

Aryloxyacetic acid derivatives as inhibitors of the HPPD enzyme: A quantitative structure-activity relationship analysis

Prithvi Singh¹ and Nidhi Shekhawat^{2*}

¹Department of Chemistry, Government Science College, Sikar-332 001, India

²Department of Zoology, Vedanta Post-Graduate Girls' College, Reengus-332 404, Rajasthan, India

Received 13 March 2024; revised 07 April 2025

The present work involved a quantitative analysis of the inhibitory effects of aryloxyacetic acid derivatives against *Arabidopsis thaliana* 4-hydroxyphenylpyruvate dioxygenase (*AtHPPD*). The study is an important *in silico* approach to design potential inhibitors of *AtHPPD* on rational basis. Also, it is helpful to optimize the existing leads to improve their biological activities and to predict the biological activities of untested and sometimes yet unavailable compounds. Moreover, the study could assist to decipher the possible mechanism of action on to receptor site(s) at the molecular level. The rational approach minimizes the time and cost needed in the drug designing process in comparison to traditional drug discovery methods.

A total of 51 compounds were examined for the inhibition actions in relation to their molecular structures. The investigation revealed 15 prominent regression models, sharing 20 chemometric descriptors. Each of these models was validated for its internal and external robustness, while the randomization test was performed to avoid any coincidental correlation. The importance of these descriptors was further elaborated through the partial least squares analysis. The highest significant model was employed to calculate/predict the inhibition activities of all analogs of the series. The applicability domain analysis suggested that the model, under consideration, has satisfactory predictability. The guidelines given in the discussion may be helpful in exploring new potential analogs of the series.

Keywords: 4-Hydroxyphenylpyruvate dioxygenase, Aryloxyacetic acid derivatives, Enzyme inhibitors, Quantitative structure-activity relationship

The enzyme 4-hydroxyphenylpyruvate dioxygenase (HPPD) is a non-heme enzyme which is found in a variety of organisms¹. The amino acid sequences of HPPD in plants and mammals differ significantly^{2,3}, which affects the binding stability between the enzyme and inhibitors. Thus, inhibition activities can vary among species, leading to the potential development of selective and safe inhibitors⁴. Within plants, HPPD inhibitors act as competitive controllers of 4-hydroxyphenylpyruvic acid (HPPA), preventing it from chelating to Fe^{II}. When the transformation of HPPA to homogentisic acid (HGA) is hindered by an HPPD inhibitor, plastoquinone production is inhibited and phytoene accumulates^{5,6}. This can result in bleaching symptoms followed by necrosis and death^{7,8}, after the plant has been exposed to sunlight. Despite this, HPPD inhibiting herbicides are still useful due to their low toxicities, broad-spectrum weed control and safety toward crops and the

environment⁹⁻¹¹. The main objective of agriculture research is to explore effective HPPD inhibitors (herbicides) that are safe for crops and the environment while controlling resistant weeds. Herbicides are usually classified into three categories: triketones, pyrazoles isoxazoles^{5,11,12}.

Recent research has shown that inhibitors of *Arabidopsis thaliana* HPPD (*AtHPPD*) interact with the active center metal through 1,3-dicarbonyl bidentate chelation and with the Phe360 and Phe403 residues of the active site through favorable sandwich π - π stacking interactions between aromatic rings. It implies that compounds with 1,3-dicarbonyl fragments and aromatic rings are crucial for effective HPPD inhibition and interact with adjacent residues in *AtHPPD*¹²⁻¹⁵. In view of this, a new series, containing 51 aryloxyacetic acid derivatives with their *in vitro* inhibition activities against *AtHPPD* have recently been reported¹⁶.

The aim of present communication is to explore the mathematical relationships between structural features and inhibition activities of reported compounds. The

*Correspondence:
E-mail: nidhi14sikar@gmail.com

structural features are expressed in terms of molecular descriptors, while their inhibition activity profiles are analyzed in relation to linear free energy. The strategy is commonly known as quantitative structure-activity relationship (QSAR) analysis, in which correlation equations/models are developed employing statistical methods. The interpolative and extrapolative capabilities of the highest significant model are made explicit in the discussion.

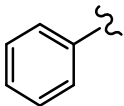
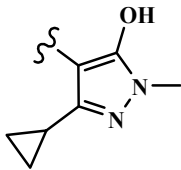
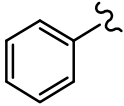
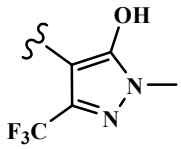
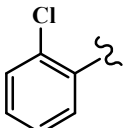
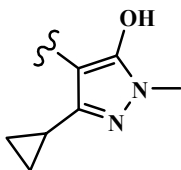
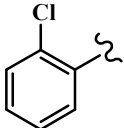
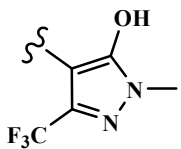
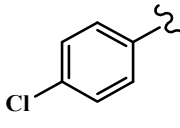
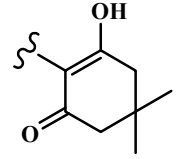
Materials and Methods

We have taken data from literature¹⁶ on 51 different aryloxyacetic acid derivatives and their inhibition activity regarding *Ar*HPPD. The inhibition activity is evaluated through binding constants K_i s, which are then expressed as pK_i s ($= -\log K_i$ s) on a

molar basis. Table 1 provides information on substituent variations at R and R₁ of the parent skeleton. The experimental activity profiles, pK_i s being the dependent variable, are listed in this Table.

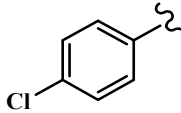
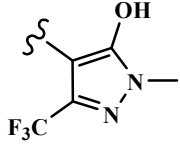
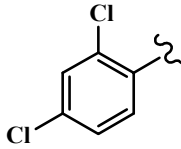
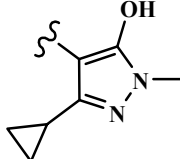
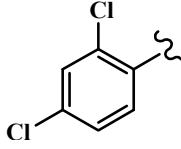
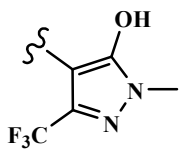
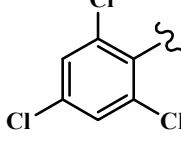
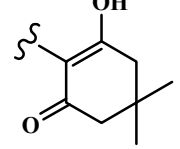
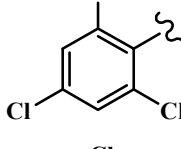
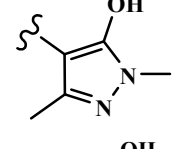
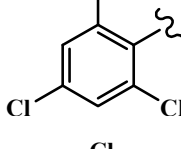
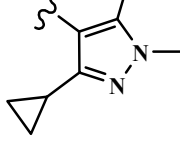
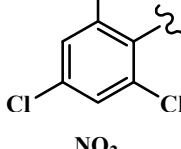
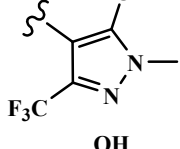
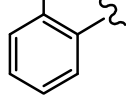
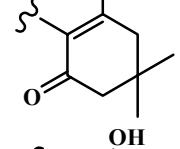
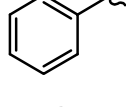
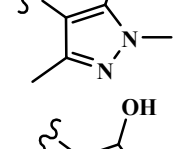
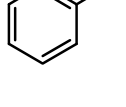
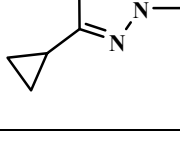
Initially, a dataset consisting of the most relevant molecular descriptors of the 51 analogs was employed to establish a correlation between their respective activity profiles. Subsequently, the most significant model equation was applied to calculate the pK_i values of these compounds. It is imperative that the calculated pK_i values must concur with the observed values. If any congener's calculated pK_i value deviates significantly, it is necessary to provide an explanation. Additionally, the application of an applicability domain analysis can facilitate the identification of the most influential and outlier compounds in the series.

