chemistry proceedings

MDPI

Proceedings Paper

Interspecies Quantitative Structure-Toxicity-Toxicity Relationships for Predicting the Acute Toxicity of Organophosphorous Compounds †

Gheorghe Ilia, Ana Borota and Simona Funar-Timofei *

Coriolan Dragulescu Institute of Chemistry, Romanian Academy, Bul. Mihai Viteazu 24, 300223 Timisoara, Romania; e-mail1 (G.I.); e-mail2 (A.B.)

* Correspondence: timofei@acad-icht.tm.edu.ro; Tel.: +40-256-491818

† Presented at the 25th International Electronic Conference on Synthetic Organic Chemistry, 1–30 November 2021, https://ecsoc-25.sciforum.net/.

Abstract: Median lethal concentration values are commonly used to express the relative risk related to the acute toxicity of chemicals. In this paper, we considered rat and mouse acute toxicity (LD_{50}) data of organophosphorous compounds (OPs) with diverse structures. Interspecies QSTTR (quantitative structure-toxicity–toxicity relationships) models were developed to predict the mouse oral acute toxicity using the multiple linear regression (MLR) approach. Descriptors were calculated from the OPs structures optimized by molecular mechanics calculations. Model validation was performed using several statistical parameters. The results suggest the suitability of the developed QSTTR models to reliably predict the acute toxicity of OPs.

Keywords: organophosphourus compounds; acute toxicity; omega; quantitative structure-toxicity-toxicity relationships

Citation: Ilia, G.; Borota, A.; Funar-Timofei, S. Interspecies Quantitative Structure-Toxicity-Toxicity Relationships for Predicting the Acute Toxicity of Organophosphorous Compounds. *Chem. Proc.* **2021**, *3*, x. https://doi.org/10.3390/xxxxx

Academic Editor: Julio A. Seijas

Published: 15 November 2021

1. Introduction

Organophosphate compounds (OPs) are commonly used as pesticides and were developed as nerve gases for chemical wars [1–3]. OPs have been utilized as insecticides, helminthicides, ascaricides, nematocides, and to a lesser degree as fungicides and herbicides for several decades. Despite their worldwide application as crop protection agents, their wide usage has led to many intoxications of nontarget species, including human death. The inhibition of the enzyme acetylcholinesterase is usually the cause of the OPs acute mammalian toxicity [4]. In addition, other OP life-threatening toxicities have been observed, which are not always related to the acetylcholinesterase inhibition.

The oral acute toxicity assessment is very important because the oral route is a very common, convenient, safe, and inexpensive route of drug administration [5]. The importance of predicting rat acute oral toxicity is closely related to the knowledge of biological activity and mechanism of a potential drug, as well as its hazard identification and risk management [6]. This toxicity is often measured using the 50% lethal dose (LD_{50}), the amount of chemical that is expected to cause death in 50% of treated animals in a period of time. These expensive and time-consuming studies use large numbers of animals.

Information about toxicity to multiple species is important to assess the threat, and for the protection of ecological populations. When chemicals cause toxicity in a different genus of living organisms following a similar mechanistic path, there might be a correlation existing between the toxicities of these organisms [7]. Because such data is available for a limited number of species, to address these data gaps in species, alternative methods such as in silico models have been accepted to determine the acute toxicity.

Quantitative Structure-Activity/Toxicity Relationships (QSAR/QSTR) correlate the activity/toxicity of chemicals to their physicochemical properties and structural descriptors. They may reduce or even replace the need for animal testing and are most powerful when applied in a mechanistic hypothesis [8]. It is considered that as acute toxicity (LD_{50}) is related to whole body information it will be difficult to model it and may require knowledge on metabolism, bioaccumulation, excretion, etc. In addition, all data must be reliable, preferably obtained for the same sex and species [9].

To reduce the in vivo use of animals in toxicology, substitute species are useful in the risk assessment of chemicals [10]. They are based on results obtained using direct or indirect relationships from different toxicity tests [11,12].

