Chemometric 3D-Molecular Descriptors in Correlation with Insecticidal Activity: The Derivatives of Novel Semicarbazide and Chalcone

[1]Nidhi Shekhawat, [2]Prithvi Singh

Department of Zoology
Vedanta Post-Graduate Girls' College
Reengus-332 404, India

***Abstract*- Two novel series of compounds bearing insecticidal properties were quantitatively examined using chemometric 3D-descriptors. The semicarbazide, thiosemicarbazide, urea, and thiourea derivatives make up the first series, while the chalcone counterparts make up the second series. These series have demonstrated insecticidal actions against Spodoptera littoralis and Spodoptera frugiperda larvae, respectively. To correlate their insecticidal effects, a total of 675 descriptors for the compounds in the first series and 651 descriptors for the compounds in the second series were initially deemed suitable. The DRAGON software was used to construct these characteristics for the energy-minimized 3D-structures of each compound individually for the two series. In order to produce statistically significant models explaining the highest significant correlations between insecticidal actions and the most influential descriptors, the descriptors were then subjected to filtering processes embedded in the combinatorial protocol in multiple linear regression, CP-MLR, computational software. Twenty and eleven descriptors in all were determined to be significant in the first and second series' statistical models, respectively. Six significant models for the larvae of S. littoralis and the same number for the larvae of S. frugiperda have been included in their respective Tables. One most crucial model, however, has finally been taken into consideration for further discussion for both 2nd and 4th instar larvae of S. littoralis and S. frugiperda. The 3D-Morse signal 22 (Mor22m) and R autocorrelation of lag 2 (R2m) were the two descriptors that had the most bearing on the first series. Atomic masses were used to weigh each of these descriptors. The radial distribution functions-4.0 and -11.5, each weighted by atomic polarizabilities (RDF040p and RDF115p), the unweighted radial distribution function-5.5 (RDF055u), and the atomic Sanderson electronegativities weighted radial distribution function-11.0 (RDF110e) were the important descriptors filtered for the second series. The direction of their influence on the activity profile for the 2nd and 4th instar larvae of S. littoralis and S. frugiperda, respectively, has been revealed by the sign of the regression coefficient linked to these descriptors. As all the compounds from both series were present in their respective domains and the significant models properly predicted the insecticidal activities of all the compounds from both series, the applicability domain (AD) analysis has revealed that the models under consideration had appropriate predictability. The guidelines given in the discussion may be useful for looking into new potential analogues of each series.**

***Keywords*: Chemometric 3D-descriptors, QSAR, Insecticidal activity, Semicarbazide, thiosemicarbazide, urea and thiourea derivatives, Chalcone analogues.**

INTRODUCTION

Insecticidal impact against Spodoptera littoralis (Boisd.) and Spodoptera frugiperda (Lepidoptera: Noctuidae) have recently been reported for two novel series of compounds [1,2]. The first series includes derivatives of urea, thiourea, semicarbazide, and thiosemicarbazide, whereas the second series includes chalcone analogues. The first series of compounds have demonstrated insecticidal activity against S. littoralis, a species of moth in the Noctuidae family. The larval stage of this species or pest primarily causes the cotton crop's enormous and astounding harm [3]. Its larvae have at least seven developmental stages and can harm a variety of other plants and crops during the cotton season [4-7]. The insecticidal action against S. frugiperda, one of the most dangerous insects that might harm maize and other crops, was triggered by the second series chalcone derivatives [8–10]. Cotton, sugar cane, rice, sorghum, and other significant crops can all be quickly destroyed by S. frugiperda insects in a couple of days [11,12]. Researchers in the fields of agriculture and plant protection have periodically created a number of novel compounds with improved activity profiles to stop the spread of various dangerous pests and/or insects. The main issue they have is coming up with new chemical compounds that may have no or minimum adverse effects and are also safe for humans, animals, and the environment in addition to synthesizing and evaluating biological functions.

The first series of analogues of the semicarbazide, thiosemicarbazide, urea, and thiourea were described as novel, efficient pesticides against S. littoralis cotton [1] and their biological activities in opposition to cotton leafworm were assessed. The insecticidal effects of these analogues were tested on S. littoralis larvae in their 2nd and 4th instars.

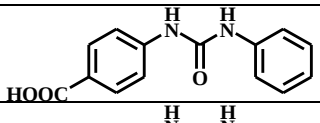
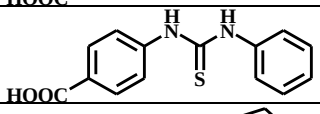
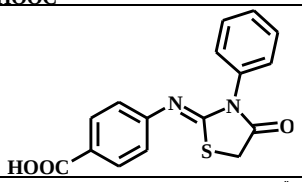
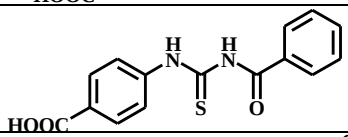
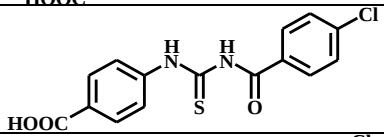
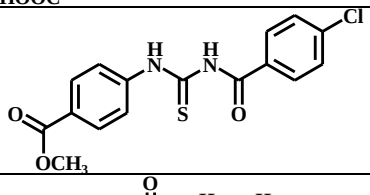
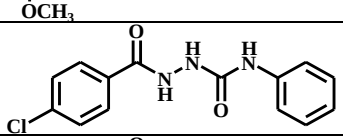
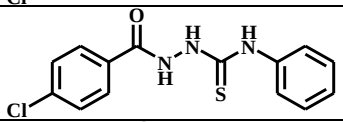
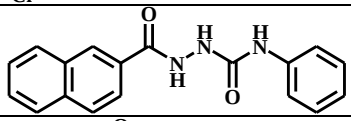
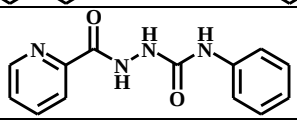
The second series of chalcones is a significant class of natural compounds that is a member of the flavonoid family [13] and has demonstrated a range of biological consequences. Some of these include antiparasitic [14], antileishmanial [14], anticancer [15], anti-inflammatory [15], anti-cardiovascular [15], and antitumor [15] actions. Chalcones are extremely reactive chemicals that are also used as protective agents in agricultural fields [17,18] because they include a conjugated enone moiety. According to the recent publication [2], the newly synthesized novel chalcones are particularly efficient compounds against the 2nd and 4th instar larvae of S. frugiperda.