Table 1 — Molecular descriptors, observed and modeled herbicidal activities of aryloxyacetic acid derivatives

Compd.	R	R ₁	Scaled descriptors				Obsd. ^a	pK _i (M)			
			TIE	CIC1	GGI6	JGI5		Calcd.			
								Eq. (1)	Eq. (2)	Eq. (7)	PLS
1			0.058	0.795	0.000	0.632	5.82	5.73	5.83	5.91	6.04
2			0.000	0.468	0.136	0.105	5.89	6.01	6.00	5.92	5.86
3			0.085	0.632	0.000	0.684	6.44	6.10	6.21	6.27	6.46
4 ^b			0.131	0.296	0.136	0.211	6.23	6.56	6.57	6.49	6.46
5			0.212	0.705	0.339	0.632	6.00	6.03	5.97	5.94	5.83

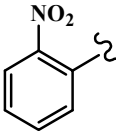
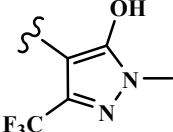
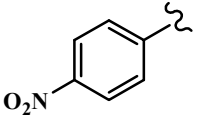
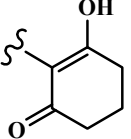
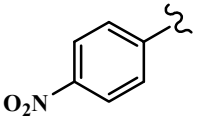
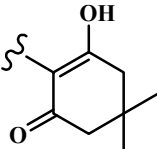
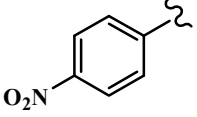
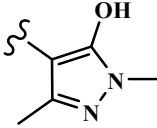
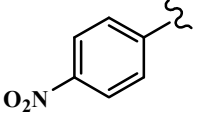
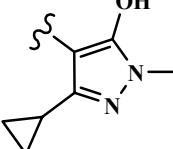
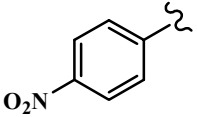
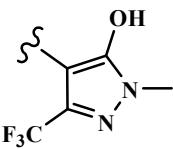
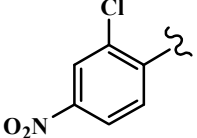
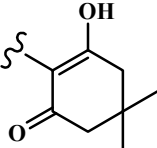
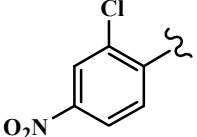
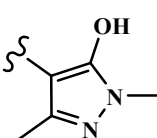
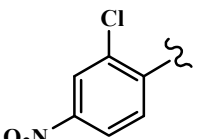
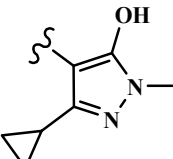
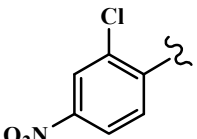
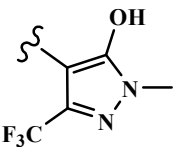
(Contd.)

Table 1 — Molecular descriptors, observed and modeled herbicidal activities of aryloxyacetic acid derivatives (Contd.)

6			0.049	0.296	0.217	0.211	6.03	6.43	6.40	6.28	6.20
7			0.133	0.546	0.081	0.737	6.44	6.33	6.42	6.44	6.56
8			0.350	0.214	0.217	0.263	6.66	7.03	7.01	6.91	6.77
9			0.485	0.584	0.339	0.737	6.51	6.69	6.65	6.63	6.42
10			0.296	0.348	0.134	0.263	6.62	6.72	6.74	6.68	6.77
11			0.272	0.496	0.081	0.789	7.09	6.64	6.74	6.77	6.86
12			0.705	0.172	0.217	0.368	7.96	7.65	7.65	7.58	7.31
13 ^b			0.200	0.701	0.425	0.737	6.35	6.05	5.96	5.91	6.05
14			0.693	0.513	0.220	0.316	6.77	6.96	6.94	6.92	6.93
15			0.223	0.653	0.167	0.842	6.57	6.28	6.33	6.37	6.62

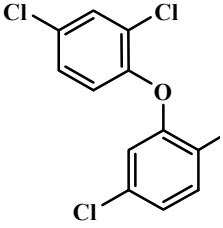
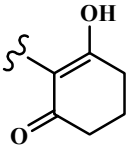
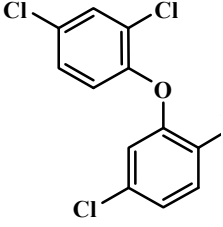
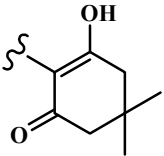
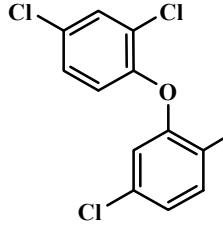
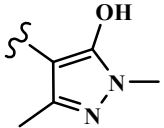
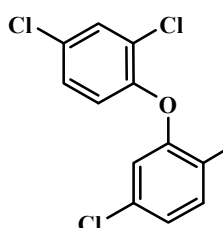
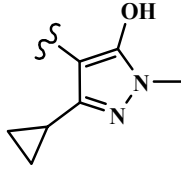
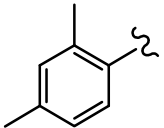
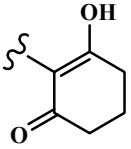
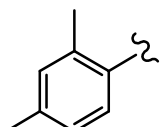
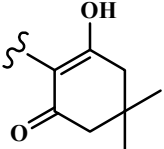
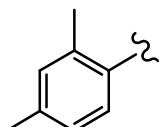
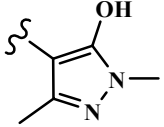
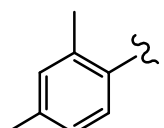
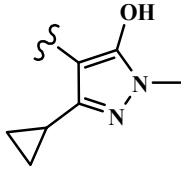
(Contd.)

