Indian Journal of Chemistry
Vol. 65, July 2026, pp. 697-709
DOI: 10.56042/ijc.v65i7.27713

In silico investigation of novel arylthioacetic acid derivatives as inhibitors of the 4-hydroxyphenylpyruvate dioxygenase

Prithvi Singh[#]

Department of Chemistry, Government Science College, Sikar 332 001, Rajasthan, India

E-mail: psingh_sikar@rediffmail.com, singhprithvi700@gmail.com

Received 18 January 2026; accepted (revised) 16 June 2026

This study investigates the inhibitory potential of arylthioacetic acid derivatives against 4-hydroxyphenylpyruvate dioxygenase from *Arabidopsis thaliana* (*At*HPPD) using *in silico* methods. The goal is to design and optimize potential *At*HPPD inhibitors through QSAR modeling and molecular analyses. These computational approaches provide a cost-effective and time-efficient alternative to traditional drug discovery, enabling the prediction of biological activity for untested compounds. A dataset of sixty-nine arylthioacetic acid derivatives has been analyzed to develop five significant regression models using eleven chemometric molecular descriptors. Model robustness has been validated through internal checks and randomization tests. Partial least-squares (PLS) analysis highlights the importance of the descriptors. The best model has been used to predict the activities of additional analogues, and applicability domain analysis has confirmed its reliable performance. These findings offer insights for identifying and designing new potent analogues in this chemical series. Furthermore, ADME-T (Absorption, Distribution, Metabolism, Excretion, and Toxicity) properties of the most active compounds have been evaluated and compared with mesotrione, a commercially established HPPD-inhibiting herbicide. Three compounds (11, 40, and 49) have been selected for detailed assessment. Similar to mesotrione, all three compounds satisfy Lipinski's rule of five and exhibit favorable predicted human intestinal absorption. However, unlike mesotrione, they exhibit strong plasma protein binding. Toxicity predictions indicate that compounds 11 and 40 are non-carcinogenic in mice, whereas compound 49 shows carcinogenic potential. In rats, only compound 40 exhibits possible carcinogenicity, while compounds 11, 49, and mesotrione are predicted to be non-carcinogenic. Additionally, the compounds are classified as inhibitors or non-inhibitors of key cytochrome P450 enzymes, namely CYP2C19, CYP2D6, and CYP3A4. Molecular docking studies have been performed on the most active compounds and compared with mesotrione to elucidate binding interactions within the *At*HPPD active site. Notably, compound 49 exhibits a binding affinity comparable to mesotrione and demonstrates stronger intermolecular interactions, highlighting its potential as a promising lead candidate. Overall, this study provides valuable molecular insights and supports further exploration of arylthioacetic acid derivatives as potent *At*HPPD inhibitors.

Keywords: Arylthioacetic acid derivatives, Inhibitors of *At*HPPD, Chemometric molecular descriptors, CP-MLR analysis, Partial least-squares (PLS), Molecular docking

Weed resistance has increased rapidly in recent years, posing a serious challenge to the agricultural sector and threatening sustainable crop production. Consequently, the discovery and development of new herbicides have become an urgent priority for modern agriculture[1,2]. The 4-hydroxyphenylpyruvate dioxygenase (HPPD) is an essential enzyme found in microbes, mammals, and plants, where it performs distinct physiological functions depending on the organism. During HPPD-catalyzed reactions, 4-hydroxyphenylpyruvic acid (HPPA) coordinates with Fe^{2+} to form a chelate complex, enabling the conversion of HPPA into homogentisic acid (HGA)[3-8].

HPPD-inhibiting herbicides exert their activity by competitively preventing HPPA from chelating Fe^{2+}, thereby blocking its conversion to HGA[6,9-12]. This inhibition disrupts plastoquinone biosynthesis, leading to bleaching symptoms in treated plants and ultimately causing plant death. HPPD inhibitors are particularly attractive as herbicides because of their high efficacy, low mammalian toxicity, broad-spectrum weed control, and favorable crop and environmental safety profiles. However, resistance or reduced sensitivity to certain commercial herbicides, such as mesotrione, has been reported in weeds including *Setaria faberi*. As a result, identifying novel

Former Emeritus Research Fellow, UGC, New Delhi

HPPD inhibitors with enhanced herbicidal activity and improved crop selectivity has become essential.

Several highly effective HPPD-inhibiting herbicides have been successfully commercialized and widely applied in agriculture. Based on their chemical structures, these inhibitors are primarily classified into triketones, pyrazoles, and isoxazoles[13,14]. Among them, mesotrione is recognized as a particularly potent triketone HPPD inhibitor[13]. Extensive research over the past decades has provided valuable insights into enzyme–inhibitor interactions, revealing that 1,3-dicarbonyl systems and aromatic moieties constitute key pharmacophoric elements required for strong HPPD inhibition. Numerous studies[15-25] have demonstrated that structural modification of the aromatic moiety represents an effective strategy for designing new HPPD inhibitors with improved herbicidal performance.

Recently, Wang *et al*.[21] reported the synthesis and herbicidal evaluation of aryloxyacetic acid derivatives as inhibitors of *Arabidopsis thaliana* 4-hydroxyphenylpyruvate dioxygenase (*At*HPPD). In a continued effort to identify novel *At*HPPD inhibitors, Huang *et al*.[26] employed a bioisosterism strategy to design a new series of compounds. Sulfur-containing compounds are ubiquitous in natural products, pharmaceuticals, agrochemicals, photoelectric materials, fragrances, and food substances, making sulfur a valuable element in the design of biologically active molecules. Following this rationale, the introduction of a sulfur atom into the extended side chain led to the development of a new class of arylthioacetic acid derivatives as potential *At*HPPD inhibitors.

In a recent publication, we reported our findings on aryloxyacetic acid–pyrazole derivatives[27]. The molecular structures of these compounds were characterized using chemometric two-dimensional (2D) molecular descriptors, several of which showed strong correlations with inhibitory activity against *At*HPPD. Multiple statistically robust models were developed, and the most significant was used to predict the activity profiles of compounds within the series. A high level of agreement was observed between experimental and predicted activities, providing valuable guidelines for the rational design of more potent analogues. In addition, the possible interaction modes of these compounds at the *At*HPPD active site were investigated.

The present study aims to elucidate the relationship between the structural features of sixty-nine arylthioacetic acid derivatives and their inhibitory activities against *At*HPPD. The generalized chemical structures of these compounds are shown in Fig. 1, and their detailed structural characteristics are summarized in Table 1. Molecular structures were encoded using 2D molecular descriptors, and inhibitory activities were analyzed using linear free-energy relationships. A statistical modeling approach was applied to correlate the dependent biological activity with independent molecular descriptors, thereby establishing a quantitative structure–activity relationship (QSAR) and enabling a precise interpretation of structure–activity trends within this compound series.

Materials and Methods

The dataset used in this study was obtained from the literature[26] and comprised sixty-nine arylthioacetic acid derivatives and their reported inhibitory activities against *At*HPPD. The inhibitory activities were expressed as binding constants (K_i), which were converted to their logarithmic form, pK_i ($-\log K_i$), on a molar basis. This dataset, containing relevant 0D to 2D molecular descriptors, was initially employed to explore correlations between molecular structure and biological activity.