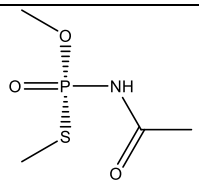
Interspecies quantitative structure–toxicity–toxicity (QSTTR) modeling allows the prediction of toxicity to several other species using the experimental toxicity values to one species [13]. This type of modeling can thus promote a reduction in the use of higher organisms and understanding of the mechanism of toxic action.

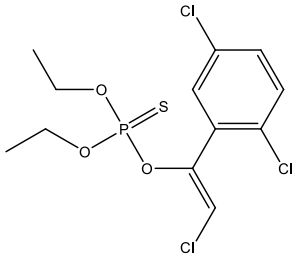
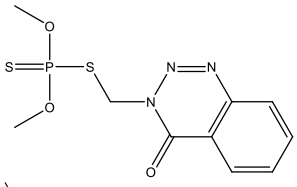
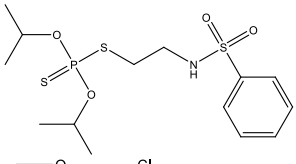
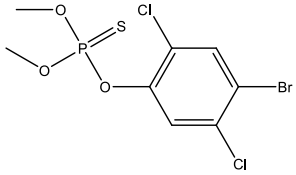
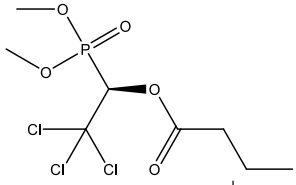
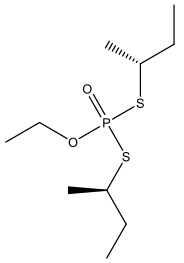
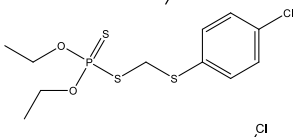
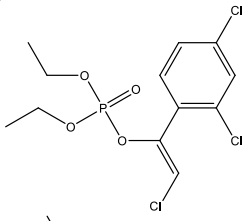
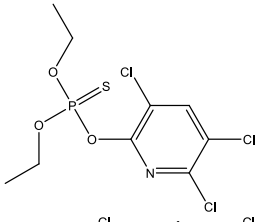
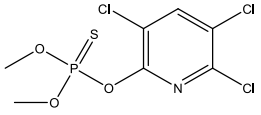
The interspecies QSTTRs extrapolate the data for one toxicity endpoint to those for another toxicity endpoint and can be used to determine the species-specific toxicity of a chemical [10,13–15].

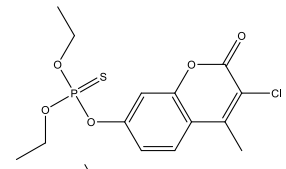
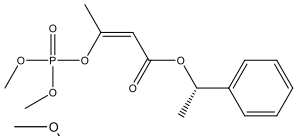
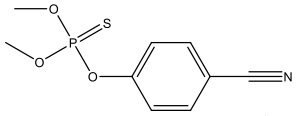
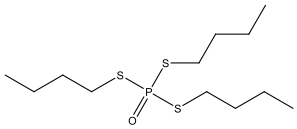
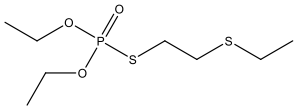
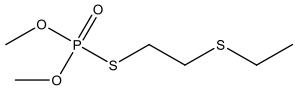
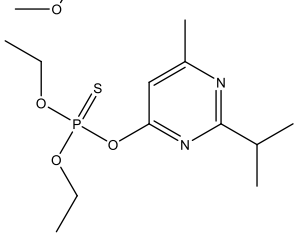
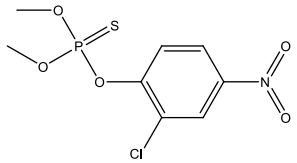
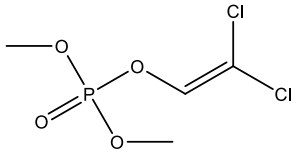
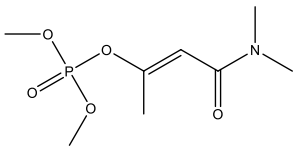
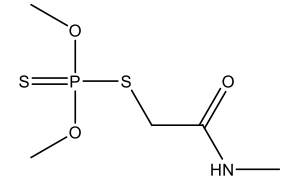
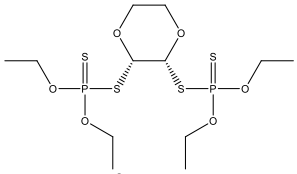
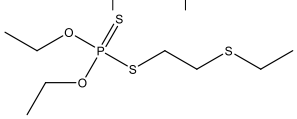
Using the underlying principle of taxonomic relationship, the development of predictive quantitative structure–toxicity–toxicity relationship (QSTTR) models allows predicting the toxicity of chemicals to a particular species using available experimental toxicity data towards a different species. Such studies may employ, along with the available experimental toxicity data to a species, molecular features and physicochemical properties of chemicals as independent variables for prediction of the toxicity profile against another closely related species [16].