Table 1 — Molecular descriptors, observed and modeled herbicidal activities of aryloxyacetic acid derivatives (*Contd.*)

16			0.033	0.331	0.303	0.368	6.36	6.39	6.33	6.20	6.32
17			0.264	0.558	0.256	0.526	6.64	6.36	6.34	6.30	6.35
18			0.204	0.701	0.423	0.895	6.03	6.12	6.05	6.02	5.99
19			0.140	0.513	0.217	0.474	6.20	6.26	6.25	6.20	6.28
20			0.163	0.653	0.163	0.947	6.30	6.24	6.31	6.35	6.46
21			0.181	0.331	0.299	0.526	5.82	6.67	6.63	6.53	6.36
22			0.154	0.563	0.423	0.947	5.99	6.34	6.28	6.22	6.23
23			0.558	0.352	0.217	0.526	7.68	7.17	7.18	7.14	7.08
24 ^b			0.408	0.505	0.163	1.000	6.59	6.89	6.97	7.01	7.04
25			0.083	0.178	0.299	0.579	6.68	6.84	6.81	6.69	6.64

(Contd.)

Table 1 — Molecular descriptors, observed and modeled herbicidal activities of aryloxyacetic acid derivatives (*Contd.*)

26			0.505	0.754	0.833	0.526	5.72	6.19	5.88	5.70	5.69
27			0.452	0.857	1.000	0.842	5.66	6.02	5.65	5.47	5.41
28			0.335	0.715	0.795	0.474	7.49 ^e	6.01	---	---	---
29 ^b			0.338	0.810	0.742	0.895	5.89	6.03	5.80	5.69	5.68
30			0.355	0.822	0.173	0.211	5.66	5.87	5.84	5.85	5.77
31			0.298	1.000	0.339	0.684	5.51	5.62	5.55	5.58	5.32
32			0.071	0.810	0.134	0.211	5.70	5.50	5.49	5.49	5.46
33			0.108	0.872	0.081	0.737	5.55	5.68	5.75	5.83	5.87

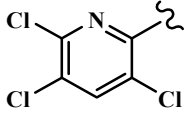
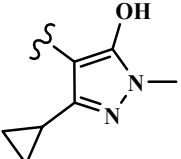
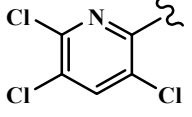
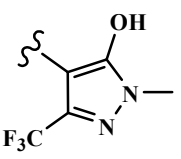
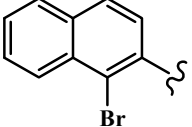
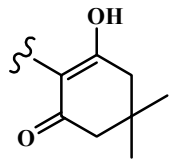
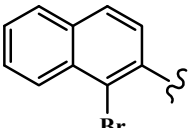
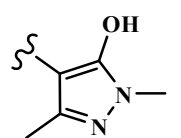
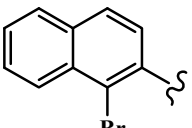
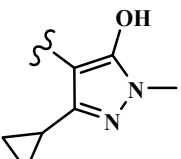
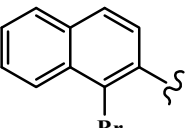
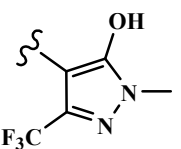
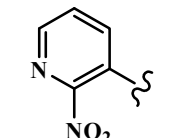
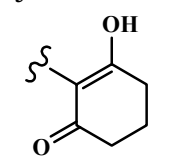
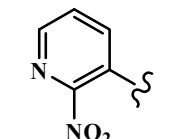
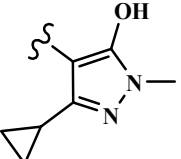
(Contd.)

Table 1 — Molecular descriptors, observed and modeled herbicidal activities of aryloxyacetic acid derivatives (*Contd.*)

34			0.059	0.571	0.217	0.263	5.67	5.94	5.91	5.84	5.70
35			0.327	0.838	0.256	0.211	5.48	5.78	5.71	5.69	5.61
36			0.271	0.985	0.423	0.632	5.47	5.56	5.45	5.44	5.29
37			0.087	0.760	0.217	0.158	5.57	5.58	5.52	5.48	5.55
38 ^b			0.117	0.878	0.163	0.684	5.54	5.64	5.66	5.71	5.74
39			0.266	0.727	0.093	0.053	5.89	5.88	5.87	5.87	5.93
40			0.018	0.669	0.055	0.000	5.92	5.63	5.63	5.61	5.74
41 ^b			0.207	0.268	0.256	0.158	7.03	6.67	6.62	6.49	6.75
42 ^b			0.145	0.480	0.423	0.579	7.01	6.33	6.23	6.12	6.16
43 ^b			0.462	0.170	0.217	0.105	7.68	7.20	7.17	7.05	7.20

(Contd.)

Table 1 — Molecular descriptors, observed and modeled herbicidal activities of aryloxyacetic acid derivatives (Contd.)

44			1.000	0.338	0.163	0.632	7.64	7.87	7.93	7.96	7.82
45			0.152	0.000	0.299	0.211	6.92	7.12	7.06	6.87	7.01
46			0.240	0.860	0.380	0.526	5.60	5.73	5.63	5.60	5.52
47			0.073	0.700	0.173	0.105	5.66	5.66	5.62	5.58	5.73
48			0.111	0.807	0.120	0.579	5.66	5.73	5.76	5.81	5.88
49 ^b			0.087	0.523	0.256	0.211	5.70	6.04	5.98	5.89	5.80
50 ^b			0.172	0.445	0.260	0.368	6.64	6.38	6.34	6.25	6.53
51			0.142	0.506	0.167	0.842	6.68	6.45	6.50	6.51	6.83

^aTaken from Ref.16, ^bcompound in test set, ^coutlier compound in present study

The dataset was split into training and test subsets to ensure accurate modeling. External validation of the correlation model was conducted, while internal consistency was assessed through the LOO and L5O procedures. The test subset was selected carefully to include 20% of the population, with a range of pK_i values and structural diversity. The software used in this study will be discussed in the upcoming sections.

Molecular descriptors

The *At*HPPD inhibitor structures (as shown in Table 1) were initially drawn in 2D ChemDraw using standard procedures¹⁷. To ensure a defined conformer relationship among the compounds being studied, these structures were converted into 3D modules and underwent energy minimization in MOPAC using the AM1 procedure for a closed-shell system. The

DRAGON software was then used to compute descriptors belonging to 0D-, 1D-, and 2D-classes as outlined in (Table 2) to address different structural features¹⁸. To ensure equal influence in the QSAR models, the descriptors were scaled¹⁹ to values between 0 and 1. This approach avoids the dominance of pre-scaled descriptors with larger or smaller values. Finally, the QSAR models were developed using the combinatorial protocol in multiple linear regression (CP-MLR) computational procedure²⁰, based on various scaled descriptors.

