Exploring Regression-based QSTR and i-QSTR Modeling for Ecotoxicity Prediction of Diverse Pesticides on Multiple Avian Species

Trina Podder[1], Ankur Kumar[1], Arnab Bhattacharjee[1], Probir Kumar Ojha[1]*

[1]Drug Discovery and Development (DDD) Laboratory,

Department of Pharmaceutical Technology, Jadavpur University,

Kolkata 700032, India

*Corresponding author: P K Ojha, Email: probirojha@yahoo.co.in, Phone: +91 8777677004;

Materials and methods

Definition of different statistical parameters

$$R^2 = 1 - \frac{\Sigma\left(Y_{obs(train)} - Y_{calc(train)}\right)^2}{\Sigma\left(Y_{obs(train)} - \overline{Y}_{train}\right)^2}$$

$$R_{adj}^2 = \frac{(n-1) \times R^2 - p}{(n-p-1)}$$

$$RMSE_c = \sqrt{\frac{\left(\Sigma Y_{obs(train)} - Y_{calc(train)}\right)^2}{n}}$$

$$Q_{LOO}^2 = 1 - \frac{\Sigma\left(Y_{obs(train)} - Y_{calc(train)}\right)^2}{\Sigma\left(Y_{obs(train)} - \overline{Y}_{train}\right)^2}$$

Where, $R^2 =$ Co-efficient of determination

$R_{adj}^2 =$ Adjusted co-efficient of determination

$Q_{LOO}^2 =$ Cross-validated correlation coefficient

$Y_{obs(train)} =$ Observed response value of training set

$Y_{calc(train)} =$ Calculated response value of training set

$\overline{Y}_{train}=$ Average of all response of training set

$RMSE_c =$ Root mean square errors of calibration

$n =$ Number of compounds

$p =$ Number of descriptors

$$Q_{F1}^2 = 1 - \frac{\Sigma(Y_{obs(test)} - Y_{calc(test)})^2}{\Sigma(Y_{obs(test)} - \overline{Y}_{train})^2}$$

$$Q_{F2}^2 = 1 - \frac{\Sigma(Y_{obs(test)} - Y_{calc(test)})^2}{\Sigma(Y_{obs(test)} - \overline{Y}_{test})^2}$$

$$r_m^2 = r^2 \times \left(1 - \sqrt{(r^2 - r_0^2)}\right)$$

$$r'^2_m = r'^2 \times \left(1 - \sqrt{(r'^2 - r'^2_0)}\right)$$

$$\overline{r}_m^2 = \frac{(r_m^2 + r'^2_m)}{2}$$

$$\Delta r_m^2 = \left| r_m^2 - r'^2_m \right|$$

Where, $Y_{obs(test)} =$ Observed response value of training set

$Y_{calc(test)} =$ Calculated response value of training set

$\overline{Y}_{train}=$ Average of all response of training set

$\overline{Y}_{test}=$ Average of all response of test set

Applicability domain: Standardization technique

The equation to calculate AD is:

$$S_{ki} = \frac{|X_{ki} - \overline{X}_i|}{\sigma X_i}$$

Where, k=1, 2, 3 … nComp (here, nComp = total number of compounds)

i= 1, 2, 3 … nDes (here, nDes = total number of descriptors)

S_{ki}= Standardized descriptor i for compound k (from the training or test set)

X_{ki} = original descriptor i for compound k (from the training or test set)

$\overline{X_{ki}}$ = mean value of the descriptor

X_i = for the training set compounds only

σX_i = standard deviation of the descriptor
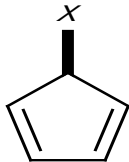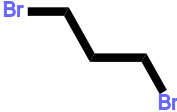
Xi for the training set compounds only

The standardization approach of applicability domain is based on the ideal data distribution; 99.7% of the compounds would stay within the range of mean ± 3 standard deviations (SDs). As a result, this range (i.e., mean ± 3SDs) is considered as the area of the majority of the training set compounds. Outside this area, a compound is examined as a diverse from the rest of the compounds. So, one should compute the maximum Si(k) value ([Si]max(k)) for the compound k. If the SD value for descriptor i of compound k (Ski) is greater than 3 then the compound is an X-outlier (if it is in the training set) or outside the AD (if it is in the test set).

Results and discussion

*Here, S denotes supporting information and number denotes to corresponding table.