For model development and validation, the dataset was divided into training and test sets. The training set was used to derive QSAR models, while the test set served for external validation. Internal consistency and robustness of the models were assessed using leave-one-out (LOO) and leave-five-out (L5O) cross-

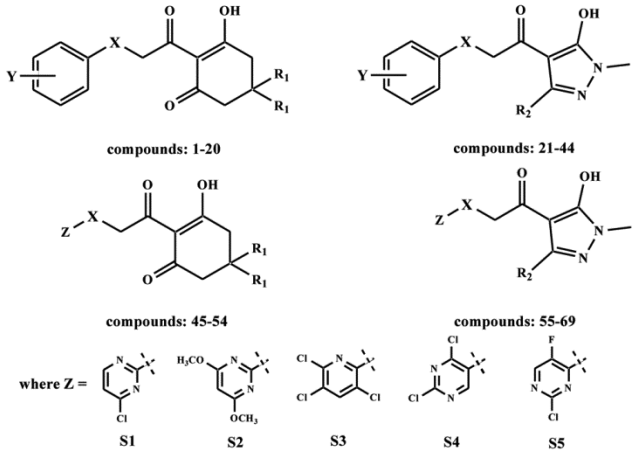

Fig. 1 — Generalized chemical structures of arylthioacetic acid derivatives (features in Table 1)

Table 1 — Chemical structure and *At*HPPD inhibitory activities of arylthioacetic acid derivatives (Fig. 1 for details)

S.No.	X	Y	R_1	R_2	Z	pK_i (M): *At*HPPD Obsd.[a]	Calcd. Eq. (6)	Calcd. PLS
1	$–SO_2-$	4-Me	H	–	–	5.374	5.390	5.226
2	$–SO_2-$	4-Me	Me	–	–	5.000	4.897	4.897
3	$–SO_2-$	H	H	–	–	5.536	5.845	5.608
4	$–SO_2-$	H	Me	–	–	5.000	5.353	5.269
5[b]	S	4-Me	H	–	–	5.776	6.281	6.081
6	S	4-Me	Me	–	–	5.540	5.764	5.868
7	S	4-Cl	H	–	–	7.284	7.258	7.141
8	S	4-Cl	Me	—	–	7.051	6.590	6.705
9	S	4-NO_2	H	—	–	6.932	6.862	6.737
10	S	4-NO_2	Me	–	–	6.697	6.445	6.514
11	S	2-NO_2	H	–	–	7.678	7.351	7.186
12	S	2-NO_2	Me	–	–	7.180	6.887	6.912
13	S	2-NO_2, 4-NO_2	H	–	–	6.374	6.938	6.806
14	S	2-NO_2, 4-NO_2	Me	–	–	6.516	6.701	6.725
15	S	2-CF_3, 4-NO_2	H	–	–	6.446	6.915	6.810
16[b]	S	2-CF_3, 4-NO_2	Me	–	–	6.535	6.782	6.772
17[b]	S	2-NO_2, 4-Cl	H	–	–	7.174	7.566	7.446
18	S	2-NO_2, 4-Cl	Me	–	–	7.108	7.108	7.156
19	S	2-NO_2, 4-CF_3	H	–	–	6.943	6.759	6.600
20	S	2-NO_2, 4-CF_3	Me	–	–	6.804	6.643	6.588
21	$–SO_2-$	4-Me	–	Me	–	5.274	5.195	5.312
22[b]	$–SO_2-$	4-Me	–	Cyclopropyl	–	5.000	4.670	4.884
23	$–SO_2-$	H	–	Me	–	5.312	5.491	5.535
24	$–SO_2-$	H	–	Cyclopropyl	–	5.176	4.894	5.016
25	$–SO_2-$	H	–	CF_3	–	5.726	5.739	5.626
26	S	4-Me	–	Me	–	5.699	6.345	6.408
27[b]	S	4-Me	–	Cyclopropyl	–	5.525	6.159	6.201
28	S	4-Cl	–	Me	–	7.357	6.918	7.094
29	S	4-Cl	–	Cyclopropyl	–	7.018	6.812	6.884
30[b]	S	4-NO_2	–	Me	–	6.959	6.647	6.806
31	S	4-NO_2	–	Cyclopropyl	–	6.821	6.488	6.602
32	S	4-NO_2	–	CF_3	–	6.955	6.524	6.710
33	S	2-NO_2	–	Me	–	7.432	7.117	7.219
34	S	2-NO_2	–	Cyclopropyl	–	7.187	6.962	7.052
35[b]	S	2-NO_2	–	CF_3	–	7.252	6.888	7.058
36	S	2-NO_2, 4-NO_2	–	Me	–	6.750	6.852	6.985
37	S	2-CF_3, 4-NO_2	–	Me	–	6.801	6.860	6.983
38	S	2-CF_3, 4-NO_2	–	Cyclopropyl	–	6.535	6.788	6.964
39[b]	S	2-CF_3, 4-NO_2	–	CF_3	–	6.801	6.618	6.822
40[b]	S	2-NO_2, 4-Cl	–	Me	–	7.658	7.316	7.473
41	S	2-NO_2, 4-Cl	–	Cyclopropyl	–	7.420	7.225	7.349
42	S	2-NO_2, 4-Cl	–	CF_3	–	7.469	7.030	7.270
43	S	2-NO_2, 4-CF_3	–	Me	–	6.575	6.692	6.772
44	S	2-NO_2, 4-CF_3	–	CF_3	–	6.411	6.462	6.623
45	S	–	H	–	S1	6.812	7.283	6.942
46	S	–	Me	–	S1	6.562	6.760	6.659
47	S	–	H	–	S2	7.155	6.878	6.595
48	S	–	Me	–	S2	7.032	6.695	6.598
49	S	–	H	–	S3	7.602	7.712	7.407
50	S	–	Me	–	S3	7.244	7.153	7.064
51[b]	S	–	H	–	S4	7.076	7.868	7.635
52	S	–	Me	–	S4	7.027	7.288	7.275
53[b]	S	–	H	–	S5	7.347	7.588	7.300
54	S	–	Me	–	S5	7.301	7.151	7.044
55	S	–	–	Me	S1	6.879	6.960	6.944

(*Contd.*)

Table 1 — Chemical structure and *At*HPPD inhibitory activities of arylthioacetic acid derivatives (Fig. 1 for details) (*Contd*.)

S.No.	X	Y	R_1	R_2	Z	pK_i (M): *At*HPPD Obsd.[a]	Calcd. Eq. (6)	Calcd. PLS
56	S	–	–	Cyclopropyl	S1	6.724	6.879	6.759
57	S	–	–	CF_3	S1	6.762	7.255	7.230
58	S	–	–	Me	S2	6.860	6.993	6.943
59[b]	S	–	–	Cyclopropyl	S2	6.876	6.846	6.824
60[b]	S	–	–	CF_3	S2	6.703	7.168	7.128
61	S	–	–	Me	S3	7.432	7.413	7.463
62	S	–	–	Cyclopropyl	S3	7.319	7.414	7.361
63	S	–	–	CF_3	S3	7.149	7.247	7.385
64	S	–	–	Me	S4	7.602	7.691	7.801
65	S	–	–	Cyclopropyl	S4	7.456	7.414	7.361
66	S	–	–	CF_3	S4	7.237	7.247	7.385
67	S	–	–	Me	S5	7.284	7.312	7.348
68[b]	S	–	–	Cyclopropyl	S5	7.432	7.253	7.186
69	S	–	–	CF_3	S5	7.131	7.106	7.232

[a]Binding constant, K_i is taken from the reference[26] and is expressed as pK_i on molar basis, [b]Compound in the test set

validation procedures. The test set was carefully selected to represent approximately 20% of the total dataset, ensuring adequate coverage of structural diversity and a wide range of pK_i values. Details of the software and computational tools employed are described in the subsequent sections.