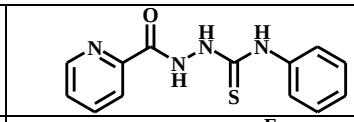
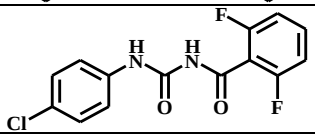
The current study's objective is to quantitatively analyze the molecular parameters associated with the insecticidal effects for each of two series of chemicals separately. Quantitative structure-activity relationship analysis (QSAR) is the term used to describe the method. To forecast more potent analogues of the series' insecticidal activities, the generated QSAR between those activities and molecular descriptors may be further utilized. Furthermore, the generated statistical models may be useful in predicting the molecular mechanism of their action at molecular level.

MATERIALS AND METHODS

The values for the compounds' insecticidal activity against S. littoralis and S. frugiperda larvae in their 2nd and 4th instars are reported as LC_{50}(ppm), where LC_{50}(ppm) denotes the lethal concentration in parts per million (ppm) required to elicit 50% of the desired action. The substances under investigation and their estimates of insecticidal activity are taken from published works [1,2]. For a compound, the estimate is expressed on logarithmic scale as logLC_{50}(ppm). The semicarbazide, thiosemicarbazide, urea, and thiourea representative compounds from the first series, which has a total of 12 compounds, are included in Table 1.

Table 1. Observed and calculated lethal concentrations of semicarbazide, thiosemi-carbazide, urea and thiourea derivatives against 2nd and 4th instar larvae of Spodoptera littoralis

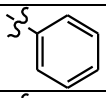
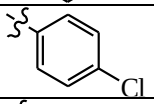
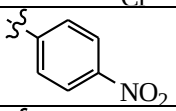
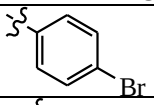
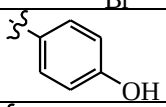
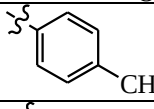
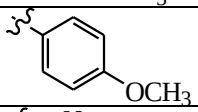
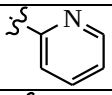
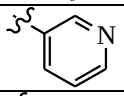
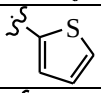
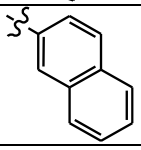
S. No.	Compound	3D-Descriptor[a]		log LC_{50} (ppm)[b]			
		Mor22m	R2m	Obsd. 2nd instar larvae	Calcd. Eq.(3)	Obsd. 4th instar larvae	Calcd. Eq.(6)
1		0.598	0.000	1.867	2.024	2.215	2.239
2		0.722	0.497	1.654	1.592	2.088	2.133
3		0.484	0.069	1.933	1.889	2.179	2.188
4		1.000	0.419	1.777	1.831	2.214	2.232
5		0.857	1.000	1.245	1.160	2.046	2.029
6		0.562	0.865	0.995	1.127	2.011	1.985
7		0.383	0.497	1.329	1.397	2.077	2.039
8		0.838	0.486	1.734	1.670	2.186	2.169
9		0.665	0.065	1.970	1.998	2.221	2.239
10		0.425	0.004	2.010	1.920	2.238	2.190

11		0.482	0.047	1.984	1.910	2.200	2.193
Dimilin		0.000	0.912	0.774	0.754	1.778	1.816

[a]Scaled 3D-descriptors.[b]Obsd. LC50 values are taken from ref. (1) and are expressed on logarithmic scale.

The second series, which consists of 11 chalcone derivatives, is listed in Table 2 and in Figure 1, the generalized molecular structure of these congeners is depicted.

Table 2. Observed and calculated lethal concentrations of chalcone derivatives against 2nd and 4th instar larvae of Spodoptera frugiperda (see Figure 1 for general structure)

S. No.	R	3D-Descriptor[a]				log LC_{50} (ppm)[b]			
		RDF040p	RDF115p	RDF055u	RDF110e	Obsd. 2nd instar larvae	Calcd. Eq.(9)	Obsd. 4th instar larvae	Calcd. Eq.(12)
1		0.126	0.257	0.338	0.630	1.552	1.610	1.949	1.968
2		0.745	0.144	0.041	0.000	0.983	1.020	1.868	1.882
3		1.000	0.522	1.000	0.483	0.976	0.944	1.826	1.821
4		0.173	0.172	0.041	0.490	1.591	1.535	1.981	1.989
5		0.093	0.168	0.000	0.359	1.633	1.604	1.983	1.967
6		0.164	0.944	0.129	0.916	1.800	1.846	2.057	2.066
7		0.236	0.289	0.346	0.369	1.419	1.525	1.877	1.910
8		0.000	0.280	0.284	0.827	1.750	1.730	2.033	2.020
9		0.022	0.128	0.226	0.550	1.692	1.650	2.006	1.970
10		0.032	0.000	0.362	0.395	1.608	1.592	1.937	1.912
11		0.404	1.000	0.491	1.000	1.710	1.658	2.010	2.022

[a]Scaled 3D-descriptors.[b]Obsd. LC50 values are taken from ref. (2) and are expressed on logarithmic scale.

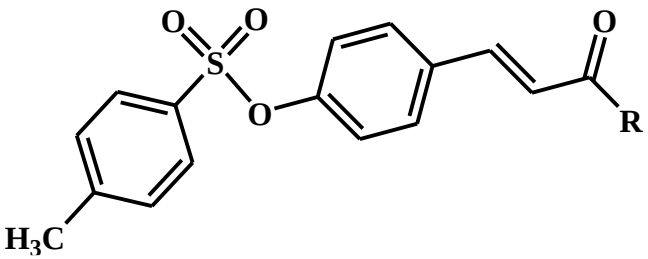


Figure 1. General structure of chalcone derivatives

The internal robustness of created statistical models has been evaluated using the leave-one-out (LOO) and leave-five-out (L5O) techniques.

Descriptors

The compounds' structures (Tables 1 and 2) were drawn using the conventional method in 2D-ChemDraw [19]. Next, using DRAGON software [20], the energy-minimized 3D-modules of these structures were used to compute the molecular descriptors pertaining to the 0D- through 2D- and 3D-classes and gathered in separate files for two sets of compounds. In the current investigation, the models created using 3D-descriptors seemed to be more significant than the models created using 0D-2D-descriptors. The examination of the insecticidal activity of both series of compounds resulted in the identification of a total of 27 3D-descriptors. Table 3 lists many classes of these descriptors that address structural aspects, as well as their definition and scope.