Results and discussion

Table S1. Mechanistic interpretation and definition of descriptors related to **electronegativity**: BQ (Bobwhite quail) oral pLD_{50}

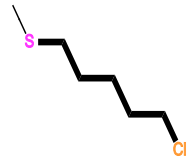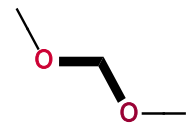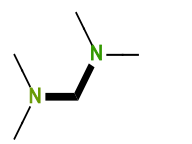| Sl no	Descriptors	Class	Definition of descriptors	Contribution	Fragments	Mechanistic interpretation
1	**P-117**	Atom-centered fragments	Presence of phosphate group	Positive	X3-P=X (X=O)	Presence of phosphate group makes pesticides more toxic
2	**nCXr**	Functional group count	Number of electro-negative atom (X) on ring C (sp3)	Positive	(X= Cl, O, S, F etc.)	Presence of this fragment makes pesticides more toxic
3	**F04[Br-Br]**	2D atom pair	Frequency of Br-Br at topological distance 4	Positive		Presence of bromine atoms make pesticides more toxic

4	**F06[S-Cl]**	2D atom pair	Frequency of S-Cl at topological distance 6	Positive		Presence of this fragment makes pesticides more toxic
5	**F02[O-O]**	2D atom pair	Frequency of O-O at topological distance 2	Positive		Presence of oxygen atoms make pesticides more toxic

Table S2. Mechanistic interpretation and definition of descriptors related to **hydrophilicity:** BQ (Bobwhite quail) oral pLD_{50}

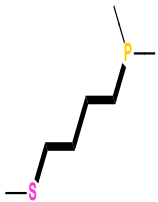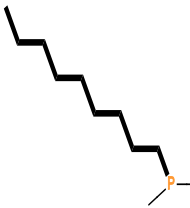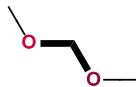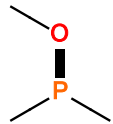| Sl no	Descriptors	Class	Definition of descriptors	Contribution	Fragments	Mechanistic interpretation
1	**B02 [N-N]**	2D atom pair	Presence/absence of N - N at topological distance 2	Negative		Attachment of this fragment in ring system makes pesticides non-toxic

2	**F05[S-P]**	2D atom pair	Frequency of S-P at topological distance 5	Negative		Presence of this fragment makes pesticides non-toxic
3	**F09[C-P]**	2D atom pair	Frequency of C-P at topological distance 9	Negative		Presence of this fragment makes pesticides non-toxic

Table S3. Mechanistic interpretation and definition of descriptors related to **hydrophobicity:** BQ (Bobwhite quail) oral pLD_{50}

Sl no	Descriptors	Class	Definition of descriptors	Contribution	Fragments	Mechanistic interpretation
1	**X2A**	Connectivity indices	Average connectivity index of order 2	Negative	Size & Shape	Increasing value of this descriptor makes the pesticides non-toxic
2	**minssCH2**	Minimum atom-type E-state	Presence of methylene ($-CH_2-$) groups in aliphatic chains	Negative	$-CH_2-$	Increasing value of this descriptor makes the pesticides non-toxic
3	**X4v**	Connectivity indices	Valence connectivity index of order	Positive	Size & Shape	Increasing value of this descriptor makes the pesticides toxic

Table S4. Mechanistic interpretation and definition of descriptors related to **electronegativity:** MD (Mallard duck) oral pLD_{50}

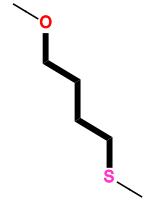| Sl no	Descriptors	Class	Definition of descriptors	Contribution	Fragments	Mechanistic interpretation
1	**nRSR**	Functional group count	Number of sulfide groups present in a compound	Positive	Number of Sulfur	Presence of sulfide group makes pesticides more toxic
2	**F02[O-O]**	2D atom pair	Frequency of O-O at topological distance 2	Positive		Presence of this fragment makes pesticides more toxic
3	**B01[O-P]**	2D atom pair	Presence/ Absence of O−P at topological distance 1	Positive		Presence of this group make pesticides more toxic

| 4 | **B05[O-S]** | 2D atom pair | Presence/ Absence of O–S at topological distance 5 | Positive | | Presence of this fragment makes pesticides more toxic |

Table S5. Mechanistic interpretation and definition of descriptors related to **hydrophilicity:** MD (Mallard duck) oral pLD_{50}

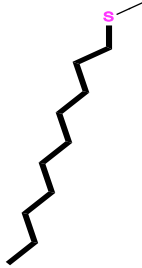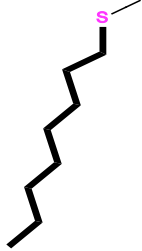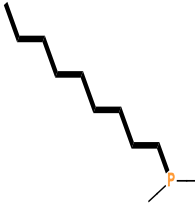| Sl no	Descriptors	Class	Definition of descriptors	Contribution	Fragments	Mechanistic interpretation