Molecular Descriptors

The chemical structures of the *At*HPPD inhibitors (Table 1) were initially drawn in two-dimensional (2D) format using ChemDraw, following standard procedures[28]. To ensure a consistent conformational framework, all structures were subsequently converted into three-dimensional (3D) models. Geometry optimization was performed using the AM1 semi-empirical method implemented in MOPAC for closed-shell systems.

A comprehensive set of 0D, 1D, and 2D molecular descriptors was calculated using the DRAGON software, capturing diverse physicochemical and topological characteristics of the compounds[29] (Table 2). To prevent descriptors with large numerical values from disproportionately influencing QSAR models, all descriptors were scaled[30] to a range of 0-1. Based on these scaled descriptors, QSAR models were developed using the combinatorial protocol in multiple linear regression (CP-MLR) methodology[31].

Model Development

The CP-MLR approach employs a filter-based variable selection strategy to generate QSAR models that are statistically significant and interpretable. Detailed descriptions of this methodology and its implementation have been reported in our earlier studies[32-36]. The CP-MLR procedure integrates four intrinsic filters that systematically refine descriptor selection and improve model quality.

Filter-1 allows only those descriptors whose intercorrelation with biological activity is below a predefined threshold (≤ 0.79) to enter the model. Filter-2 regulates variable inclusion based on the statistical significance of regression coefficients, requiring a minimum *t*-value of ≥ 2.0. Filter-3 ensures comparability among equations containing different numbers of variables by employing the square root of the adjusted multiple correlation coefficient ($\bar{r}$). Filter-4 evaluates model consistency *via* LOO cross-validation, yielding acceptable Q^2 values of 0.3-1.0. To enhance descriptor relevance and explanatory strength, the Filter-3 threshold is progressively increased across successive model generations, using the $\bar{r}$ value of the preceding optimal model as the new threshold.

To exclude the possibility of chance correlations, each validated model was subjected to a randomization (Y-scrambling) test[37,38] involving 100 simulation runs. In this process, activity values were randomly shuffled while keeping the descriptor matrix intact. The percentage of randomized models yielding regression statistics equal to or better than those of the original model was calculated to assess statistical significance. Model quality was evaluated using the multiple correlation coefficient (r), standard deviation (s), and F-ratio (F). Internal validation was assessed using cross-validated indices (Q^2_{LOO} and Q^2_{L5O}), where Q^2 values greater than 0.5 indicate

Table 2 — Identified descriptors[a] and their physical meaning, average regression coefficient, and incidence[b], in the modeling of *At*HPPD inhibition activity

No.	Name	Class	Physical meaning	Average regression coefficient and incidence
1	Jhetv	TOPO	Balaban-type index from van der Waals volume weighted distance matrix	–1.322 (1)
2	Jhete	TOPO	Balaban-type index from electronegativity-weighted distance matrix	–0.756 (1)
3	Jhetp	TOPO	Balaban-type index from polarizability-weighted distance matrix	–1.050 (1)
4	X2A	TOPO	Average connectivity index chi-2	–0.655 (1)
5	PCR	TOPO	Ratio of multiple path counts to path counts	0.870 (2)
6	BELe4	BCUT	Lowest eigenvalue no. 4 of Burden matrix/weighted by atomic Sanderson electronegativities	–2.093 (5)
7	MATS6m	2D-AUTO	Moran autocorrelation-lag 6/weighted by atomic masses	1.051 (2)
8	MATS6p	2D-AUTO	Moran autocorrelation-lag 6/weighted by atomic polarizabilities	1.120 (1)
9	GATS2v	2D-AUTO	Geary autocorrelation-lag 2/weighted by atomic van der Waals volumes	–1.475 (2)
10	GATS7e	2D-AUTO	Geary autocorrelation-lag 7/weighted by atomic Sanderson electronegativities	–1.172 (1)
11	GATS8e	2D-AUTO	Geary autocorrelation-lag 8/weighted by atomic Sanderson electronegativities	–1.483 (3)

[a]Descriptors were identified from the models and surfaced from the CP-MLR protocol with a training set of 55 compounds for *At*HPPD inhibition activity,[b]Average regression coefficient of the descriptor relating to all models and the total number of its incidence; the arithmetic sign of the coefficient represents the actual sign of the regression coefficient in the models

robust models. External validation was performed using the test set, and predictive ability was quantified by the squared correlation coefficient (r^2_{Test}). Models with r^2_{Test} values exceeding 0.5 were considered predictive.

Partial Least-Squares Analysis

Partial least-squares (PLS) regression is a powerful multivariate technique that overcomes the limitations of conventional multiple linear regression (MLR), particularly in the presence of noisy, collinear, or incomplete descriptor matrices (X) and response variables (Y)[39-41]. In PLS analysis, matrix X is projected onto a reduced number of latent variables (LVs), which are linear combinations of the original descriptors, while simultaneously maximizing covariance with matrix Y.

Prior to PLS modeling, all descriptors were auto-scaled to unit variance and mean-centered to eliminate bias arising from differences in descriptor magnitudes. Cross-validation was performed to determine the optimal number of LVs by sequentially excluding samples from the calibration set and predicting their activity values. This procedure was repeated until each compound was excluded once. Model performance was evaluated using the predicted residual sum of squares (PRESS), calculated for each additional LV added during cross-validation.

Applicability Domain

The predictive reliability of a QSAR model depends on its applicability to compounds that fall within the chemical space defined by the training set. Therefore, the applicability domain (AD) of the developed models was evaluated to ensure reliable predictions for new analogues[42,43]. The AD was assessed using a Williams plot, in which standardized residuals are plotted against leverage values (h_i) for all compounds in the training set.

The AD is represented as a bounded region defined by the leverage threshold (h*), calculated as $3(k + 1)/n$, where n is the number of training compounds and k is the number of descriptors in the model. The standardized residual limits were set at $\pm\beta$ times the standard deviation, with β values typically ranging from 2 to 3. The Williams plot enables identification of response outliers (Y-outliers) and structurally influential compounds (X-outliers). Predictions are considered reliable when leverage values are below the threshold ($h_i < h^*$); compounds exceeding this limit are regarded as outside the model's domain.

ADME-T prediction

The drug-likeness of bioactive compounds is governed by their absorption, distribution, metabolism, excretion, and toxicity (ADME-T) properties. An ideal candidate should exhibit favorable pharmacokinetic behavior in addition to high biological activity. To evaluate ADME-T-related characteristics and ensure compliance with the Lipinski rule of five (Ro5), structural, physicochemical, and pharmacokinetic properties of selected compounds were analyzed using the SwissADME and PreADME-T online prediction servers[44,45].

These tools provided key parameters, including molecular weight (MW), number of rotatable bonds (NRB), partition coefficient ($iLogP_{o/w}$), number of hydrogen bond acceptors (NHA), number of hydrogen bond donors (NHD), topological polar surface area (TPSA), inhibition potential toward cytochrome P450 enzymes (CYP2D6, CYP2C19, and CYP3A4), human intestinal absorption (HIA%), and carcinogenicity. Compounds satisfying specified criteria were considered to possess favorable drug-like properties. ADMET predictions were performed by submitting either the 2D structures or SMILES representations of the compounds to the respective online servers.