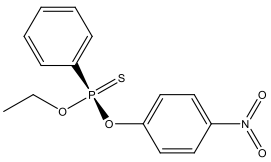
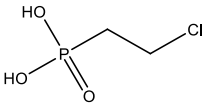
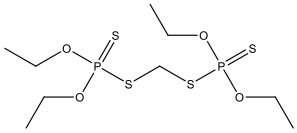
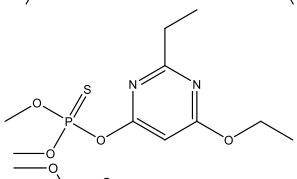
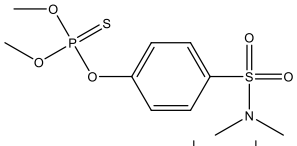
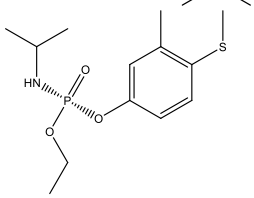
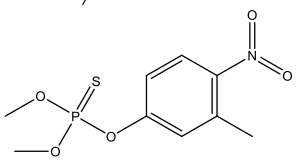
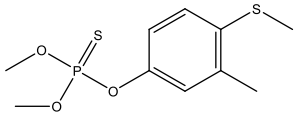
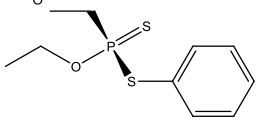
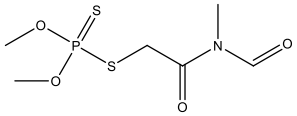
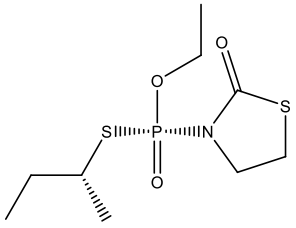
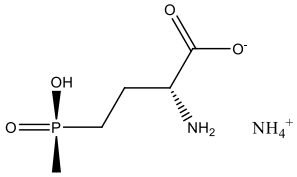
In this paper, we considered experimental rat and mouse acute toxicity data (LD_{50} values) of a series of 76 organophosphorous compounds (OPs) with diverse structures (Table 1). Interspecies QSTTR models were developed to predict the oral acute toxicity to a particular species using available experimental data towards a different species. The multiple linear regression approach was applied to extrapolate the known toxicity of chemicals of interest to species missing toxicity data. OP structures were optimized employing molecular mechanics calculations using the MMFF94s force field. Structural parameters were calculated based on the optimized structures. The mouse acute toxicity data of OPs was related to the rat acute toxicity using the multiple linear regression (MLR) approach. Additional descriptors improved the fitting quality of the MLR models. Model validation was performed using several statistical parameters to test the model predictive power. The results suggest the suitability of the developed QSTTR models to reliably predict the acute toxicity of organophosphorous chemicals.