Model development

The CP-MLR method uses a filter-based variable selection approach to develop highly significant QSAR models. The process and execution of this method are detailed in our earlier publications²¹⁻²⁵. The computational software incorporates four intrinsic filters that ensure efficient variable selection, resulting in unique and significant QSAR models. Filter-1 allows only variables with inter-parameter correlations to activity profiles below a predefined upper limit (≤ 0.79) to enter the equation. Filter-2 controls variable entry through t-values of coefficients (≥ 2.0 threshold value). Filter-3 ensures comparability

of equations with varying numbers of variables by using the square root of the adjusted multiple correlation coefficient, \bar{r} . Filter-4 evaluates equation consistency using leave-one-out (LOO) cross-validation as the default option and a Q^2 threshold value of $0.3 \leq Q^2 \leq 1.0$. To obtain descriptors with higher information content and explanatory power, the threshold of filter-3 is incrementally increased with each generation, using the \bar{r} value of the preceding optimum model as the new threshold.

To determine any potential correlations, each cross-validated model underwent a randomization test involving repeated randomization of the activity^{26,27}. The activity was scrambled for each model 100 times in simulation runs. The percentage of scrambled activity models with regression statistics equal to or better than the original activity model was calculated to express the chance correlation of the model under consideration. The statistical significance of a model was evaluated using the multiple correlation coefficient (r), standard deviation (s) and the F-ratio between the variances of calculated to observed activities (F). The model's internal validation was determined through the cross-validated index, Q^2 ,

Table 2 — Identified descriptors along with their physical meaning, average regression coefficient and incidence^a, in modeling the inhibition actions of aryloxyacetic acid derivatives against *At*HPPD

S. No.	Descriptor	Class	Physical meaning	Average regression coefficient (incidence)
1	TIE	TOPO	E-state topological parameter	1.333 (15)
2	CENT	TOPO	Centralization parameter	-0.948 (2)
3	VAR	TOPO	Variation parameter	-0.882 (2)
4	IC1	TOPO	Information content index (neighborhood symmetry of 1-order)	1.360 (4)
5	SIC1	TOPO	Structural information content (neighborhood symmetry of 1-order)	1.978 (2)
6	CIC1	TOPO	Complementary information content (neighborhood symmetry of 1-order)	-1.854 (4)
7	BIC1	TOPO	Bond information content (neighborhood symmetry of 1-order)	1.962 (2)
8	CIC2	TOPO	Complementary information content (neighborhood symmetry of 2-order)	-1.771 (3)
9	IC4	TOPO	Information content index (neighborhood symmetry of 4-order)	-0.774 (1)
10	IC5	TOPO	Information content index (neighborhood symmetry of 5-order)	-0.792 (2)
11	BIC5	TOPO	Bond information content (neighborhood symmetry of 5-order)	-0.838 (1)
12	piID	TOPO	Conventional bond-order ID number	-0.989 (1)
13	PCD	TOPO	Difference of multiple path counts to path counts	-0.985 (1)
14	BELe4	BCUT	Lowest eigenvalue no. 4 of Burden matrix/weighted by atomic Sanderson electronegativities	-1.093 (1)
15	GGI1	GALVZ	Topological charge index of order 1	-0.862 (1)
16	GGI6	GALVZ	Topological charge index of order 6	-1.053 (1)
17	JGI5	GALVZ	Mean topological charge index of order 5	0.621 (3)
18	MATS8m	2DAUTO	Moran autocorrelation-lag 8/weighted by atomic masses	-0.771 (9)
19	MATS1e	2DAUTO	Moran autocorrelation-lag 1/weighted by atomic Sanderson electronegativities	0.470 (1)
20	GATS6p	2DAUTO	Geary autocorrelation-lag 6/ weighted by atomic polarizabilities	0.580 (4)

^aThe average regression coefficient of the descriptor relating to all models and the total number of incidents.

using leave-one-out (Q^2_{LOO}) and leave-five-out (Q^2_{L5O}) procedures. The Q^2 value, greater than 0.5 represents an internally robust model. The developed model's external validation or predictive power was assessed using test set compounds. The squared correlation coefficient (r^2_{Test}) between the observed and predicted data of the test set was computed for this purpose. The training set is deemed to have reliable predictive power if the r^2_{Test} index exceeds 0.5.

Partial least squares analysis

PLS (Partial Least Squares) linear regression is a powerful modeling technique that solves the issues found in MLR, such as noisy, collinear, and incomplete variables in the descriptor matrix X and activity profile matrix Y²⁸⁻³⁰. This technique projects matrix X onto a small number of latent variables (LVs) called PLS components, which are linear combinations of the original variables. Meanwhile, matrix Y estimates the LVs in X that are most appropriate for predicting the Y variables. Before applying the PLS approach, the descriptors are auto-scaled using weights based on the variables' standard deviation, ensuring the data is mean-centered. Scaling the matrix X is critical because preprocessed descriptors often have a different order of magnitude.

Cross-validation was performed to determine the optimal number of LVs by excluding some samples from calibration and using them for prediction. This was repeated until each sample was left out once. The predicted values for the left-out samples were then compared with the observed values using the predicted residual sum of squares (PRESS). The PRESS was calculated for each new LV added to the model during cross-validation.

Applicability domain

The importance of a QSAR model is decided by its exact prediction ability for new analogs of the series. A model is applicable only within its training domain, and new analogs must be evaluated as belonging to this domain. The applicability domain (AD) is resolved through the Williams plot in which standardized residuals are plotted against the leverage values, h_i of all compounds belonging to the training domain^{31,32}. The AD is ascertained through a square area within the Williams plot. This area is obtained by considering a measure, $\pm\beta \times (s.d.)$, and the leverage threshold h^* . The h^* is usually fixed at $3(k+1)/n$, where n is the number of compounds considered in the training set and k is the number of independent

descriptors of the model under consideration. The value of β varies from 2 to 3. The plot may be used for graphical detection of the Y outlier (response outlier) and the X outlier (structurally influential compound) in a QSAR model under investigation. The prediction becomes reliable for a compound when its leverage value is lower than the threshold value ($h < h^*$). In such a case, the predicted and the observed activity values of the training set compounds mutually agree. Conversely, the prediction is unreliable when $h > h^*$.