Molecular Docking

Molecular docking is a structure-based computational approach used to predict the binding orientation and affinity of small molecules within the active site of a target protein. This technique enables the visualization of ligand–receptor interactions at the atomic level and provides insights into the stability of the resulting complexes.

Docking simulations rely on two complementary components: search algorithms, which explore the ligand's conformational space within the binding pocket, and scoring functions, which evaluate each pose by estimating the binding free energy. In the present study, molecular docking was performed for the three most active compounds in the series, and their binding interactions were compared with those of the reference herbicide mesotrione. This comparative analysis was undertaken to elucidate binding modes and assess the relative inhibitory potential of the selected compounds.

Results and Discussion

A total of 490 molecular descriptors spanning 0D–2D classes were calculated for sixty-nine compounds using DRAGON software. All descriptors were normalized to a 0–1 range prior to model development. Statistical models were generated using the combinatorial protocol in multiple linear regression (CP-MLR). Descriptor selection was performed using Filter-3, in which the best model's $\bar{r}$ threshold at each stage was iteratively used to expand the descriptor pool.

Only four-descriptor models were found to be statistically significant in explaining the inhibition activity. Among these, Equation (1) exhibited the most favorable statistical characteristics and was therefore selected as the optimal global model:

$$p\mathrm{K_i}\ (\mathrm{M}) = 7.477 - 1.130(0.190)\mathrm{BELe4} + 1.204(0.154)\mathrm{MATS6m} - 1.176(0.169)\mathrm{GATS7e} - 1.375(0.226)\mathrm{GATS8e}$$
$$n = 69,\ r = 919,\ s = 0.298,\ F(4, 64) = 86.446,$$
$$\mathrm{AIC} = 0.102,\ \mathrm{LOF} = 0.105,\ \mathrm{FIT} = 4.068,$$
$$Q^2_{\mathrm{LOO}} = 0.819,\ Q^2_{\mathrm{L5O}} = 0.814 \quad \ldots (1)$$

Values in parentheses represent the standard errors of the regression coefficients. Along with conventional statistical parameters (r, s, F, Q^2_{LOO}, and Q^2_{L5O}), additional indices, including FIT, AIC, and LOF, were employed to evaluate model robustness and predictive reliability.

The FIT function[46,47] is a model assessment metric closely related to the F-statistic but exhibits reduced sensitivity to the number of descriptors (k) when k is small, and increased sensitivity as k increases. Consequently, models with higher FIT values are considered superior. Akaike's Information Criterion (AIC)[48,49] evaluates model goodness of fit while penalizing excessive parameterization; lower AIC values indicate more parsimonious and informative models. Friedman's Lack-of-Fit (LOF)[50] factor also accounts for model complexity and is unbiased toward larger equations; statistically sound models yield lower LOF values. In comparative QSAR studies involving heterogeneous descriptor classes, these indices are essential for objective model selection.

External Validation

To assess external predictivity, structurally diverse compounds spanning a wide range of $p\mathrm{K_i}$ values were selected to construct test sets. Six independent test sets (14 compounds each) were generated *via* random selection, while the remaining 55 compounds constituted the training sets. All resulting models demonstrated acceptable predictive performance, with r^2_{Test} values exceeding 0.5. The test set yielding the highest r^2_{Test} comprised compounds 5, 16, 17, 22, 27, 30, 35, 39, 40, 51, 53, 59, 60, and 68 (Table 1).

Models incorporating one, two, and three descriptors showed gradual improvements in statistical quality. A marked enhancement was observed for four-descriptor models when the $\bar{r}$ threshold was set at ≥ 0.911. Five statistically significant models containing a total of 11 unique descriptors were obtained and are presented in ascending order of significance as Equations (2)–(6).

pK_i (M) = 8.845 -0.655(0.144)X2A
+0.820(0.142)PCR -2.904(0.191)BELe4
-1.387(0.160)GATS2v
n = 55, r = 0.928, s = 0.282, F(4, 50) = 77.404,
AIC = 0.095, LOF = 0.099, FIT = 4.361,
Q^2_{LOO} = 0.833, Q^2_{L5O} = 0.828, r^2_{Test} = 0.633 … (2)

pK_i (M) = 8.782 -0.756(0.160)Jhete
+0.921(0.138)PCR -2.637(0.186)BELe4
-1.5663(0.160)GATS2v
n = 55, r = 0.929, s = 0.279, F(4, 50) = 79.412,
AIC = 0.093, LOF = 0.097, FIT = 4.474,
Q^2_{LOO} = 0.839, Q^2_{L5O} = 0.836, r^2_{Test} = 0.640 … (3)

pK_i (M) = 8.255 -1.322(0.158)Jhetv
-1.951(0.190)BELe4 +1.120(0.160)MATS6p
-1.433(0.244)GATS8e
n = 55, r = 0.930, s = 0.277, F(4, 50) = 80.108,
AIC = 0.092, LOF = 0.096, FIT = 4.513,
Q^2_{LOO} = 0.843, Q^2_{L5O} = 0.843, r^2_{Test} = 0.656 … (4)

pK_i (M) = 7.908 -1.050(0.156)Jhetp
-1.586(0.194)BELe4 +1.118(0.170)MATS6m
-1.438(0.242)GATS8e
n = 55, r = 0.931 s = 0.276, F(4, 50) = 80.924,
AIC = 0.092, LOF = 0.095, FIT = 4.559,
Q^2_{LOO} = 0.840, Q^2_{L5O} = 0.840, r^2_{Test} = 0.737 … (5)

pK_i (M) = 7.788 -1.387(0.200)BELe4
+0.984(0.173)MATS6m -1.172(0.174)GATS7e
-1.577(0.245)GATS8e
n = 55, r = 0.931, s = 0.276, F(4, 50) = 80.960,
AIC = 0.092, LOF = 0.095, FIT = 4.561,
Q^2_{LOO} = 0.838, Q^2_{L5O} = 0.842, r^2_{Test} = 0.705 … (6)

All reported models exhibit F-values significant at the 99% confidence level [$F_{4,50}(0.01)$ = 3.738], and the standard errors associated with the regression coefficients are significant at the 95% level or higher. Q^2_{LOO} and Q^2_{L5O} values above 0.5 confirm strong internal robustness, while r^2_{Test}> 0.5 validates the external predictive capability of the models. The sign of each regression coefficient reflects the direction of influence of the corresponding descriptor on biological activity.

Table 2 summarizes the descriptor classes, definitions, average regression coefficients, and frequency of occurrence across the five models. Within the TOPO class, PCR shows a positive contribution, whereas Jhetv, Jhete, Jhetp, and X2A show negative contributions. PCR reflects the ratio of multiple path counts to total path counts, while the Jhet descriptors represent Balaban-type indices derived from weighted distance matrices based on van der Waals volume, electronegativity, and polarizability. X2A corresponds to the average connectivity index.

In the BCUT class, BELe4 represents the fourth lowest eigenvalue of the Burden matrix weighted by Sanderson electronegativities and consistently exhibits a negative contribution to activity.

Descriptors from the 2DAUTO class include MATS6m and MATS6p (Moran autocorrelation at lag 6, weighted by atomic mass and polarizability), both of which contribute positively. In contrast, GATS2v, GATS7e, and GATS8e (Geary autocorrelations weighted by van der Waals volume and electronegativity) exert negative effects on activity.