Table 3. 3D-descriptors, their classes and physical meanings, used for analysis of insecticidal activity of both series of compounds

No.	Descriptor	Class	Physical meaning
1	DISPp	Geometrical	d COMMA2 value/weighted by atomic polarizabilities
2	H6e	GETAWAY	H autocorrelation of lag6/weighted by atomic Sanderson electro-negativities
3	H6m	GETAWAY	H autocorrelation of lag 6 / weighted by atomic masses
4	HATSe	GETAWAY	Leverage-weighted total index/weighted by atomic Sanderson electronegativities
5	HATS1m	GETAWAY	Leverage-weighted autocorrelation of lag 1/weighted by atomic masses
6	HATS2m	GETAWAY	Leverage-weighted autocorrelation of lag 2/weighted by atomic masses
7	HATS1p	GETAWAY	Leverage-weighted autocorrelation of lag 1/weighted by atomic polarizabilities
8	HATS1v	GETAWAY	Leverage-weighted autocorrelation of lag 1/weighted by atomic van der Waals volumes
9	Mor09e	3D-MoRSE	3D-Morse–signal09/weighted by atomic Sanderson electronegativities
10	Mor14e	3D-MoRSE	3D-Morse–signal14/weighted by atomic Sanderson electronegativities
11	Mor15m	3D-MoRSE	3D-Morse–signal15/weighted by atomic masses
12	Mor22m	3D-MoRSE	3D-Morse–signal22/weighted by atomic masses
13	Mor09u	3D-MoRSE	3D-Morse–signal09 / unweighted
14	Mor14u	3D-MoRSE	3D-Morse–signal14/unweighted
15	RDF055e	RDF	Radial Distribution Function–5.5/weighted by atomic Sanderson electronegativities
16	RDF110e	RDF	Radial Distribution Function–11.0/weighted by atomic Sanderson electronegativities
17	RDF130m	RDF	Radial Distribution Function–13.0/weighted by atomic masses
18	RDF140m	RDF	Radial Distribution Function–14.0/weighted by atomic masses
19	RDF040p	RDF	Radial Distribution Function–4.0/weighted by atomic polarizabilities
20	RDF115p	RDF	Radial Distribution Function–11.5/weighted by atomic polarizabilities
21	RDF055u	RDF	Radial Distribution Function–5.5/unweighted
22	RDF110u	RDF	Radial Distribution Function–11.0/unweighted
23	RDF040v	RDF	Radial Distribution Function–4.0/weighted by atomic van der Waals volumes
24	RDF115v	RDF	Radial Distribution Function–11.5/weighted by atomic van der Waals volumes
25	R1m	GETAWAY	R autocorrelation of lag 1/weighted by atomic masses
26	R2m	GETAWAY	R autocorrelation of lag 2/weighted by atomic masses
27	R5e+	GETAWAY	R maximal autocorrelation of lag 5/weighted by atomic Sanderson electronegativities

As the magnitude of the identified descriptors differs greatly from one another, the imbalanced regression coefficients and the intercept of the regression model would reflect this. The data-set descriptors were further scaled [21] between 0 and 1 to prevent this. Comparatively to unscaled descriptors with higher or lower values, such descriptors would assign equal weights in a particular model.The QSAR models in various scaled descriptors were then developed using the combinatorial protocol in multiple linear regression (CP-MLR) computational process [22].

Regression Models

One of the important tasks in a QSAR analysis is to choose highly significant descriptors from the multivariate space in order to derive meaningful models. The CP-MLR is one of a vast variety of techniques that uses a 'filter'-based variable selection method to organize the selection process and produce unique statistically significant models. Our earlier papers [23–28] covered

the specifics and application of this strategy. Four filters have been embedded in the CP-MLR analysis computation software. Only those descriptors having inter-descriptor correlations with the dependent variable that are greater than or equal to 0.79 can be entered using the first filter. The second filter regulates the entry of descriptors into a regression model by thresholding their coefficients' t-values at 2.0. The third filter enables comparison of models with various descriptor counts using the model equation's r-bar, which is the square root of the adjusted multiple correlation coefficient. The fourth filter measures the model's internal robustness using the leave-one-out index Q^2_{LOO}, such that $0.3 \leq Q^2_{LOO} \leq 1.0$. As a new upper limit for subsequent model creation, the r-bar value (third filter) of the prior optimum model was improved with an increase in the number of useful descriptors. Each cross-validated model was put through the randomization test by repeatedly randomizing the activity profile in order to look for any chance correlations [29,30]. Every model was put through 100 simulation runs with random activity for this. To describe the percent chance correlation of the model under discussion, the scrambled activity models with regression statistics better than or equal to those of the original activity model were counted. The multiple correlation coefficient (r), standard deviation (s), and F-ratio between the variances of calculated to observed activities were used to evaluate a model's statistical significance ($F_{n, n-k-1}$, n is number of compounds and k is number independent descriptors). The leave-one-out and leave-five-out procedures were used to determine the internal validation, and the resulting statistical indices Q_{LOO} and Q_{L5O}, which are both greater than 0.5, indicate an internally reliable model. The Kubinyi function, FIT [33,34], the Friedman's lack of fit, LOF [35], and the Akaike's information criteria, AIC [31,32], have also been derived to evaluate the best model. Furthermore, the model can be externally validated if the data set includes a sizable number of substances. External validation is not feasible for the current investigation, however, because only a small number of chemicals are present in both series. A study on the applicability domain (AD) was also conducted to make sure that all congeners of a given series remained inside its domain. This means that the compounds used to construct a robust model have a greater chance of predicting future analogues of the series if they also share the same AD.

Applicability Domain

The significance of a statistical model is judged by its capacity to accurately anticipate new analogues of the series. A model is only usable inside its training area, and new analogues must be reviewed to ensure that they belong to this domain. The applicability domain (AD) is determined using the Williams plot, which plots standardized residuals against the leverage values, h, of all compounds in the training domain [36,37]. The AD is determined using the Williams plot, which yields an area by taking into account a measure (s.d.) and the leverage threshold h*.. The h* is commonly specified at 3(k + 1)/n, where n is the number of compounds included in the training-set and k is the number of independent variables of the model under consideration. The value of β ranges between 2 and 3. The figure can then be used to visually discover the model's Y-outlier (response outlier) and X-outlier (structurally influential compound). When the leverage value of a compound is less than the threshold value h*, the prediction becomes trustworthy. In this situation, the calculated and observed activity values of the training-set chemicals agree. When h > h*, the forecast becomes unreliable.

RESULTS AND DISCUSSION

A total of 675 descriptors for compounds in Table 1 and 651 descriptors for compounds in Table 2 belonging to 3D-classes were initially used to correlate their insecticidal capabilities using the CP-MLR technique. Furthermore, the descriptors of the 0D- to 2D-classes have been attempted to associate the biological activity profiles of the substances under inquiry. However, these descriptors revealed poor models when compared to 3D-descriptors.

