemotherapeutic agents augment the antitumor effects of tumor necrosis factor: involvement of the pattern shift of cytokines from Th2 to Th1 in tumor lesions. *C Anticancer Res*. 1998;18(5D):3957-64.
- Newman DJ, Cragg GM, Snader KM. The influence of natural products upon drug discovery. *Nat Prod Rep*. 2000;17(3):215-34.
- Bae KH. The medicinal plants of Korea. Seoul: Kyo-Hak Press. 1999;139.
- Martin ML, Roman LS, Dominguez A. *In vitro* activity of protoanemonin, an antifungal agent. *Planta Med*. 1990;56(01):66-9.
- Campbell WE, Cragg GML, Powrie AH. Anemonin, protoanemonin and ranunculin

- from *Knowltonia capensis*. Phytochemistry. 1979;18:323-4.
10. Lee HS, Beon MS, Kim MK. Selective growth inhibitor toward human intestinal bacteria derived from *Pulsatilla cernua* root. J Agric Food Chem. 2001;49(10):4656-61.
 11. Bang SC, Lee JH, Song GY, Kim DH, Yoon MY, Ahn BZ. Antitumor activity of *Pulsatilla koreana* saponins and their structure-activity relationship. Chem Pharm Bull (Tokyo). 2005;53(11):1451-4.
 12. Kim Y, Kim SB, You YJ, Ahn BZ. Deoxypodophyllotoxin; The cytotoxic and antiangiogenic component from *Pulsatilla koreana*. Planta Med. 2002;68(03):268-71.
 13. Kim SY, Kim SB. Anti-tumor effects of extracts of *Pulsatilla koreana* (SB-31) *in vitro*. J Kor Cancer Assoc. 1994;26(6):959-63.
 14. Yong K, Seong-Cheol B, et al. *Pulsatilla* Saponin D: the Antitumor Principle from *Pulsatilla koreana*. Arch Pharm Res. 2004;27(9):915-8.
 15. Hong SW, Jung KH, Lee HS, et al. Apoptotic and anti-angiogenic effects of *Pulsatilla koreana* extract on hepatocellular carcinoma. Int J Oncol. 2012;40(2):452-60.
 16. Castagnoli N, Kier L. Looking to the Future: Computational Methods in Drug Design Chem. Biodivers. 2005;2(11):1409-10.
 17. Kiss T, Erdi P. From electric patterns to drugs: perspectives of computational neuroscience in drug design. Biosystems. 2006;86(1-3):46-52.
 18. Ortiz AR, Gomez-Puertas P, Leo-Macias A, Lopez-Romero P, et al. Computational Approaches to Model Ligand Selectivity in Drug Design. Curr Top Med Chem. 2006;6(1):41-55.
 19. Fujita T, Timmerman H. QSAR and Drug Design, New Developments and Applications. 1st ed., Elsevier, Amsterdam. 1995;435.
 20. Hansch C, Hoekman D, Gao H. Comparative QSAR: Toward a Deeper Understanding of Chemicobiological Interactions. Chem Rev. 1996;96(3):1045-76.
 21. Hemmateenejad B. Optimal QSAR analysis of the carcinogenic activity of drugs by correlation ranking and genetic algorithm-based PCR. J Chemometr. 2004;18(11):475-85.
 22. Seierstad M, Agrafiotis DK. A QSAR Model of hERG Binding Using a Large, Diverse and Internally Consistent Training Set. Chem Biol Drug Des. 2006;67(4):284-96.
 23. Hansch C, Maloney PP, Fujita T, Muir RM. Correlation of Biological Activity of Phenoxyacetic Acids with Hammett Substituent Constants and Partition Coefficients. Nature. 1962;194(4824):178-80.
 24. Hemmateenejad B, Sancholi M. Substituent electronic descriptors for fast QSAR/QSPR. J Chemometr. 2007;21(3-4):96-107.
 25. Jain AK, Ravichandran V, Singh R, Mourya VK, Agrawal RK, et al. Three-dimensional qsar study of 2,4-disubstituted-phenoxy acetic acid derivatives as a CRTh2 receptor antagonist: using the k-nearest neighbor method. Dig J Nano Biostruc. 2008;3(2):63-73.
 26. Prasad RK, Narsinghani T, Sharma R. QSAR analysis of novel N-alkyl substituted isatins derivatives as anticancer agents. J Chem Pharm Res. 2009;1(1):199-206.
 27. Hudson BD, Hyde RM, Rahr E, Wood J. Parameter based methods for compounds selection from chemical databases. Quant Struct Act Relat. 1996;15(4):285-9. doi:10.1002/qsar.19960150402.
 28. V-LifeMolecular Design Suite 3.0, VLife Sciences Technologies Pvt. Ltd; Baner Road: Pune, Maharashtra, India. www.Vlifesciences.com
 29. Shen M, Xiao Y, Golbraikh A, Gombar VK, Tropsha A. Development and validation of k-nearest-neighbor QSPR models of metabolic stability of drug candidates. J Med Chem. 2003;46(14):3013-20.
 30. Croux C, Joossens K. Influence of observations on the misclassification probability in quadratic discriminant analysis. J Multivar Anal. 2005;96(2):348-403.
 31. Devillers J. Neuronal network in QSAR and drug design. Academic Press, London. 1996.
 32. Kubyani H. Variable selection in QSAR studies: an evolutionary algorithm. Quant Struct Act Relat. 1994;13(3):285-94.