Results and Discussion

A set of 488 descriptors was calculated for 51 compounds in DRAGON, spanning 0D-2D classes. These descriptors were further scaled between 0 and 1 and used to generate statistical models *via* CP-MLR. Filter-3 was employed, gradually increasing the number of descriptors considered for each equation by using the r-bar value of the previous best model as the new threshold. Ultimately, only models utilizing four descriptors proved statistically significant in explaining the activity profiles of the compounds. Among these models, Eq. (1) was determined to have the most significant statistical parameters.

$$pK_i = 6.886 + 1.396(\pm 0.288)TIE - 1.901(\pm 0.247)CIC1 - 0.232(\pm 0.288)GGI6 + 0.430(\pm 0.209)JGI5$$

$$n = 51, r = 0.836, s = 0.377, F(4,46) = 26.769,$$

$$AIC = 0.173, FIT = 1.598, LOF = 0.180,$$

$$Q^2_{LOO} = 0.606, Q^2_{L5O} = 0.589 \quad \dots(1)$$

Table 1 presents the calculated pK_i values, as per the equation above, for comparative purposes with the observed ones. The observed *versus* calculated pK_i s chart is displayed in Figure 1, where compound 28 exhibits a negative deviation of approximately 20% from the normal behavior of the remaining series compounds. To obtain the subsequent model, this

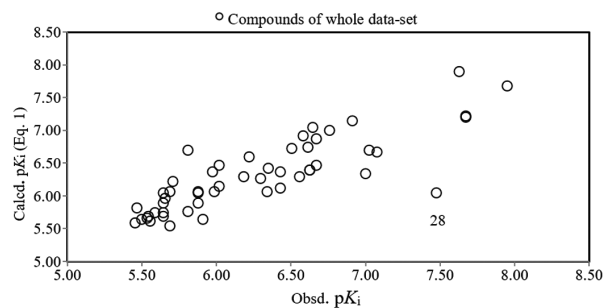


Fig. 1 — Plot of observed *versus* calculated pK_i values, using Eq. (1), for inhibition of *AtHPPD* for all compounds listed in Table 1

compound was excluded from the dataset and the resulting Eq. (2) is presented.

$$\begin{aligned} pK_i = & 6.944 + 1.417(\pm 0.225)\text{TIE} - 1.937(\pm 0.193)\text{CIC1} \\ & - 0.731(\pm 0.242)\text{GGI6} + 0.544(\pm 0.165)\text{JGI5} \\ n = & 50, r = 0.898, s = 0.294, F(4,45) = 47.128, \\ \text{AIC} = & 0.106, \text{FIT} = 2.856, \text{LOF} = 0.111, \\ Q^2_{\text{LOO}} = & 0.766, Q^2_{\text{L50}} = 0.773 \quad \dots(2) \end{aligned}$$

The statistical parameters exhibit significant improvement compared to Eq. (1), with the r^2 -value accounting for almost 81% of the variance in observed activity values. The pK_i calculated through Eq. (2) and listed in (Table 1) are consistent with observed values. However, the behavior of compound 28 is peculiar and warrants thorough investigation to uncover its mode of action at receptor sites and, or meticulously evaluate its herbicidal activity.

After a random selection process, a test set was chosen for external validation of models developed from the training set compounds. Table 1 highlights the ten compounds (4, 13, 24, 29, 38, 41, 42, 43, 49, and 50) selected for the test set, while the remaining 40 compounds make up the training set. With an r -bar ≥ 0.915 , a total of 15 significant models with four descriptors and 20 shared descriptors were obtained. Below are the top five models listed in order of increasing significance.

$$\begin{aligned} pK_i = & 5.897 + 1.301(\pm 0.195)\text{TIE} + 1.411(\pm 0.173)\text{IC1} \\ & - 0.985(\pm 0.210)\text{PCD} - 0.891(\pm 0.152)\text{MATS8m} \\ n = & 40, r = 0.926, s = 0.253, F(4,35) = 52.779, \\ \text{AIC} = & 0.082, \text{FIT} = 3.770, \text{LOF} = 0.088, \\ Q^2_{\text{LOO}} = & 0.817, Q^2_{\text{L50}} = 0.781, r^2_{\text{Test}} = 0.565 \quad \dots (3) \end{aligned}$$

$$\begin{aligned} pK_i = & 5.899 + 1.301(\pm 0.195)\text{TIE} + 1.417(\pm 0.172)\text{IC1} \\ & - 0.989(\pm 0.210)\text{piID} - 0.890(\pm 0.152)\text{MATS8m} \\ n = & 40, r = 0.926, s = 0.253, F(4, 35) = 52.781, \\ \text{AIC} = & 0.082, \text{FIT} = 3.770, \text{LOF} = 0.088, \\ Q^2_{\text{LOO}} = & 0.817, Q^2_{\text{L50}} = 0.827, r^2_{\text{Test}} = 0.567 \quad \dots (4) \end{aligned}$$

$$\begin{aligned} pK_i = & 6.217 + 1.278 (\pm 0.191)\text{TIE} + 1.318(\pm 0.172)\text{IC1} \\ & - 1.093(\pm 0.220)\text{BELe4} - 0.852(\pm 0.149)\text{MATS8m} \\ n = & 40, r = 0.929, s = 0.248, F(4, 35) = 55.499, \\ \text{AIC} = & 0.079, \text{FIT} = 3.964, \text{LOF} = 0.084, \\ Q^2_{\text{LOO}} = & 0.818, Q^2_{\text{L50}} = 0.828, r^2_{\text{Test}} = 0.678 \quad \dots (5) \end{aligned}$$

$$\begin{aligned} pK_i = & 7.188 + 1.336 (\pm 0.189)\text{TIE} - 2.176(\pm 0.180)\text{CIC1} \\ & - 0.862(\pm 0.189)\text{GGI1} - 0.730(\pm 0.158)\text{JGI5} \\ n = & 40, r = 0.931, s = 0.245, F(4, 35) = 56.723, \\ \text{AIC} = & 0.077, \text{FIT} = 4.052, \text{LOF} = 0.082, \\ Q^2_{\text{LOO}} = & 0.821, Q^2_{\text{L50}} = 0.818, r^2_{\text{Test}} = 0.657 \quad \dots (6) \end{aligned}$$

$$\begin{aligned} pK_i = & 6.823 + 1.497(\pm 0.192)\text{TIE} - 1.773(\pm 0.170)\text{CIC1} \\ & - 1.053(\pm 0.220)\text{GGI6} + 0.646(\pm 0.150)\text{JGI5} \\ n = & 40, r = 0.933, s = 0.241, F(4, 35) = 59.290, \\ \text{AIC} = & 0.074, \text{FIT} = 4.235, \text{LOF} = 0.079, \\ Q^2_{\text{LOO}} = & 0.826, Q^2_{\text{L50}} = 0.829, r^2_{\text{Test}} = 0.529 \quad \dots (7) \end{aligned}$$

In all of the aforementioned equations, the F-values hold significant at the 99% level [$F_{4,35}(0.01) = 3.908$] and the standard errors linked with regression coefficients (data within the parentheses) are significant at over a 95% level. The indices Q^2_{LOO} and Q^2_{L50} (>0.5) demonstrated the internal robustness of the derived models, while the index r^2_{Test} greater than 0.5 confirmed that the selected test set was responsible for the external validation of the said models. The signs of the regression coefficients represent the direction of influence of explanatory variables; a positive regression coefficient related to a descriptor will enhance the activity of a compound, while a negative coefficient will cause a detrimental effect. Table 2 contains the class, brief description, average regression coefficients, and total incidences of the 20 identified descriptors.