As a result, 3D-descriptors were only employed to obtain models in one-descriptor and two-descriptor increments. The two-descriptor models remained statistically significant in explaining the variance in observed activities of the congeners in two series. To examine prediction models, compounds, from Table 1 and Table 2, were used to measure their insecticidal activity against 2nd and 4th instar larvae in terms of scaled 3D-descriptors. The most acceptable descriptors used to create the final highest significant models of two series are included in respective Tables for convenience.

For the novel compounds (Table 1) which have shown insecticidal activities against 2nd instar larvae and 4th instar larvae of S. littoralis, only three highest significant models, for each of them, have been included in Table 4.

Table 4. Regression equations and statistical parameters for the 2nd and 4th instar larvae of Spodoptera littoralis

Eq. No.	Regression Equation (n = 12 and LC_{50} in ppm)	r	s	F(2, 9)[a]	AIC	LOF	FIT	Q^2_{LOO}	Q^2_{L5O}
2nd instar larvae									
1	$logLC_{50}$= –0.486(0.089)RDF140m –0.939(0.097)HATS2m+ 2.161	0.973	0.108	78.779	0.019	0.020	9.847	0.897	0.693
2	$logLC_{50}$ = 0.513(0.108)Mor09e –1.142(0.087)R2m + 1.820	0.975	0.103	85.967	0.018	0.018	10.746	0.910	0.897
3	$logLC_{50}$ = 0.577(0.109)Mor22m –1.014(0.077)R2m + 1.679	0.979	0.095	102.238	0.015	0.015	12.780	0.927	0.929

4th instar larvae									
4	$logLC_{50}$ = 0.301(0.050)Mor22m –0.312(0.041)R1m + 2.084	0.954	0.044	45.478	0.003	0.003	5.685	0.743	0.699
5	$logLC_{50}$= –0.399(0.038)RDF130m –0.156(0.035)HATS1p + 2.254	0.970	0.036	71.356	0.002	0.002	8.920	0.911	0.823
6	$logLC_{50}$ = 0.277(0.038)Mor22m –0.282(0.027)R2m + 2.073	0.974	0.033	83.656	0.002	0.002	10.457	0.856	0.866

[a]The critical F-value, significant at 99%, is $F_{2, 9}(0.01) = 8.022$

These models with requisite statistical parameters are given through Eqs. (1)-(3) and Eqs. (4)-(6). In all these Equations, the F-values remain significant at 99% level and the standard errors associated with regression coefficients (data within the parentheses) are significant at more than 95% level. The indices Q^2_{LOO} and Q^2_{L5O} (> 0.5) accounted for internal robustness of the derived models. In a model, the signs of the regression coefficients revealed the direction of influence of explanatory variables, the negative regression coefficient associated to a descriptor will improve the insecticidal activity of a compound while the positive regression coefficient will cause detrimental effect to it. A total number of 20 and 11 descriptors have participated in the developed models for data-set in Table 1 and Table 2 respectively.

The further discussion is, however, confined only to one highest significant model Eq. (3) and Eq. (6) of Table 4 for insecticidal activities, respectively, against 2nd instar and 4th instar larvae of S. littoralis. The descriptors, namely, the Mor22m and R2m have participated in both Eq. (3) and Eq. (6) but differ in their statistical parameters. The descriptor Mor22m is from the 3D-MoRSE class, denoting the 3D-Morse signal 22 and R2m is from the GETAWAY class, representing the R autocorrelation of lag 2. Each of these descriptors is weighted by atomic masses.The resulting descriptors are, therefore, highly influential to address the insecticidal actions of the compounds. The sign of regression coefficient associated to these descriptors have indicated the direction of their influence on the $logLC_{50}$(ppm) for 2nd instar and 4th instar larvae of S. littoralis. For a compound to be more active, its lethal concentration LC_{50} (or $logLC_{50}$) should be low. In other words, the insecticidal activity of a compound is inversely proportional to its lethal concentration. In Eq. (3) and Eq. (6), the regression coefficient of the descriptor Mor22m is positive, indicates that its more positive value will result into higher $logLC_{50}$ value (or lower activity). On the other hand, the regression coefficient of descriptor R2m is negative which suggests that its more positive value will results into lower $logLC_{50}$ value (or higher activity). Thus, to improve the effectiveness a compound against 2nd instar and 4th instar larvae, it is desirable to have lower (positive) value of the Mor22m and higher (positive) value of R2m.

The molecular bulk (mass) appeared to play important role for an insecticide to be active against 2nd and 4th instar larvae of S. littoralis.

In Table 5, the highest significant regression Eq. (9) and Eq. (12) with their statistical parameters stands, respectively, for the 2nd and 4th instar larvae of S. frugiperda.

Table 5. Regression equations and statistical parameters for the 2nd and 4th instar larvae of Spodoptera frugiperda

Eq. No.	Regression Equation (n = 11 and LC_{50} in ppm)	r	s	F(2, 8)[a]	AIC	LOF	FIT	Q^2_{LOO}	Q^2_{L5O}
2nd instar larvae									
7	$logLC_{50}$ = –0.686(0.069)RDF040v +0.473(0.073)RDF110e + 1.414	0.980	0.063	97.261	0.007	0.007	12.968	0.931	0.723
8	$logLC_{50}$ = 0.398(0.060)RDF115v –0.878(0.060)RDF040p + 1.615	0.982	0.060	110.252	0.006	0.006	14.700	0.929	0.887
9	$logLC_{50}$ = –0.882(0.060)RDF040p +0.394(0.059)RDF115p + 1.620	0.982	0.060	111.262	0.006	0.006	14.835	0.933	0.928
4th instar larvae									
10	$logLC_{50}$ = 0.222(0.030)RDF110u +0.114(0.027)R5e+ + 1.767	0.939	0.028	29.789	0.001	0.001	3.972	0.836	0.800
11	$logLC_{50}$ = –0.171(0.029)RDF055e +0.210(0.028)RDF110e + 1.889	0.953	0.025	40.038	0.001	0.001	5.338	0.845	0.864

12	$logLC_{50}$ = –0.173(0.027)RDF055u +0.218(0.026)RDF110e + 1.889	0.960	0.023	46.667	0.001	0.001	6.224	0.875	0.830

[a]The critical F-value, significant at 99%, is $F_{2,8}(0.01) = 8.649$

In Eq. (9), the participated descriptors, RDF040p and RDF115p are from RDF-class account, in that order, for radial distribution functions-4.0 and -11.5. Both these functions are weighted by atomic polarizabilities. In Eq. (12), the important emerged descriptors, the RDF055u and the RDF110e are also from RDF-class. The former descriptor is unweighted radial distribution function-5.5 while the later descriptor is atomic Sanderson electronegativities weighted radial distribution function-11.0. From the Eq. (9) it became apparent that the lower (positive) value of RDF115p and the higher (positive) value of RDF040p are helpful in augmenting the activity profile of a compound against 2nd instar larvae. The polarizability of an insecticidal molecule has crucial role against 2nd instar larvae of S. frugiperda. Likewise, the lower value of the RDF110e and the higher value of the RDF055u of Eq. (12) are beneficial in improving the activity of a compound against 4nd instar larvae. The electronegativity weighted descriptor has imparted significant contribution against 4nd instar larvae of S. frugiperda.