Among Equations (2)–(6), Equation (6) emerged as the most statistically significant and predictive model. It explains 87% of the variance in observed activity (r^2 = 0.867) and exhibits optimal values of FIT, AIC, LOF, and r^2_{Test}. This model was therefore used to predict pK_i values for all 69 compounds (Table 1). The close agreement between observed and calculated values is evident from Table 1 and Fig. 2.

According to Equation (6), enhanced activity is associated with higher values of MATS6m and lower values of BELe4, GATS7e, and GATS8e, providing valuable guidance for the rational design of new analogues.

Partial Least Squares (PLS) analysis was performed using the 11 descriptors identified from CP-MLR. After autoscaling (zero mean, unit variance), two optimal components were selected *via* cross-validation, explaining 87% of the variance in activity (r^2 = 0.872). The corresponding PLS equations and MLR-like coefficients are presented in Table 3. Predicted activities closely matched experimental values for both training and test sets (Table 1), with a near-linear relationship illustrated in

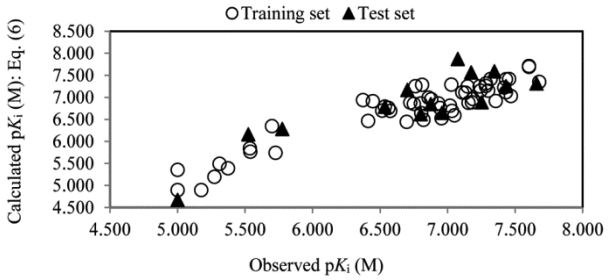

Fig. 2 — Plot of observed *versus* calculated pK_i (M) values (Equation 6) for inhibition of the *At*HPPD for training and test sets compounds

Table 3 — PLS and MLR-like PLS equations derived from the descriptors of CP-MLR identified models for *At*HPPD inhibition activity

A: PLS equation

Components	Coefficient (s. e.)[a]
TA	–0.347 (0.020)
TB	0.161 (0.023)
Constant	6.726

B: PLS regression statistics

Symbol	Estimate	Symbol	Estimate
n	55	LOF	0.077
r	0.934	FIT	6.056
s	0.264	Q^2_{LOO}	0.862
F	178.639	Q^2_{L5O}	0.862
AIC	0.078	r^2_{Test}	0.827

C: MLR-like PLS equation

S.No.	Descriptor	MLR-like coefficient[b]	F. C. (Order)[b]	S.No.	Descriptor	MLR-like coefficient[b]	F. C.(Order)[b]
1	Jhetv	–0.255	–0.050 (9)	7	MATS6m	0.783	0.146 (3)
2	Jhete	–0.115	–022 (11)	8	MATS6p	0.385	0.075 (5)
3	Jhetp	–0.306	–060 (7)	9	GATS2v	–0.339	–0.066 (6)
4	X2A	0.115	0.025 (10)	10	GATS7e	–0.545	–0.102 (4)
5	PCR	0.265	0.058 (8)	11	GATS8e	–1.276	–0.166 (2)
6	BELe4	–1.388	–0.230 (1)	–	Constant	7.719	–

[a]Regression coefficient of the PLS component and its standard error. [b]Coefficient of MLR-like PLS equation in terms of descriptors for their original values; F. C. is the fraction contribution of the regression coefficient computed from the normalized regression coefficients obtained from the auto-scaled (zero mean and unit standard deviation) data

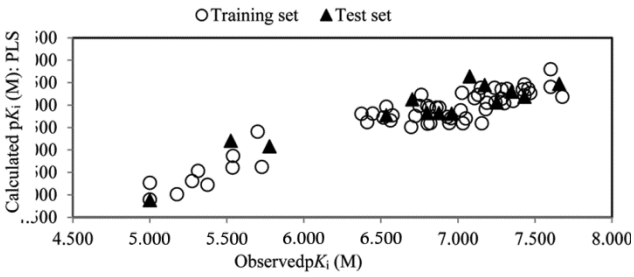

Fig. 3 — Plot of observed *versus* calculated pK_i (M) values (PLS) for inhibition of the AtHPPD for training and test sets compounds

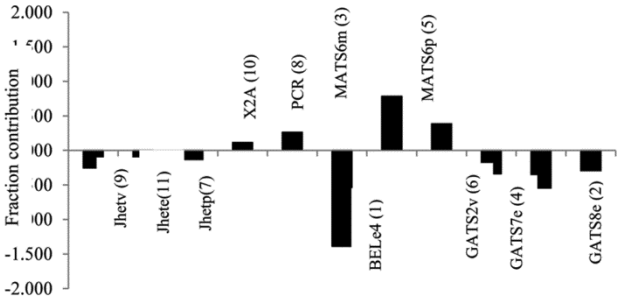

Fig. 4 — Plot of fraction contribution of MLR-like PLS coefficients (normalized) against 11 identified descriptors (Table 3) associated with the *At*HPPD inhibition activities of the compounds

Fig. 3. Fig. 4 depicts the fractional contributions of normalized regression coefficients, highlighting the relative importance of each descriptor. Descriptor ranking in Table 3 further elucidates their influence on herbicidal activity.

The applicability domain (AD) of Equation (1) was assessed using a Williams plot (Fig. 5), which plots standardized residuals against leverage (h_i). As mentioned earlier, the leverage threshold was calculated as $h^* = 3(k + 1)/n$. All compounds were found to lie within the defined AD, confirming the reliability of predictions for both training and test sets.

Mesotrione is a potent herbicide that inhibits the enzyme HPPD and is commonly used in agriculture. As a triketone HPPD herbicide, it contains 1,3-dicarbonyl and aromatic groups, both of which are critical for its activity. Recent studies on modified compounds, particularly arylthioacetic acid derivatives, have identified compounds 11, 40, and 49 (Table 1), featuring cyclohexenone and pyrazolyl

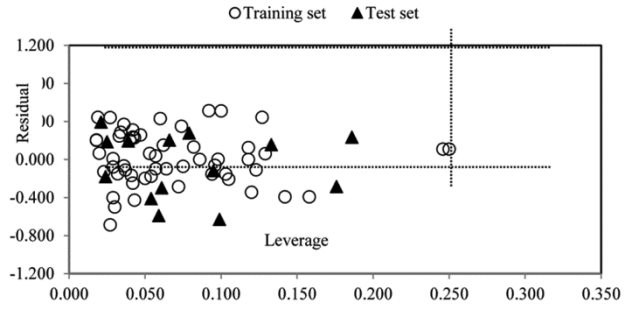

Fig. 5 — Williams plot for the whole dataset for *At*HPPD inhibition activities of the title compounds, listed in Table 1 (leverage threshold $h^* = 0.273$ and residual limits = ±0.828)

groups, as the most active analogues, with pK_i values similar to mesotrione. A comparison of the physicochemical properties and toxicity of compounds 11, 40, and 49 with mesotrione is provided in Table 4.