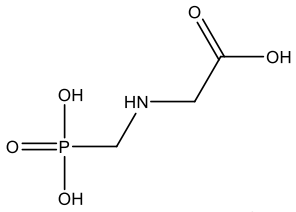
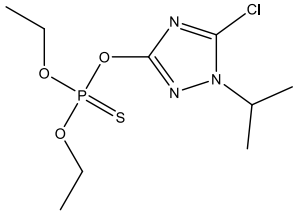
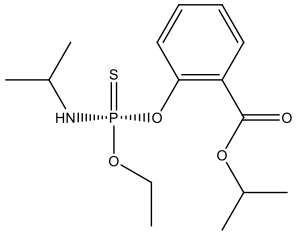
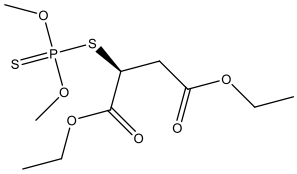
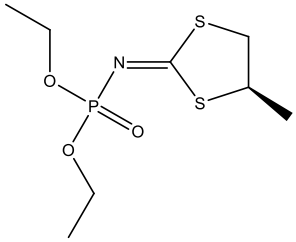
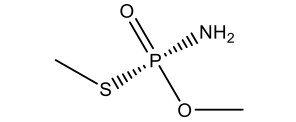
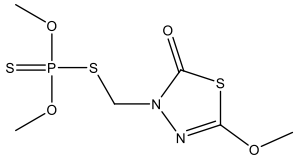
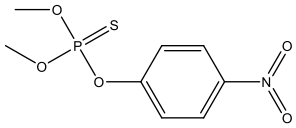
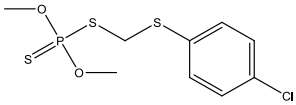
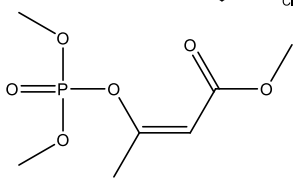
Table 1. The organophosphorous structures, the pLD_{50} values derived from experimental oral acute toxicity data of mouse and rat, CAS number, the predicted oral mouse pLD_{50} values model, and descriptors used in the best MLR1 model.

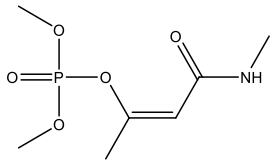
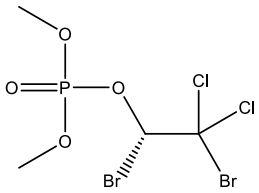
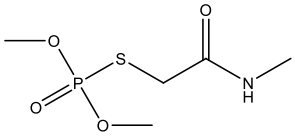
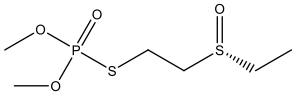
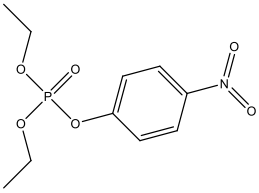
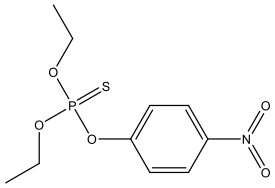
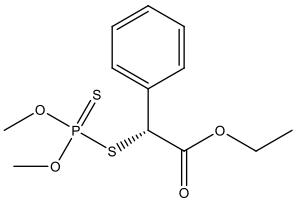
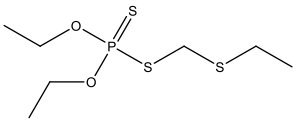
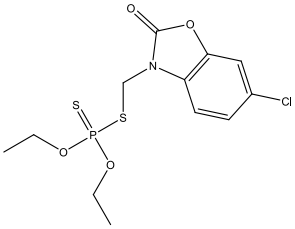
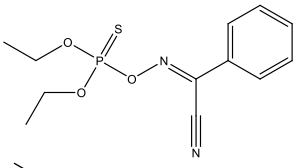
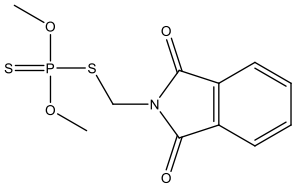
No	Structure	Experimental pLD_{50} (Oral Mouse, mole/kg)	Experimental pLD_{50} (Oral Rat, mole/kg)	CAS	Predicted pLD_{50} (Oral Mouse, MLR1)	Mor06m	TPSA(NO)	Mor26m
1 *		2.90	2.42	30560-19-1	2.82	1.069	55.4	0.119