The average regression coefficients and incidences of all 31 descriptors are included in Table 6.

Table 6. Identified descriptors[a] with average participation of regression coefficient and incidence[b], in modeling of the insecticidal activity

Series I : Semicarbazide, thiosemicarbazide, urea and thiourea derivatives				**Series II : Chalcone derivatives**			
Larvae of S. littoralis				**Larvae of S. frugiperda**			
2nd instar		**4th instar**		**2nd instar**		**4th instar**	
Descriptor	**AvgP (incidence)**	**Descriptor**	**AvgP (incidence)**	**Descriptor**	**AvgP (incidence)**	**Descriptor**	**AvgP (incidence)**
DISPp	-1.237 (1)	RDF130m	-0.409(3)	RDF040v	-0.686 (1)	RDF055u	-0.173 (1)
RDF140m	-0.486 (1)	Mor15m	0.133 (1)	RDF115v	0.399 (1)	RDF110u	0.190 (2)
Mor09u	0.497 (1)	Mor22m	0.269 (3)	RDF110e	0.473 (1)	RDF055e	-0.171 (1)
Mor22m	0.577 (1)	H6m	-0.302(2)	RDF040p	-0.880 (2)	RDF110e	0.214 (2)
Mor09e	0.513 (1)	HATS1m	-0.245 (1)	RDF115p	0.394 (1)	Mor14u	-0.143 (1)
Mor14e	0.746 (1)	HATS2m	-0.241 (1)			R5e+	0.114 (1)
HATS2m	-1.006(2)	HATS1v	-0.118 (1)				
H6e	-0.555 (1)	HATSe	-0.337 (1)				
R2m	-1.091 (3)	HATS1p	-0.156 (1)				
		R1m	-0.312 (1)				
		R2m	-0.282 (1)				

[a]The descriptors were identified from the models, surfaced from CP-MLR protocol with training set of 12 (first series) and 11 (second series) compounds for insecticidal activity. [b]The average regression coefficient of descriptor relating to the most significant models and total number of its incidence. The arithmetic sign of the coefficient represents the actual sign of the regression coefficient in the models.

The squared correlation coefficient r^2, obtained in conjunction with Eq. (3) and Eq. (6), has explained 96% and 95% of variance in observed activity profiles, respectively, against 2nd instar and 4th instar larvae of S. littoralis. Also, the explained variance, in terms of r^2 value of Eq. (9) and Eq. (12), were 96% and 92%, respectively, for the 2nd and 4th instar larvae of S. frugiperda. The other statistical parameters of these four Equations tune to the most significant models.

Next, Eq. (3) and Eq. (6) for all compounds of first series and Eq. (9) and Eq. (12) for compounds of second series were used to calculate the insecticidal activities and the same are included, respectively, in Table 1 and Table 2 for the sake of comparison with observed ones. Moreover, a close agreement between observed and calculated $logLC_{50}$s for the compounds of these Tables, are apparent through the graphical representations, shown in Figure 2.

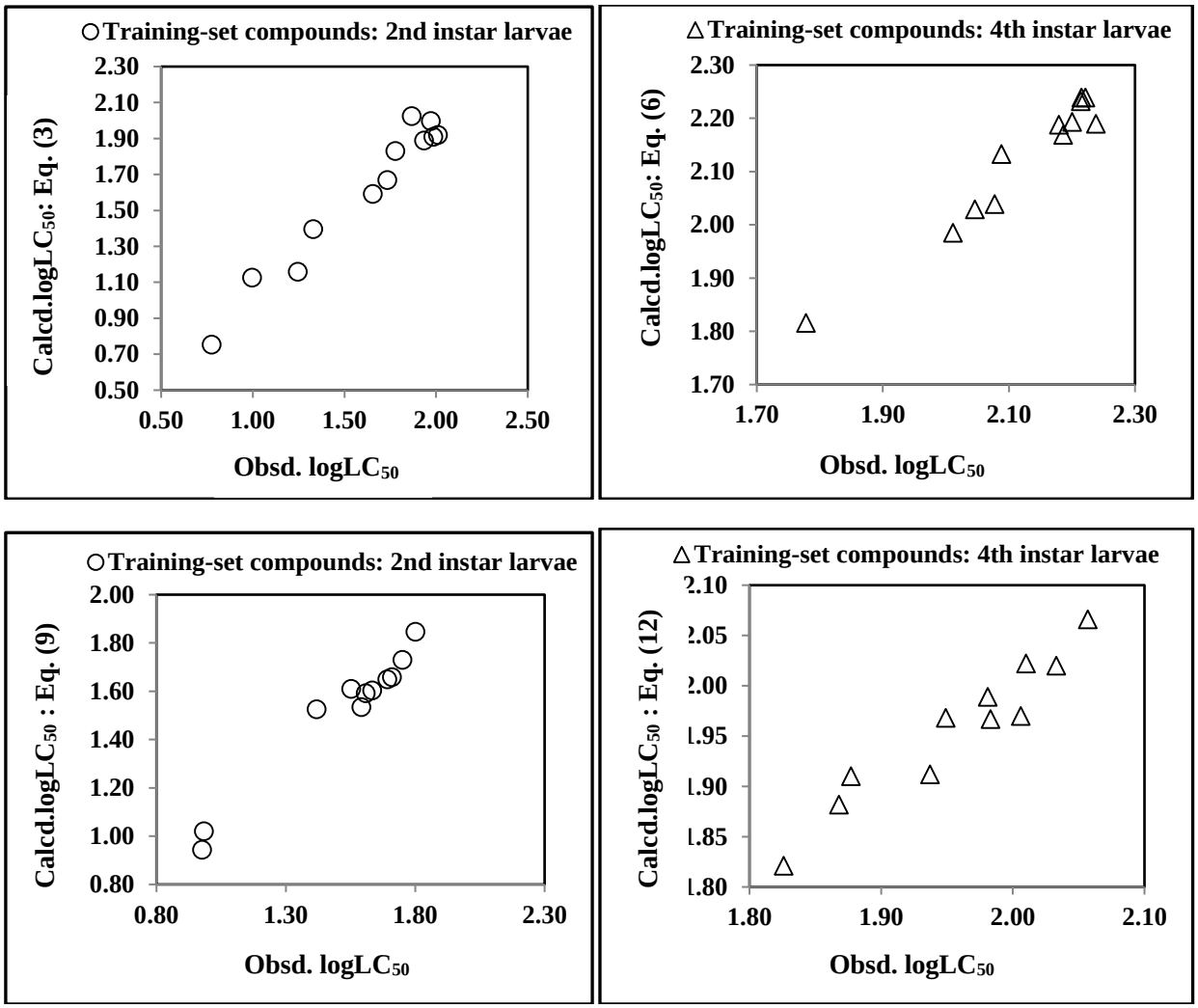


Figure 2. Plot between observed and calculated logLC50 values using Eq. (3) for 2nd larvae, Eq. (6) for 4th larvae of S. littoralis, and Eq. (9) for 2nd larvae, Eq. (12) for 4th larvae of S. frugiperda