Table 4 — Drug-likeness and toxicity-related properties of mesotrione, and the most active compounds obtained from SwissADME and preADME-T servers

No.[a]	Property		Mesotrione[b]	Most active compound[c] 11	40	49
1	Experimental pK_i (M)		7.699	7.678	7.658	7.602
2	Molecular weight (MW) $gmol^{-1}$		339.32	307.32	341.77	366.65
3	Fraction Csp^3		0.36	0.29	0.23	0.31
4	No. of rotatable bonds (NRB)		4	5	5	4
5	No. of hydrogen bond acceptors (NHA)		7	5	5	4
6	No. of hydrogen bond donors (NHD)		0	1	1	1
7	Molar Refractivity (MR) m^3mol^{-1}		80.02	109.98	85.75	84.02
8	Topological polar surface area (TPSA) $Å^2$		125.49	103.07	126.24	92.56
9	Partition coefficient ($iLogP_{o/w}$)		0.49	1.79	2.14	2.83
10	Lipinski rule followed (violation)		yes (0)	yes (0)	yes (0)	yes (0)
11	Inhibition of cytochrome P450 subtypes	CYP 2C19	yes	yes	no	yes
		CYP 2D6	no	no	no	no
		CYP 3A4	yes	yes	no	yes
12	Human intestinal absorption (HIA %)		74.51	84.83	92.63	96.88
13	Plasma protein binding (PPB %)		83.77	89.79	100	91.86
14	Carcinogenicity	Mouse	negative	negative	negative	positive
		Rat	negative	negative	positive	negative

[a]Properties numbered 1 to 11 were obtained from SwissADME, and properties 12 to 14 were acquired from PreADMET servers, [b]For drug-likeness property, the predicted compound should have MW < 500, NRB ≤ 10, $iLogP_{o/w}$ < 5, NHA ≤ 10, NHD ≤ 5,TPSA ≤ 140 $Å^2$, Fraction Csp^3 ≤ 0.45, MR between 40-130, HIA: 0-20% poorly absorbed; 21-70% moderately absorbed; and 71-100% strongly absorbed, PPB: > 90% strongly bound, and < 90% weakly bound, carcinogenicity: + carcinogenic, and – non-carcinogenic, [c]For most active compound number, see Table 1

The initial eleven properties listed in Table 4 for these compounds and mesotrione comply with the Lipinski rule of five, as none of the specified properties is violated.

The data on HIA% indicates that the human intestine tends to absorb these three compounds similarly. Regarding PPB%, mesotrione exhibits a weak affinity for plasma proteins, whereas compounds 11, 40, and 49 show a strong affinity.

The findings regarding toxicity indicate that, except for compound 49, the other three compounds are non-carcinogenic in mice. However, compound 40 may have carcinogenic potential, whereas compounds 11, 49, and mesotrione are considered non-carcinogenic in rat models.

Cytochrome P450 enzymes (CYP) are a diverse group of heme-containing proteins essential for metabolising various substances, including drugs, steroids, fatty acids, and toxins. These enzymes are classified into families and subfamilies based on similarities in their amino acid sequences. In humans, there are 18 families of CYP enzymes, encompassing 57 identified genes that encode active enzymes.

Among these, six CYP enzymes are particularly notable in drug metabolism: CYP1A2, which processes various drugs and toxins, including caffeine and certain antidepressants; CYP2C9, which metabolizes a wide range of drugs such as warfarin, ibuprofen, and phenytoin; CYP2C19, responsible for processing medications like clopidogrel and some proton pump inhibitors; CYP2D6, which metabolizes numerous drugs including antidepressants, antipsychotics, and beta-blockers; CYP2E1, involved in the metabolism of alcohol and certain toxins; and CYP3A4, the most abundant enzyme in the liver, which is responsible for metabolizing a diverse array of drugs including statins, calcium channel blockers, and benzodiazepines. Additionally, notable CYP enzymes include CYP1A1, significant in the metabolism of certain procarcinogens; CYP51A1, involved in cholesterol biosynthesis; and CYP7A1, a key enzyme in bile acid biosynthesis.

Drugs that inhibit CYP enzymes can reduce their metabolic activity, leading to increased plasma levels of concurrently administered CYP enzyme substrates. This may result in drug toxicity or adverse reactions. Therefore, understanding the activity and specificity of CYP enzymes is critical for drug development, personalised medicine, and toxicity assessment. Table 4 lists three key CYP enzymes and indicates

whether the specified compound may act as an inhibitor or a non-inhibitor.

Molecular Docking

The three most active compounds (11, 40, and 49; Table 1), along with the reference herbicide mesotrione, were selected for molecular docking studies to explore their binding behavior at the target site. Docking simulations were carried out using several freely available computational tools. The two-dimensional chemical structures of all compounds (ligands) were drawn using ChemDraw and subsequently converted into SMILES format. These SMILES strings were used as inputs to OpenBabel[51] to generate the required three-dimensional ligand files, designated Lig-2, Lig-3, and Lig-4, while mesotrione was designated Lig-1.

The three-dimensional crystallographic structure of the target protein was retrieved from the Protein Data Bank (PDB)[51,52] using the accession code 5YWG. The downloaded structure comprised two protein chains (A and B) and included non-standard components such as crystallographic water molecules, a cobalt ion [Co(II)], and the co-crystallized reference ligand mesotrione.

Protein preparation for docking was performed using ChimeraX and MGLTools software[53,54]. Initially, all water molecules were removed from the structure. Subsequently, chain B, along with its associated non-standard residues, was deleted, retaining only chain A for further analysis. This prepared structure, which still contained the cobalt ion and the bound reference ligand, was labeled R_Co_Lig-1. To generate the receptor-only model, the reference ligand (Lig-1) was removed from this structure, yielding a second file named R_Co. Both receptor files were then corrected for missing atoms, assigned Kollman charges, and supplemented with polar hydrogens to ensure suitability for docking calculations.

The R_Co_Lig-1 file served two primary purposes. First, it was used to define the grid-box parameters by identifying the center coordinates and dimensions of the binding cavity based on the original position of the co-crystallized mesotrione. Second, it enabled validation of the docking protocol by calculating the root-mean-square deviation (RMSD) between the experimentally observed and re-docked poses of Lig-1. In this case, the RMSD value is zero because the derived coordinates and grid box size from the reference ligand were identical and reused for docking Lig-1. These validated grid-box parameters were subsequently applied for docking all ligands using AutoDock Vina[55]. For each ligand, nine binding conformations were generated, and the most favorable pose was selected based on the lowest binding energy (*i.e.*, the most negative value expressed in kcal/mol).

The receptor file R_Co was then combined with the best-ranked conformer of each ligand using PyMol software[56], resulting in four receptor–ligand complexes. These complexes were analyzed for intermolecular interactions using the Protein–Ligand Interaction Profiler (PLIP) web server[57]. The identified interactions are summarized in Table 5, while the corresponding three-dimensional interaction patterns are illustrated in Fig. 6.

All ligands exhibited favorable binding affinities, indicating stable accommodation within the catalytic pocket coordinated by the essential Co(II) ion. The reference ligand mesotrione (Lig-1) demonstrated strong binding affinity (−8.280 kcal/mol), forming a tetrahedral metal coordination with Co(II) supported by residues HIS226, HIS308, and GLU394. This coordination was further stabilized by multiple hydrophobic interactions and π–π stacking with PHE424.

Lig-2 (compound 11) showed a comparable binding affinity (−7.945 kcal/mol), maintaining tetrahedral coordination with Co(II) and forming an additional hydrogen bond with ASN282, which likely contributes to its stabilization despite the slightly reduced affinity. Lig-3 (compound 40) exhibited the weakest binding (−7.492 kcal/mol), which correlated with a shift in metal coordination geometry toward trigonal pyramidal and a reduced number of hydrophobic interactions. Nevertheless, its binding was partially compensated by multiple hydrogen bonds with ASN282 and π–π stacking with PHE381.