In the TOPO class, certain descriptors have positive average regression coefficients, including the e-state topological parameter, TIE, the information content index (neighborhood symmetry of 1-order), IC1, the structural information content (neighborhood symmetry of 1-order), SIC1 and the bond information content (neighborhood symmetry of 1-order), BIC1. Conversely, the descriptors with negative average regression coefficients include the centralization parameter, CENT, the variation parameter, VAR, the complementary information content (neighborhood symmetry of 1-order and 2-order), CIC1 and CIC2, the information content index (neighborhood symmetry of 4-order and order-5), IC4 and IC5, the bond information content (neighborhood symmetry of 5-order), BIC5, the conventional bond-order ID number, piID and the difference of multiple path counts to path counts, PCD.

From BCUT class, the descriptor accounting for the lowest eigenvalue no. 4 of Burden matrix/weighted by atomic Sanderson electronegativities, BELe4 has shown a negative average regression coefficient.

The mean topological charge index of order-5, JGI5, has a positive contribution, while the topological charge indices of order-1 and order-6, GGI1 and GGI6, respectively, have negative average regression coefficients in the GALVZ class.

The 2DAUTO class provides descriptors such as the Moran autocorrelation-lag 1 (weighted by atomic Sanderson electronegativities), MATS1e and the Geary autocorrelation-lag 6 (weighted by atomic polarizabilities), GATS6p. Notably, both of these descriptors yield positive average regression coefficients. Conversely, the Moran autocorrelation-lag 8 (weighted by atomic masses), MATS8m, has a negative average regression coefficient. The incidences of participation of all above-mentioned descriptors in 15 developed models are mentioned, within parentheses, in the last column of (Table 2).

After analyzing Eqs. (3)-(7), Eq. (7) was found to be the most significant and, thus, will be discussed further. The correlation coefficient obtained from this equation explains 87% ($r^2 = 0.870$) of the variance in observed inhibition activity profiles. Additionally, other statistical parameters of this equation also suggest that it is the most significant model. Therefore, Eq. (7) was used to calculate the *AtHPPD* inhibition activities of all 50 congeners in the series, which are included in (Table 1). Table 1 and Figure 2 show a close agreement between observed and calculated pK_i 's for both the training and test set compounds. From Eq. (7), it is apparent that higher values of descriptors TIE and JGI5 and lower values of descriptors CIC1 and GGI6, are desirable to enhance the activity of a new compound.

Table 3 includes the results of the PLS analysis on 20 descriptors associated with the *AtHPPD* inhibition activity of the compounds. To ensure equal weight, the descriptors were auto-scaled (zero mean and unit standard deviation) for this study. Three components were optimal for cross-validation and explained 87% ($r^2 = 0.874$) of the variance in observed activity values. The PLS equations of the three components and MLR-like PLS coefficients of these descriptors for *AtHPPD* inhibition activities are provided in

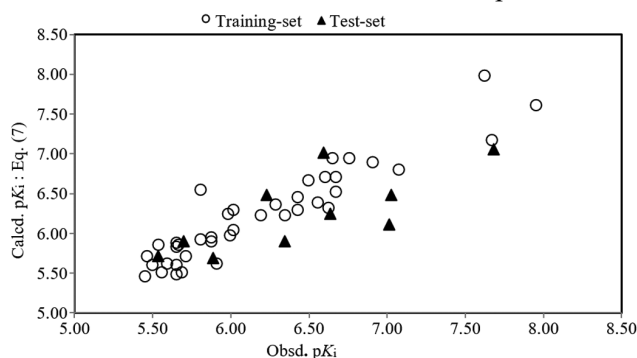


Fig. 2 — Plot between observed and calculated pK_i values, using Eq. (7), for inhibition of *AtHPPD*

(Table 3). The calculated pK_i values for the training and test set compounds closely matched the observed values and are listed in (Table 1). Figure 3 depicts the observed *versus* calculated pK_i values for both sets of compounds. Figure 4 displays a graphical representation of the fraction contribution of normalized regression coefficients of identified descriptors to the activity. Table 3 includes different orders indicating the significance level of the 20 descriptors. Lower orders indicate higher significance in addressing the herbicidal activity for a given descriptor. Descriptors with positive contributions will enhance the activity, and their higher values are desirable for further improvement. Conversely, descriptors with negative contributions will reduce the activity. Therefore, lower or more negative values of such descriptors may enhance the activity of a compound. These guidelines may be beneficial in future synthesis of new potential congeners of the series.

The model based on the complete dataset was analyzed for its applicability domain (AD), which was demonstrated through the Williams plot. This plot displayed standardized residuals plotted against

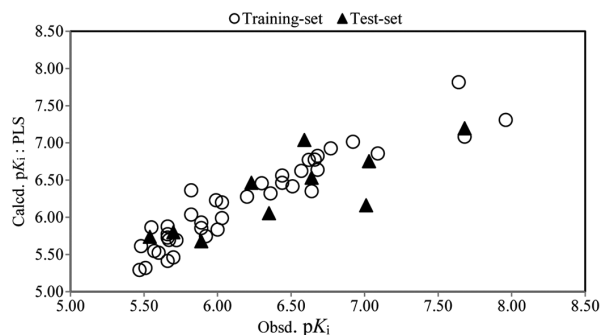


Fig. 3 — Plot between observed and calculated pK_i values, using PLS analysis, for inhibition of *AtHPPD*

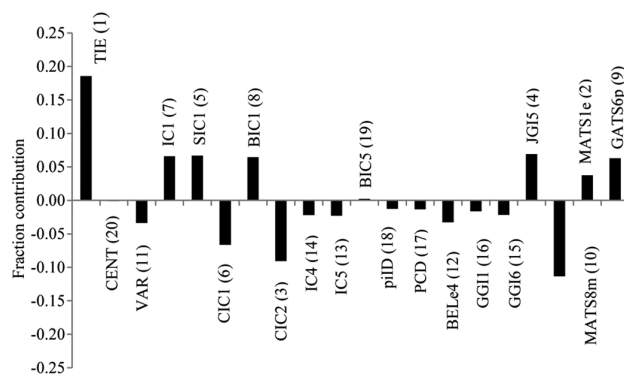


Fig. 4 — Plot of fraction contribution of MLR-like PLS coefficients (normalized) against 20 identified descriptors (Table 2) associated with herbicidal activities of the compounds