The applicability domains (ADs) were analyzed for the models based on data-set in Table 1 and Table 2. The same are shown through the Williams plots, separately for first series and second series, in which standardized residuals were plotted against leverage (h_i) values. For this purpose, the most influential descriptors, Mor22m and R2m, were considered for compounds of first series relating to the activity profiles for 2nd instar and 4th instar larvae of S. littoralis (Eq. 3 and Eq. 6, Table 4). Similarly the descriptors, RDF040p and RDF115p, of Eq. (9) and the descriptors, RDF055u and RDF110e, of Eq. (12), from Table 5 were used to obtain standardized residuals and leverage values for the larvae of S. frugiperda. The standardized residuals and leverage values were further used to establish the ADs individually for 2nd instar and 4th instar larvae of S. littoralis and S. frugiperda. In each case, the limits of standardized residuals (Y-outliers) were considered as $\pm\beta \times$s.d. and the leverage threshold as $h^* (= 3(k + 1)/n)$. The value of β was taken equal to 2 for all four models. For convenience, the graphical representations, for the models obtained in their influential descriptors, depicting the data-set compounds are given in Figure 3.

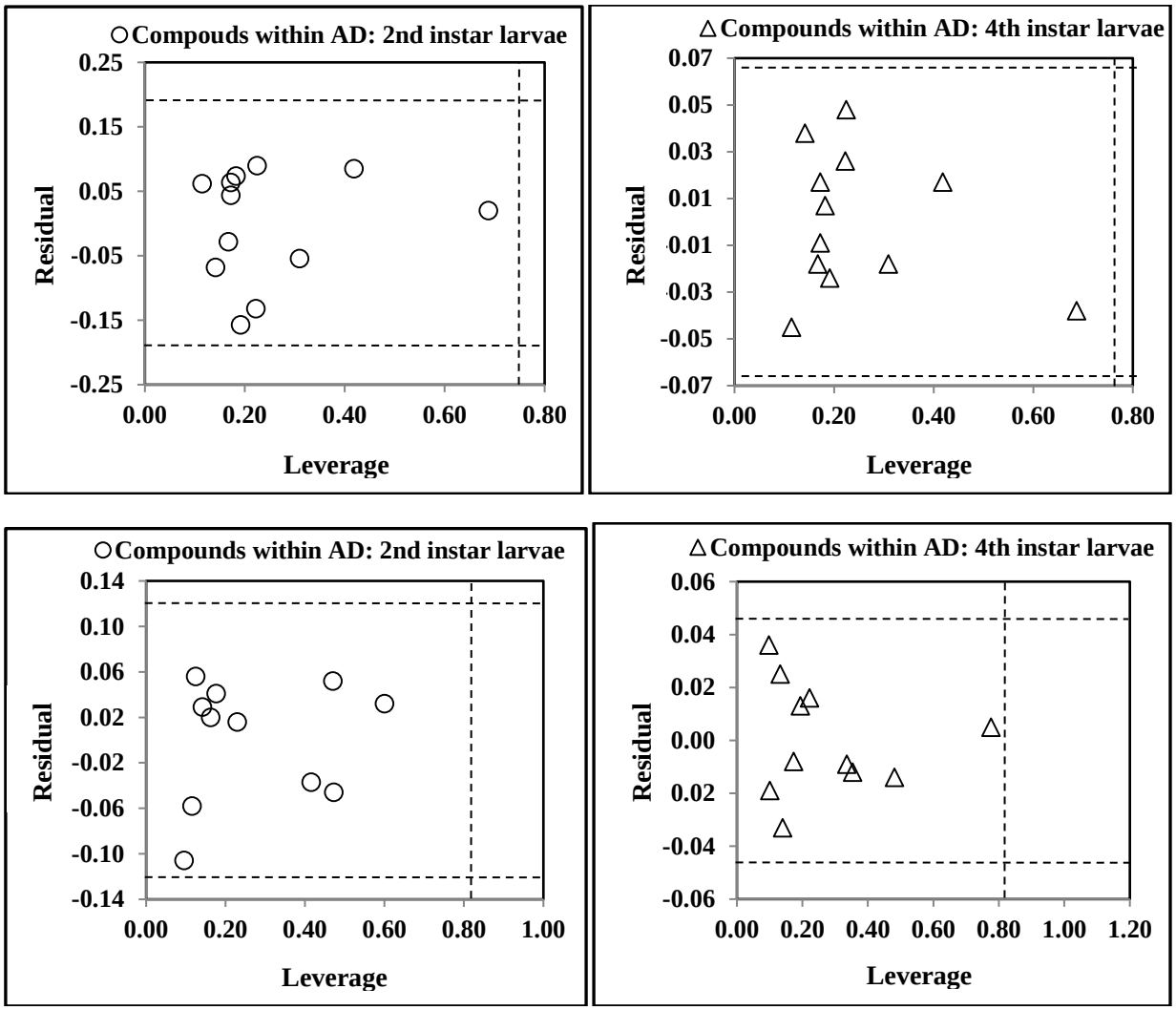


Figure 3. Plot between leverage and residual values for 2nd and 4th instar larvae of S. littoralis (h* = 0.75) and 2nd and 4th instar larvae of S. frugiperda (h* = 0.82)

For given data-set compounds, the suggested models match the most significant parameters with good fitting power. Further, all of the compounds were remained within the ADs which denote that the models under consideration were able to evaluate the data-set compounds, correctly.