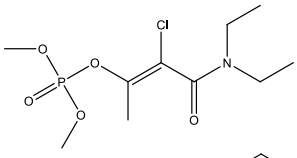
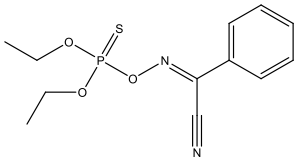
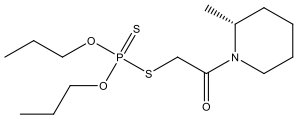
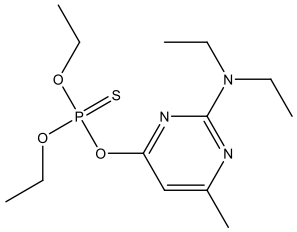
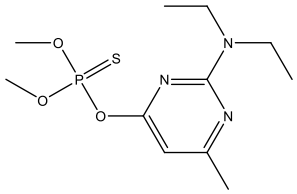
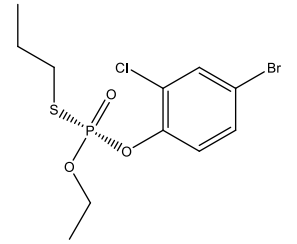
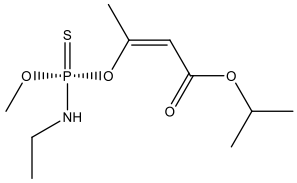
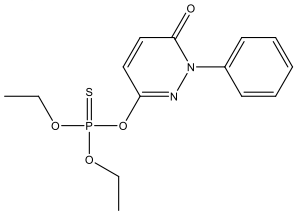
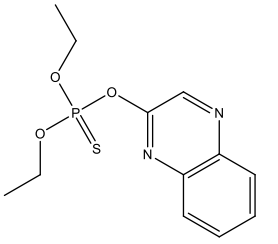
2		3.63	3.95	1757-18-2	3.64	4.117	27.69	−0.127
3 *		4.57	4.66	86-50-0	4.50	0.635	66.24	−0.415
4		2.41	3.17	741-58-2	3.03	3.657	64.63	−0.574
5 *		2.11	2.36	2104-96-3	2.33	2.746	27.69	−0.412
6		2.63	2.47	126-22-7	2.66	5.226	61.83	0.198
7		3.58	3.86	95465-99-9	3.95	0.352	26.3	0.242
8 **		3.20	4.70	786-19-6		1.971	18.46	−0.202
9		3.74	4.56	470-90-6	4.20	3.927	44.76	−0.171
10		3.77	3.63	2921-88-2	3.34	5.254	40.58	−0.2
11 *		2.20	2.25	5598-13-0	2.23	4.958	40.58	−0.215