Table 3 — PLS and MLR-like PLS equations from the descriptors of CP-MLR identified models for herbicidal Activity

A.PLS equation				B.PLS Regression statistics			
PLS components	PLS coefficient (s.e.) ^a			Symbol	Estimate		
Component-1	-0.219 (0.015)			n	40		
Component-2	-0.116 (0.019)			r	0.935		
Component-3	0.068 (0.019)			s	0.234		
Constant	6.212			F	84.088		
				Q ² _{LOO}	0.843		
				Q ² _{L50}	0.805		
				r ² _{Test}	0.680		
C.MLR-Like PLS equation							
S. No.	Descriptor	MLR-like Coefficient	FC(order) ^b	S. No.	Descriptor	MLR-like Coefficient	FC(order) ^b
1	TIE	1.024	0.186(1)	11	BIC5	0.011	0.002(19)
2	CENT	-0.008	-0.001(20)	12	piID	-0.076	-0.013(18)
3	VAR	-0.203	-0.034(11)	13	PCD	-0.080	-0.013(17)
4	IC1	0.320	0.066(7)	14	BELe4	-0.211	-0.033(12)
5	SIC1	0.333	0.067(5)	15	GGI1	-0.084	-0.016(16)
6	CIC1	-0.330	-0.066(6)	16	GGI6	-0.133	-0.022(15)
7	BIC1	0.317	0.064(8)	17	JGI5	0.302	0.069(4)
8	CIC1	-0.532	-0.091(3)	18	MATS8m	-0.500	-0.113(2)
9	IC4	-0.107	-0.022(14)	19	MATS1e	0.163	0.037(10)
10	IC5	-0.111	-0.023(13)	20	GATS6p	0.233	0.063(9)
				Constant	6.287		

^aRegression coefficient of PLS factor and its standard error. ^bCoefficients of MLR-like PLS equation in terms of descriptors for their original values, FC is fraction of contribution of regression coefficient, computed from the normalized regression coefficients obtained from the autoscaled (zero mean and unit standard deviation) data.

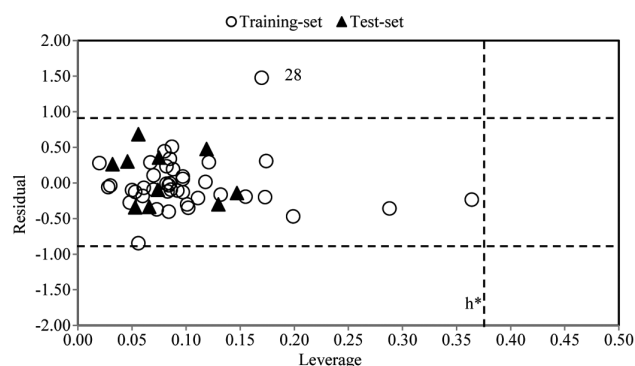


Fig. 5 — William plot based on whole data set for *A*HPPD inhibition activity of compounds given in Table 1 (h^* value is 0.375 and residual limits are $\pm 3.0 \times s.d.$)

leverage (h_i) values for the most influential descriptors - TIE, CIC1, GGI6 and JGI5. A model related to the complete dataset ($n = 51$) was developed using these descriptors and is represented in Eq. (1). The AD was established using the standardized residuals and leverage values calculated in conjunction with it. Y-outliers were considered as $\pm \beta \times s.d.$ and the leverage threshold as $h^* (= 3(k + 1)/n)$. Figure 5 provides a graphical representation of

the suggested model for the influential descriptors, depicting both the training and test set compounds. The model shows good fitting power and can assess external data, matching the most significant parameters for both sets of compounds.

All compounds, except compound 28, were found within the AD, indicating that the model in question can accurately evaluate both training and test set compounds. Compound 28, however, stands out as a clear Y-outlier, as it is not within the AD. It is possible that this compound is operating through a different mechanism or interacting with the receptor uniquely.

Conclusion

Using of chemometric descriptors, the inhibition activities of aryloxyacetic acid derivatives against the *Arabidopsis thaliana* 4-hydroxyphenylpyruvate dioxygenase (*A*HPPD) were quantitatively analyzed. The resulting models, developed statistically, offer insight into the inhibitory properties of these congeners. In all, 15 models were generated from 20 influential descriptors. The models that were created underwent external validation using a specified test

set. Additionally, their internal robustness was assessed by implementing leave-one-out and leave-five-out procedures. In a given model, both negative and positive regression coefficients associated with descriptors were obtained. The higher positive value of the descriptor, bearing a positive regression coefficient, and a more negative value of the descriptor, having a negative regression coefficient, are desirable to further improve the activity profile of a compound.

The findings of the PLS study support the importance of the descriptors that were identified by the CP-MLR filters. Moreover, the AD analysis confirms the reliability of the suggested model, as it successfully predicts the properties of both the training and test sets within the defined domain. One analogue was identified as an outlier, potentially interacting with the receptor in dissimilar modes. It is worth noting that the model in question has accurately predicted the *A*HPPD inhibition activities of all compounds in the series.

Acknowledgement

The research facilities provided to one of the author (NS) is thankfully acknowledged. We also appreciate the help from our colleagues.

Conflict of interest

All authors declare no conflict of interest.