CONCLUSION

Insecticidal activity profiles of two novel series of chemicals were quantified using chemometric 3D-descriptors. The first series consists of semicarbazide, thiosemicarbazide, urea, and thiourea derivatives that have insecticidal activity against Spodoptera littoralis larvae in the 2nd and 4th instars. The second series, consisting of chalcone analogues, has exhibited insecticidal activity for Spodoptera frugiperda larvae in the 2nd and 4th instars. The quantitative structure-activity relationship (QSAR) models built statistically gave rationale to explain the insecticidal effects of both series of chemicals. For the first series, the most influential descriptors were the 3D-Morse signal 22/weighted by atomic masses (Mor22m) and R autocorrelation of lag 2/weighted by atomic masses (R2m). The radial distribution functions-4.0 and -11.5, which are weighted by atomic polarizabilities (RDF040p and RDF115p), the unweighted radial distribution function-5.5 (RDF055u), and the atomic Sanderson electronegativities weighted radial distribution function-11.0 (RDF110e) were filtered out for the second series.The AD studies demonstrated that the proposed models have appropriate prediction because all of the chemicals from both series were present within their domains (square areas). As a result, the models under evaluation successfully predicted the insecticidal activity of all compounds in both series separately. The guidelines mentioned in the discussion may be useful in exploring new potential analogues of each series.

CONFLICT OF INTEREST

Authors declare that there is no conflict of interest between them.

ACKNOWLEDGEMENT

Help provided by the Institution and the colleagues is thankfully acknowledged.

REFERECES:

[1] Gad, M.A., Alqurashi, E.A., Alsenani, N.I., Abd El Latif, F.M., Aref, S.A., Ahmed, N.A., Abdelhamid, A.A., El-Saghier, A.M.M. (**2023**) Insecticidal activity, and SAR studies of semicarbazide, thiosemicarbazide, urea and thiourea derivatives against spodoptera littoralis (Boisd.). J. Umm. Al-Qura Univ. Appl. Sci. [https://doi.org/10.1007/s43994-023-00037-6]

[2] Ali, A.M., Salah, H., Gad, M.A., Youssef, M.A.M., Elkanzi, N.A.A. (**2022**) Design, synthesis, and SAR studies of some novel chalcone derivatives for potential insecticidal bioefficacy screening on spodoptera frugiperda (Lepidoptera: Noctuidae). ACS Omega, 7(44), 40091-40095. [https://doi.org/10.1021/acsomega.2c04814]

[3] Al-Shannaf, H.M., Desuky, W.M., Abd El-Halim,) S.M. (**2006**) Effect of some compounds on cotton leafworm, spodoptera littoralis (Boisd.) and their predators. Egypt. J. Appl. Sci. 21(4B), 646-660.

[4] Nasr, H.M., Badawy, M.E.I., Rabea, E.I. (**2010**) Toxicity and biochemical study of two insect growth regulators, buprofezin and pyriproxyfen on cotton leafworm spodoptera littoralis. Pestic. Biochem. Physiol. 98(2), 198-205. [https://doi:10.1016/j.pestbp.2010.06.007]

[5] Abdelhamid, A.A., Elwassimy, M.M., Aref, S.A., Gad, M.A. (**2019**) Chemical design and bioefficacy screening of new insect growth regulators as potential insecticidal agents against spodoptera littoralis (Boisd.). Biotechnol. Rep. 24, 394-401. [https://doi: 10.1016/j.btre.2019.e00394]

[6] Aiming, S., Andrew, P., Weiqiang, Z., Murray, E.E., Joshua, D., Li-Ting, C., Jeong-Joong, Y., Radchenko, E.V., Palyulin, V.A., Compans, R.W. (**2006**) Nonpeptide inhibitors of measles virus entry. J. Med. Chem. 49(17), 5080-5092. [https://doi.org/10.1021/jm0602559]

[7] Dan, H., Menglei, W., Siyu, Z., Yisong, S., Honglian, Z., Cheng, X., Cheng, L.,Yuanyan, L. (**2008**) Synthesis of novel 4, 6-disubstituted quinazoline derivatives, their anti-inflammatory and anti-cancer activity (cytotoxic) against U937 leukemia cell lines. Eur. J. Med. Chem. 43(4), 846-852. [https://doi:10.1016/j.ejmech.2007.06.010]

[8] Deshmukh, S., Pavithra, H.B., Kalleshwaraswamy, C.M., Shivanna, B.K., Maruthi, M.S., Mota-Sanchez, D. (**2020**) Field efficacy of insecticides for management of invasive fall armyworm, spodoptera frugiperda (Smith, J.E.) (Lepidoptera: Noctuidae) on Maize in India, Fla. Entomol. 103(2), 221-227. [https://doi.org/10.1653/024.103.0211]

[9] Maruthadurai, R., Ramesh, R. (**2020**) Occurrence, damage pattern and biology of fall armywork, spodoptera frugiperda (Smith, J.E.) (Lepidoptera: Noctuidae) on fodder crop and green amaranth in Goa, India. Phytoparasitica 48(10), 15-23. [https://doi:10.1007/s12600-019-00771-w]

[10] Assefa, F., Ayalew, D. (**2019**) Status and control measure of fall armyworm (spodoptera frugiperda) infections in maize fields in Ethiopia: A review. Cogent Food Agric. 5(1), 1641902-1641917. [https://doi.org/10.1080/23311932.2019.1641902]

[11] Storer, N.P., Babcock, J.M., Schlenz, M., Meade, T., Thompson, G.D., Bing, J.W., Huckaba, R.M. (**2010**) Discovery and Characterization of Field Resistance to Bt maize: spodoptera frugiperda (Lepidoptera: Noctuidae) in Puerto rico. J. Econ. Entomol. 103(4), 1031-1038. [https://doi.org/10.1603/EC10040]

[12] Romanelli, G.P., Virla, E.G., Duchowicz, P.R., Gaddi, A.L., Ruiz, D.M., Bennardi, D.O., del valleOrtz, E.V., Autino, J.C. (**2010**) Sustainable synthesis of flavonoid derivatives. QSAR study of insecticidal activity against fall armyworm, spodoptera frugiperda (Lep.: Noctuidae). J. Agric. Food Chem. 58(10), 6290-6295. [https://doi.org/10.1021/jf100073j]

[13] Nielsen, S.F., Christensen, S.B., Cruciani, D., Kharazmi, A., Liljefors, T. (**1998**) Antileishmanial chalcones: Statistical design, synthesis, and three-dimensional quantitative structure-activity relationship analysis. J. Med. Chem. 41(24), 4819-4832. [https://doi: 10.1021/jm980410m]

[14] Anto, R.J., Sukumaran, K., Kuttan, G., Rao, M.N.A., Subbaraju, V., Kuttan, R. (**1975**) Anticancer and antioxidant activity of synthetic chalcones and related compounds. Cancer Lett. 97(1), 33-37. [https://doi.org/10.1016/0304-3835(95)03945-S]