Notably, Lig-4 (compound 49) displayed a binding affinity (−8.263 kcal/mol) nearly identical to that of mesotrione. This strong interaction was supported by optimal tetrahedral coordination with Co(II), extensive hydrophobic contacts, dual hydrogen bonding with ASN282, π–π stacking with PHE381, and a distinctive halogen bond involving GLY420. The synergistic contribution of these interactions suggests enhanced complementarity within the active site.

Overall, the docking results underscore the importance of Co(II) coordination, hydrogen bonding

Table 5 — Interactions shown by the mesotrione (reference), and the most active compounds 11, 40, and 49 (for compound number, see Table 1) docked into the active site of the receptor protein, 5YWG

Compd	Binding affinity (kcal/mol)	Type of interaction and species involved		Distance (Å)	Additional specification
Mesotrione	–8.280	Metallic	Co(II)–Mesotrione	2.19	Tetrahedral
			Co(II)–HIS226	2.50	
			Co(II)–HIS308	2.37	
			Co(II)–GLU394	2.07	
		Hydrophobic	VAL228–Mesotrione	3.89	
			VAL228–Mesotrione	3.77	
			PRO280–Mesotrione	3.13	
			PHE381–Mesotrione	3.53	
		π-Stacking	PHE424–Mesotrione	4.66	Center distance (angle 69.52°)
11	–7.945	Metallic	Co(II)–Compd 11	2.41	Tetrahedral
			Co(II)–HIS226	2.50	
			Co(II)–HIS308	2.37	
			Co(II)–GLU394	2.07	
		Hydrophobic	PRO280–Compd 11	3.57	
			PHE419–Compd 11	3.31	
			LYS421–Compd 11	3.80	
			PHE424–Compd 11	3.76	
			PHE424–Compd 11	3.56	
		Hydrogen-bond	ASN282–Compd 11	2.25	H-donor (Nam)
				3.04	H-acceptor (O2)
40	–7.492	Metallic	Co(II)–Compd 40	2.50	Trigonal pyramidal
			Co(II)–HIS226	2.37	
			Co(II)–GLU394	2.07	
		Hydrophobic	LYS421–Compd 40	3.83	
			PHE424–Compd 40	3.68	
		Hydrogen-bond	ASN282–Compd 40	2.27	H-donor (Nam)
				3.10	H-acceptor (O3)
			ASN282–Compd 40	2.15	H-donor (O3)
				3.08	H-acceptor (O2)
		π-Stacking	PHE381–Compd 40	4.01	Center distance (angle 17.17°)
49	–8.263	Metallic	Co(II)–Compd 49	2.46	Tetrahedral
			Co(II)–HIS226	2.50	
			Co(II)–HIS308	2.37	
			Co(II)–GLU394	2.07	
		Hydrophobic	PRO280–Compd 49	3.52	
			PHE419–Compd 49	3.37	
			LYS421–Compd 49	3.97	
		Hydrogen-bond	ASN282–Compd 49	2.34	H-donor (Nam) angle 128.59°
				3.08	H-acceptor (O2)
			ASN282–Compd 49	2.56	H-donor (O2) angle 101.51°
				2.90	H-acceptor (O2)
		π-Stacking	PHE381–Compd 49	3.92	Center distance (angle 9.14°)
		Halogen-bond	GLY420–Compd 49	3.60	Donor (Cl; angle 46.73°) Acceptor (O2; angle 130.08°)

with ASN282, and aromatic stacking interactions in stabilizing ligand binding. Among the tested compounds, Lig-4 (compound 49) emerged as the most promising analogue, exhibiting binding characteristics comparable to the reference ligand mesotrione.

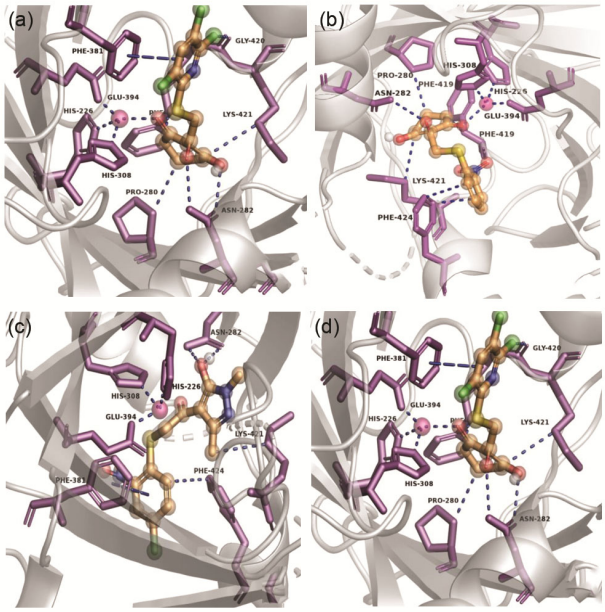

Fig. 6 — The reference compound mesotrione (A), the most active compounds, 11 (B), 40 (C), and 49 (D), showing interactions with the amino acids of the receptor protein 5YWG

Conclusion

The inhibitory activities of arylthioacetic acid derivatives against *Arabidopsis thaliana* 4-hydroxyphenylpyruvate dioxygenase (*At*HPPD) were successfully analyzed using QSAR modeling. Five statistically robust models based on 11 significant descriptors were developed, providing insight into the structural factors governing inhibition. Positive contributions to activity were associated with PCR and Moran autocorrelation descriptors (MATS6m and MATS6p), whereas Balaban-type indices (Jhetv, Jhete, Jhetp), BELe4, X2A, and Geary autocorrelation descriptors (GATS2v, GATS7e, GATS8e) showed negative effects.

All models demonstrated strong internal robustness and satisfactory external predictive performance, supported by leave-one-out and leave-five-out validation, as well as applicability domain analysis. Partial least squares analysis further confirmed the relevance of the selected descriptors, and the final model reliably predicted the inhibition activities of all compounds in the dataset.

ADME-T evaluation of the most active compounds (11, 40, and 49) revealed compliance with Lipinski's rule of five and intestinal absorption profiles comparable to mesotrione, although stronger plasma protein binding was observed. Toxicity predictions suggested generally favorable safety profiles, with compound 49 showing the most balanced characteristics.

Molecular docking highlighted the importance of Co(II) coordination, hydrogen bonding with ASN282, and aromatic stacking interactions for ligand stabilization. Among the tested analogues, compound 49 showed binding and affinity comparable to those of mesotrione, making it a promising lead for the further development of potent *At*HPPD inhibitors.

Acknowledgement

The author gratefully acknowledges the research facilities provided by the institution and appreciates the assistance of his colleagues.