References

- 1 Yang C, Pflugrath JW, Camper DL, Foster ML, Pernich DJ & Walsh TA, Structural basis for herbicidal inhibitor selectivity revealed by comparison of crystal structures of plant and mammalian 4-hydroxyphenylpyruvate dioxygenases. *Biochem*, 43 (2004) 10414.
- 2 Hall MG, Wilks MF, Provan WM, Eksborg S & Lumholtz B, Pharmacokinetics and pharmacodynamics of NTBC (2-(2-nitro-4-fluoromethylbenzoyl)-1,3-cyclohexanedione) and mesotrione, inhibitors of 4-hydroxyphenyl pyruvate dioxygenase (HPPD) following a single dose to healthy male volunteers. *Br J Clin Pharmacol*, 52 (2001) 169.
- 3 Lewis RW & Botham JW, A review of the mode of toxicity and relevance to humans of the triketone herbicide 2-(4-methylsulfonyl-2-nitrobenzoyl)-1,3-cyclohexanedione. *Crit Rev Toxicol*, 43 (2013) 185.
- 4 He B, Wang D, Yang W, Chen Q & Yang, G, Advances in research on 4-hydroxyphenylpyruvate dioxygenase (HPPD) structure and pyrazole-containing herbicides. *Chin J Org Chem*, 37 (2017) 2895.
- 5 Hawkes TR, Langford MP, Viner R, Blain RE, Callaghan FM, Mackay EA, Hogg BV, Singh S & Dale RP, Characterization of 4-hydroxyphenylpyruvate dioxygenases, inhibition by herbicides and engineering for herbicide tolerance in crops. *Pestic Biochem Physiol*, 156 (2019) 9.
- 6 Li HB, Li L, Li JX, Han TF, He JL & Zhu YQ, Novel HPPD inhibitors: Triketone 2*H*-benzo[*b*][1,4]oxazin-3(4*H*)-one analogs. *Pest Manage Sci*, 74 (2018) 579.
- 7 Neidig ML, Decker A, Choroba OW, Huang F, Kavana M, Moran GR, Spencer JB & Solomon EI, Spectroscopic and electronic structure studies of aromatic electrophilic attack and hydrogen-atom abstraction by non-heme iron enzymes. *Proc Natl Acad Sci USA*, 103 (2006) 12966.
- 8 Borowski T, Bassan A, Siegbahn PEM, 4-Hydroxyphenylpyruvate dioxygenase: A hybrid density functional study of the catalytic reaction mechanism. *Biochem*, 43 (2004)12331.
- 9 Ahrens H, Lange G, Müller T, Rosinger C, Willms L, van Almsick A, 4-Hydroxyphenylpyruvate dioxygenase inhibitors in combination with safeners: Solutions for modern and sustainable agriculture. *Angew Chem Int Edn*, 52 (2013) 9388.
- 10 Heap I, International survey of herbicide resistant weeds. *Weed Sci*, 2025
- 11 Beaudegnies R, Edmunds AJF, Fraser TEM, Hall RG, Hawkes TR, Mitchell G, Schaezter J, Wendeborn S & Wibley J, Herbicidal 4-hydroxyphenylpyruvate dioxygenase inhibitors-a review of the triketone chemistry story from a Syngenta perspective. *Bioorg Med Chem*, 17 (2009) 4134.
- 12 Zhu YQ, Hu FZ & Yang HZ, Reviews on 4-hydroxyphenylpyruvate dioxygenase enzyme and the structure-activity relationships of its inhibitors. *Huaxue Tongbao*, 67 (2004) w018/1 (in Chinese).
- 13 Freitas LM, de Assis Valadares LP, Camozzi MGM, de Oliveira PG, Ferreira Machado MR & Lima FC, Animal models in the neurotoxicology of 2,4-D. *Hum Exp Toxicol*, 38 (2019) 1178.
- 14 Ndikuryayo F, Kang WM, Wu FX, Yang WC & Yang GF, Hydrophobicity-oriented drug design (HODD) of new human 4-hydroxyphenylpyruvate dioxygenase inhibitors. *Eur J Med Chem*, 166 (2019) 22.
- 15 Fu Y, Zhang S-Q, Liu Y-X, Wang J-Y, Gao S, Zhao L-X & Ye F, Design, synthesis, SAR and molecular docking of novel green niacin-triketone HPPD inhibitor. *Ind Crops Prod*, 137 (2019) 566.
- 16 Wang MM, Huang H, Shu L, Liu JM, Zhang JQ, Yan YL & Zhang DY, Synthesis and herbicidal activities of aryloxyacetic acid derivatives as HPPD inhibitors. *Beilstein J Org Chem*, 16 (2020) 233.
- 17 ChemDraw ultra 6.0 and Chem3D ultra, Cambridge Soft Corporation, Cambridge, USA.
- 18 DRAGON software, version 3.0-2003; by Todeschini Y, Consonni V, Mauri A & Pavan M, Milano, Italy. Available from: <http://disat.unimib.it/chm/Dragon.htm>.
- 19 Golbraikh A & Tropsha A, Beware of q²! *J Mol Graph Model*, 20 (2002) 269.
- 20 Prabhakar YS, 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 (2003) 583.
- 21 Sharma S, Sharma BK, Sharma SK, Singh P & Prabhakar YS, Topological descriptors in modeling the agonistic activity of human A₃ adenosine receptor ligands: The derivatives of 2-chloro-*N*⁶-substituted-4'-thioadenosine-5'-uronamide. *Eur J Med Chem*, 44 (2009) 1377.
- 22 Sharma BK, Paliana P, Singh P & Prabhakar YS, 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 (2010) 169.
- 23 Sharma BK, Singh P, Sarbhai K & Prabhakar YS, A quantitative structure-activity relationship study on serotonin 5-HT₆ receptor ligands: Indolyl and piperidinyl sulphonamides. *SAR QSAR Environ Res*, 21 (2010) 369.
- 24 Sharma BK, Pilania P, Sarbhai K, Singh P & Prabhakar YS, Chemometric descriptors in modeling the carbonic anhydrase inhibition activity of sulfonamide and sulfamate derivatives. *Mol Divers*, 14 (2010) 371.
- 25 Sharma BK, Singh P, Shekhawat M, Sarbhai K & Prabhakar YS, Modelling of serotonin reuptake inhibitory and histamine H₃ antagonistic activity of piperazine and diazepane amides: QSAR rationales for co-optimization of the activity profiles. *SAR QSAR Environ Res*, 22 (2011) 365.
- 26 So SS & Karplus M, Three-dimensional quantitative structure-activity relationships from molecular similarity matrices and genetic neural networks. 1. Method and validations. *J Med Chem*, 40 (1997) 4347.
- 27 Prabhakar YS, Solomon VR, Rawal RK, Gupta MK & Katti SB, CP-MLR/PLS Directed structure-activity modeling of the HIV-1 RT inhibitory activity of 2,3-diaryl-1,3-thiazolidin-4-ones. *QSAR Comb Sci*, 23 (2004) 234.
- 28 Wold S, Cross-validatory estimation of the number of components in factor and principal components models. *Technomet*, 20 (1978) 397.
- 29 Kettaneh N, Berglund A & Wold S, PCA and PLS with very large data sets. *Comput Stat Data Anal*, 48 (2005) 69.
- 30 Stahle L & Wold S, Multivariate data analysis and experimental design in biomedical research. In *Progress in Medicinal Chemistry Vol 25*, (Ed. by GP Ellis & GB West; Elsevier Science Publishers, BV) 1988, 291.
- 31 Gramatica P, Principles of QSAR models validation: Internal and external. *QSAR Comb Sci*, 26 (2007) 694.
- 32 Eriksson L, Jaworska J, Worth AP, Cronin MTD, McDowell RM & Gramatica P, Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. *Environ Health Persp*, 111 (2003) 1361.