[15] Konieczny, M.T., Horowska, B.,, Kunikowski, A., Konopa, J., Wierzba, K., Yamada, Y., Asao, T. (**2001**) Synthesis of polyhydroxylated derivatives of phenyl vinyl sulfone as structural analogs of chalcones. Synthesis 9, 1363-1367. [https://doi: 10.1055/s-2001-15223]

[16] Lin, Y.M., Zhou, Y., Flavin, M.T., Zhou, L.M., Nie, W., Chen, F.C. (**2002**) Chalcones and flavonoids as anti-tuberculosis agents. Bioorg. Med. Chem. 10(8), 2795-2802. [https://doi: 10.1016/s0968-0896(02)00094-9]

[17] El-Gaby, M.S.A., Ammar, Y.A., Drar, A.M., Gad, M.A. (**2022**) Inseciticidal bioefficacy screening of some chalcone and acetophenone hydazone derivatives on spodopetra frugiperda (Leidopetra: Noctuidae). Curr. Chem. Lett. 11(3), 263-268. [http://doi: 10.5267/j.ccl.2022.4.003]

[18] Böger, M., Dur, D., Gsell, L., Hall, R.G., Karrer, F., Kristiansen, O., Malenfisch, P., Pascual, A., Rindlisbacher, A. (**2001**) Synthesis and structure-activIty relationships of benzophenone hydrazone derivative with insecticidal activity. Pest Manage. Sci. 57(2), 191-202. [https://doi.org/10.1002/1526-4998(200102)57:2<191::AID-PS275>3.0.CO,2-O]

[19] ChemDraw ultra 6.0 and Chem3D ultra. Cambridge Soft Corporation, Cambridge, USA.

[20] Todeschini, Y., Consonni, V., Mauri, A., Pavan, M. DRAGON software (version 3.0-**2003**), Milano, Italy.

[21] Golbraikh, A., Tropsha, A. (**2002**) Beware of q2! J. Mol. Graph Model. 20(4), 269-276. [https://doi:10.1016/s1093-3263(01)00123-1]

[22] Prabhakar, Y.S. (**2003**) A combinatorial approach to the variable selection in multiple linear regression: Analysis of Selwood et al. data set - A case study. QSAR Comb. Sci. 22(6), 583-595. [https://doi:10.1002/qsar.200330814]

[23] Sharma, S., Sharma, B.K., Sharma, S.K., Singh, P., Prabhakar, Y.S. (**2009**) Topological descriptors in modeling the agonistic activity of human A_3 adenosine receptor ligands: The derivatives of 2-chloro-N^6-substituted-4′-thioadenosine-5′-uronamide. Eur. J. Med. Chem. 44(4), 1377-1382. [https://doi:10.1016/j.ejmech.2008.09.022]

[24] Sharma, B.K., Pilania, P., Singh, P., Prabhakar, Y.S. (**2010**) Combinatorial protocol in multiple linear regression/partial least-squares directed rationale for the caspase-3 inhibition activity of isoquinoline-1,3,4-trione derivatives. SAR QSAR Environ. Res. 21(1-2), 169-185. [https://doi.org/10.1080/10629360903570545]

[25] Sharma, B.K., Singh, P., Sarbhai, K., Prabhakar, Y.S. (**2010**) A quantitative structure-activity relationship study on serotonin (5-HT6) receptor ligands: Indolyl and piperidinylsulphonamides. SAR QSAR Environ. Res. 21(3-4), 369-388. [https://doi.org/10.1080/10629361003773997]

[26] Sharma, B.K., Pilania, P., Sarbhai, K., Singh, P., Prabhakar, Y.S. (**2010**) Chemometric descriptors in modeling the carbonic anhydrase inhibition activity of sulfonamide and sulfamate derivatives. Mol. Divers.14(2), 371-384. [https://doi:10.1007/s11030-009-9181-5]

[27] Sharma, B.K., Singh, P., Shekhawat, M., Sarbhai, K., Prabhakar, Y.S. (**2011**) Modelling of serotonin reuptake inhibitory and histamine H_3 antagonistic activity of piperazine and diazepane amides: QSAR rationales for co-optimization of the activity profiles. SAR QSAR Environ. Res. 22(3-4), 365-383. [https://doi.org/10.1080/1062936X.2011.569895]

[28] Singh, P. (**2013**) Molecular Descriptors in modelling the tumour necrosis factor-α converting enzyme inhibition activity of novel tartrate-based analogues. Indian J. Pharm. Sci. 75(1), 36-44. [https://doi:10.4103/0250-474X.113539]

[29] So, S.-S., Karplus, M. (**1997**) Three-dimensional quantitative structure-activity relationships from molecular similarity matrices and genetic neural networks. 1. Method and validations. J. Med. Chem. 40(26), 4347-4359. [https://doi.org/10.1021/jm970487v]

[30] Prabhakar, Y.S., Solomon, V.R., Rawal, R.K., Gupta, M.K., Katti, S.B. (**2004**) CP-MLR/PLS directed structure-activity modeling of the HIV-1 RT inhibitory activity of 2,3-diaryl-1,3-thiazolidin-4-ones. QSAR Combn. Sci. 23(4), 234-244.[https://doi.org/10.1002/qsar.200330854]

[31] Akaike, H. (**1973**) Information theory and an extension of the minimum likelihood principle. In: Second international symposium on information theory, Petrov, B.N., Csaki, F., Eds., Akademiai Kiado, Budapest, pp. 267-281.

[32] Akaike H. (**1974**) A new look at the statistical identification model. IEEE Trans. Automat. Control. AC-19(6), 716-723. [https://doi: 10.1109/TAC.1974.1100705]

[33] Kubinyi, H. Variable selection in QSAR studies. I. (**1994**) An evolutionary algorithm. Quant. Struct.-Act. Relat. 13(3), 285-294. [https://doi.org/10.1002/qsar.19940130306]

[34] Kubinyi, H. (**1994**) Variable selection in QSAR studies. II. A highly efficient combination of systematic search and evolution. Quant. Struct.-Act. Relat. 13(4), 393-401. [https://doi.org/10.1002/qsar.19940130403]

[35] Friedman J. (**1990**) In: Technical report no 102. Laboratory for computational statistics, Stanford, Stanford University.

[36] Gramatica, P. (**2007**) Principles of QSAR models validation: Internal and external. QSAR Comb. Sci. 26(5), 694-701. [https://doi:10.1002/qsar.200610151]

[37] Eriksson, L., Jaworska, J, Worth A.P., Cronin, M.T., McDowell, R.M., Gramatica, P. (2003) Methods for reliability and uncertainty assessment and for applicability evaluations of classification and regression-based QSARs. Environ. Health Perspect. 111(10), 1361-1375. [https://doi.org/10.1289/ehp.5758]