References

1 Duke S O, *Pest Manage Sci*,68 (2012) 505.
2 Ruegg W T, Quadranti M & Zoschke A, *Weed Res*, 47 (2007) 271.
3 Ahrens H, Lange G, Müller T, Rosinger C, Willms L & Almsick A, *Angew Chem Int Ed Engl*,52 (2013) 9388.
4 Moran G R, *Arch Biochem Biophys*,433 (2005) 117.
5 Xu K, Racine F, He Z & Juneau P, *Environ Pollut*, 244 (2019) 295.
6 Ndikuryayo F, Moosavi B, Yang W C & Yang G F, *J Agric Food Chem*,65 (2017) 8523.
7 Lin H Y, Yang J F, Wang D W, Hao G F, Dong J Q, Wang Y X, Yang W C, Wu J W, Zhan C G & Yang G F, *The FEBS J*, 286 (2019) 975.
8 Wu C S, Huang J L, Sun Y S & Yang D Y, *J Med Chem*,45 (2002) 2222.
9 Brownlee J M, Johnson W K, Harrison D T & Moran G R, *Biochem*,43, (2004) 6370.
10 Kovaleva E G & Lipscomb J D, *Nat Chem Biol*,4 (2008) 186.
11 Grossmann K & Ehrhardt T, *Pest Manage Sci*,63 (2007) 429.
12 Raspail C, Graindorge M, Moreau Y, Crouzy S, Lefèbvre B, Robin A Y Dumas R & Matringe M, *J Biol Chem*, 286 (2011) 26061.
13 Mitchell G, Bartlett D W, Fraser T E, Hawkes T R, Holt D C, Townson J K &Wichert R A, *Pest Manage Sci*, 57 (2001) 120.
14 Yamaoka K, Nakagawa M & Ishida M, *J Pestic Sci*, 12 (1987) 209.
15 Wang D W, Lin H Y, He B, Wu F X, Chen T, Chen Q, Yang W C & Yang G F, *J Agric Food Chem*, 64 (2016) 8986.
16 Fu Y, Zhang S Q, Liu Y X, Wang J Y, Gao S, Zhao L X, & Ye F, *Ind Crop Prod*, 137 (2019) 566.
17 He B, Dong J, Lin H Y, Wang M Y, Li X K, Zheng B F, Zheng B F, Chen Q, Hao G F, Yang W C & Yang G F, *J Agric Food Chem*, 67 (2019) 10844.
18 Liu X H, Qiao L, Zhai Z W, Cai P P, Charles L C, Tan C X, Weng J Q, Han L & Wu H K, *Pest Manag Sci*,75 (2019) 2892.
19 Fu Y, Zhang D, Zhang S Q, Liu Y X, Guo Y Y, Wang M X, Gao S, Zhao L X & Ye F, *J Agric Food Chem*,67, (2019) 11839.
20 Ye F, Zhai Y, Guo K L, Liu Y X, Li N, Gao S, Zhao L X & Fu Y, *J Agric Food Chem*, 67 (2019) 11568.

21 Wang M M, Huang H, Shu L, Liu J M, Zhang J Q, Yan Y L & Zhang D Y, *Beilstein J Org Chem*, 16 (2020) 233.
22 Fu Q, Cai P P, Cheng L, Zhong L K, Tan C X, Shen Z H, Han L, Xu T M & Liu X H, *Pest Manage Sci*, 76 (2020) 868.
23 He B, Wu X F, Yu L K, Wu L, Chen Q, Hao G F, Yang W C, Lin H Y & Yang G F, *J Agric Food Chem*, 68 (2020) 5059.
24 Zhang Y Y, Gao S, Liu Y X, Wang C, Jiang W, Zhao L X, Fu Y & Ye F, *J Agric Food Chem*, 68 (2020) 3403.
25 Zhao L X, Jiang M J, Hu J J, Zou Y L, Cheng Y, Ren T, Gao S, Fu Y & Ye F, *J Agric Food Chem*, 68 (2020) 3729.
26 Huang H, Wang M M, Shu L, Yan Y L, Zhang J Q, Liu J M, Zhan X & Zhang D Y, *Pest Manage Sci*, 76 (2020) 4112.
27 Singh P & Shekhawat N, *Ind J Biochem Biophys*, 62 (2025) 882.
28 Chemdraw ultra 6.0 and Chem3D ultra, Cambridge Soft Corporation, Cambridge, USA. Available at:http://www.camsoft.com.
29 DRAGON software version, 3.0-2003 by Todeschini R, Consonni V, Mauri A, Pavan M, Milano, Italy. Available at: http//disat.unimib.it/chm/Dragon.htm.
30 Golbraikh A & Tropsha A, *J Mol Graph Model*, 20 (2002) 269.
31 Prabhakar Y S, *QSAR Comb Sci*, 22 (2003) 583.
32 Sharma S, Sharma B K, Sharma S K, Singh P & Prabhakar Y S, *Eur J Med Chem*, 44 (2009) 1377.
33 Sharma B K, Pilania P, Singh P & Prabhakar Y S, *SAR QSAR Environ Res*, 21 (2010) 169.
34 Sharma B K, Singh P, Sarbhai K & Prabhakar Y S, *SAR QSAR Environ Res*, 21 (2010) 369.
35 Sharma B K, Pilania P, Sarbhai K, Singh P & Prabhakar Y S, *Mol Divers*, 14 (2010) 371.
36 Sharma B K, Singh P, Shekhawat M, Sarbhai K & Prabhakar Y S, *SAR QSAR Environ Res*, 22 (2011) 365.
37 So S -S & Karplus M, *J Med Chem*, 40 (1997) 4347.
38 Prabhakar Y S, Solomon V R, Rawal R K, Gupta M K & Katti S B, *QSAR Combn Sci*, 23 (2004) 234.
39 Wold S, *Technomet*, 20 (1978) 397.
40 Kettaneh N, Berglund A & Wold S, *Comput Stat Data Anal*, 48 (2005) 69.
41 Stahle L, Wold S, P*rogress in Medicinal Chemistry* 25 (1988) 291.
42 Gramatica P, *QSAR Comb Sci*, 26 (2007) 694.
43 Eriksson L, Jaworska J, Worth A P, Cronin M T D, McDowell R M & Gramatica P, *Env Health Persp*, 111 (2003) 1361.
44 Morris G M, Huey R, Lindstrom W, Sanner M F, Belew R K, Goodsell D S & Olson A J, *J Comput Chem*, 30 (2009) 2785.
45 Morris G M, Goodsell D S, Halliday R S, Huey R, Hart W E, Belew R K & Olson A J, *J Comput Chem*, 19 (1998)1639.
46 Kubinyi H, *Quant Struct-Act Relat*, 13 (1994) 285.
47 Kubinyi H, *Quant Struct-Act Relat*, 13 (1994) 393.
48 Akaike H, in *Second International Symposium on Information Theory*, edited by Petrov B N, Csaki F (Budapest, Akademia Kiado), 1973, pp. 267-281.
49 Akaike H, *IEEE Trans Automat Contr* AC-19 (1974) 716.
50 Friedman J, *Technical Report No. 102*, Laboratory for Computational Statistics, Stanford (Stanford University), August 1990.
51 O'Boyle N M, Banck M, James C A, Morley C, Vandermeersch T & Hutchison G R, *J Cheminform*, 3 (2011) 33.
52 Berman H M, Westbrook J, Feng Z, Gilliland G, Bhat T N, Weissig H, Shindyalov I N & Bourne P E,*Nucleic Acids Res*, 28 (2000) 235.
53 Pettersen E F, Goddard T D, Huang C C, Meng E C, Couch G S, Croll T I, Morris J H& Ferrin T E,*Protein Sci*, 30 (2021) 70.
54 Morris G M, Huey R, Lindstrom W, Sanner M F, Belew R K, Goodsell D S & Olson A J, *J Comput Chem*, 30 (2009) 2785.
55 Trott O & Olson A J,*J Comput Chem*, 31 (2010) 455.
56 Schrödinger LLC, (2025). *The PyMOL Molecular Graphics System*, *Version 3.1*.Schrödinger, LLC, New York, NY, USA. Available at: http://www.pymol.org/pymol.
57 Salentin S, Schreiber S, Haupt V J,Adasme M F & Schroeder M, *Nucleic Acids Res*, 43 (2015) W443.