1	**F10[C-S]**	2D atom pair	Frequency of C-S at topological distance 10	Negative		Attachment of this fragment in ring system makes pesticides non-toxic
2	**F08[C-S]**	2D atom pair	Frequency of C-S at topological distance 8	Negative		Presence of this fragment makes pesticides non-toxic
3	**F06[C-P]**	2D atom pair	Frequency of C-P at topological distance 6	Negative		Presence of this fragment makes pesticides non-toxic

Table S6. Mechanistic interpretation and Definition of descriptors related to π–π **Electronegativity:** MD (Mallard duck) oral pLD_{50} oral pLD_{50}
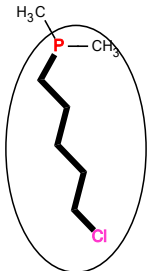
Sl no	Descriptors	Class	Definition of descriptors	Contribution	Fragments	Mechanistic interpretation
1	**T(P..Cl)**	2D atom pair	Sum of topological distances between phosphorus and chlorine	Negative		Attachment of this fragment in structure makes pesticides non-toxic

Table S7. Mechanistic interpretation and definition of descriptors related to **lipophilicity:** MD (Mallard duck) oral pLD_{50}

Sl no	Descriptors	Class	Definition of descriptors	Contribution	Fragments	Mechanistic interpretation
1	**nBridge-Head**	Ring-descriptor	How many bridgehead atoms are present in the ring structure	Positive	Hetero-atoms like $-PO_4$, $-SO_4$ etc.	Presence of this group makes pesticides more toxic

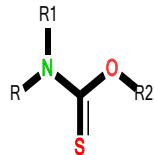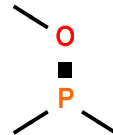|						
2	**nArOCON**	Functional group count	How many aromatic (thio-) carbamate groups are in a compound	Positive		Presence of this fragment makes pesticides more toxic

Table S8. Mechanistic interpretation and definition of descriptors related to **electronegativity:** ZF (Zebra finch) oral pLD_{50}

Sl no	Descriptors	Class	Definition of descriptors	Contribution	Fragments	Mechanistic interpretation
1	**F01[O-P]**	2D atom pair	Frequency of O−P at topological distance 1	Positive		Presence of this group makes pesticides more toxic
2	**nRSR**	Functional group count	Number of sulfide groups present in a compound	Positive	Number of sulfur	Presence of sulfur makes pesticides more toxic

3	**c-031**	Atom centered fragment	Two electro-negative atoms attached with carbon along with another group	Positive	X--CR— X(X=Cl,Br,F, O,S etc.)	Presence of this group make pesticides more toxic

Table S9. Mechanistic interpretation and definition of descriptors related to **hydrophilicity:** ZF (Zebra finch) oral pLD_{50}

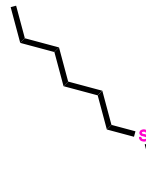| Sl no	Descriptors	Class	Definition of descriptors	Contribution	Fragments	Mechanistic interpretation
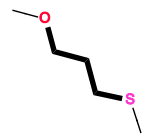
1	**F04[O-S]**	2D atom pair	Frequency of the O−S fragment in a compound at topological distance 4	Negative		Attachment of this fragment in ring system makes pesticides non-toxic
2	**B04[C-C]**	2D atom pair	Presence/absence of C−C at topological distance 4	Negative		Presence of this fragment makes pesticides non-toxic
3	**F06[C-S]**	2D atom pair	Frequency of C-S at topological distance 6	Negative		Presence of this fragment makes pesticides non-toxic
4	**F05[C-S]**	2D atom pair	Frequency of C-S at topological distance 5	Negative		Presence of this fragment makes pesticides non-toxic

Table S10. Mechanistic interpretation and definition of descriptors related to **π–π electronegativity:** ZF (Zebra finch) oral pLD$_{50}$

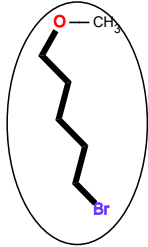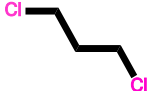| Sl no	Descriptors	Class	Definition of descriptors	Contribution	Fragments	Mechanistic interpretation
1	**T(O..Br)**	2D atom pair	Sum of topological distances between oxygen and bromine	Negative		Presence of this group makes pesticides non-toxic
2	**B04[Cl-Cl]**	2D atom pair	Presence/absence of Cl-Cl at topological distance 4	Negative		Presence of this fragment makes pesticides more non-toxic