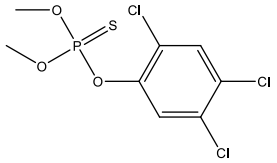
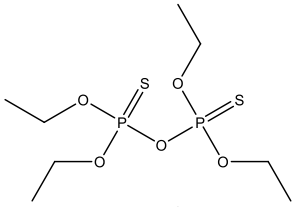
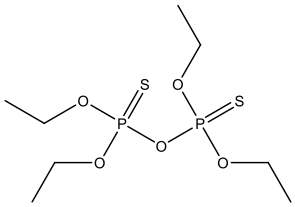
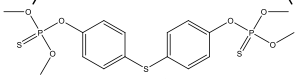
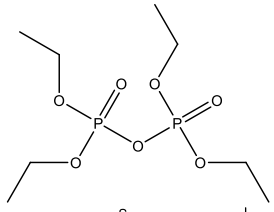
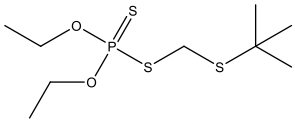
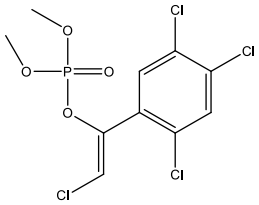
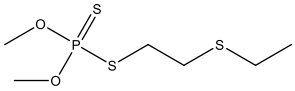
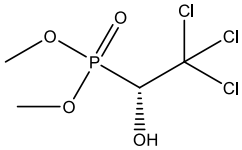
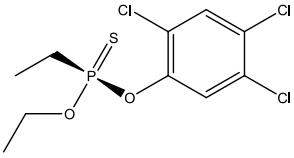
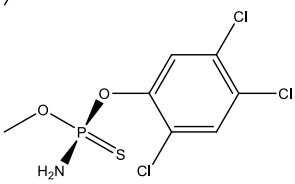
12 *		4.11	4.45	56-72-4	4.16	3.3	57.9	−0.297
13		3.90	3.91	7700-17-6	4.01	1.554	71.06	−0.032
14 *		2.53	3.05	2636-26-2	3.16	2.025	51.48	−0.119
15		3.61	3.32	78-48-8	3.59	−0.997	17.07	0.328
16		4.52	5.18	8065-48-3	4.98	2.493	35.53	0.413
17		3.95	3.49	8022-00-2	3.63	2.357	35.53	0.44
18		4.25	3.66	333-41-5	3.64	2.779	53.47	−0.046
19		2.95	3.02	2463-84-5	3.06	2.854	73.51	−0.452
20		3.56	4.11	62-73-7	3.80	4.86	44.76	−0.117
21		4.33	4.26	141-66-2	4.25	3.276	65.07	0.206
22		3.58	3.06	60-51-5	3.27	0.159	47.56	−0.115
23 *		3.41	4.36	78-34-2	4.19	2.259	55.38	−0.185
24		4.76	5.02	298-04-4	4.69	1.296	18.46	0.006

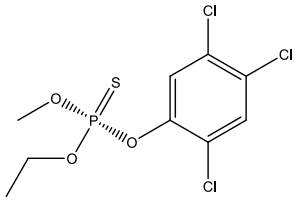
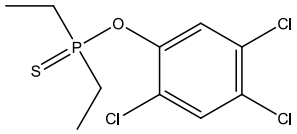
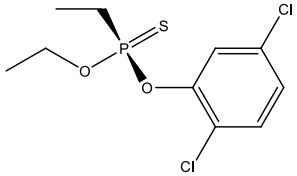
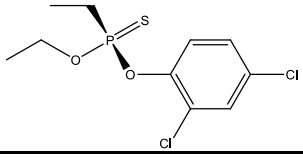
25		4.42	4.66	2104-64-5	4.45	1.519	64.28	−0.386
26		1.70	1.63	16672-87-0	2.13	2.265	57.53	0.155
27		3.98	4.47	563-12-2	4.21	1.299	36.92	−0.312
28		2.83	2.21	38260-54-7	2.62	1.709	62.7	0.046
29		4.53	4.53	52-85-7	4.10	4.272	65.07	−0.572
30		4.13	4.58	22224-92-6	4.22	3.577	47.56	−0.254
31		3.08	3.04	122-14-5	3.23	1.626	73.51	−0.253
32		3.50	3.19	55-38-9	3.24	2.117	27.69	0.107
33 *		4.25	4.36	944-22-9	4.09	0.675	9.23	−0.163
34		3.49	3.01	2540-82-1	3.20	0.689	55.84	−0.205
35 *		3.49	3.70	98886-44-3	3.76	2.941	46.61	0.29
36		2.64	2.05	77182-82-2	2.75	1.614	103.45	0.203

37		2.14	1.54	1071-83-6	2.26	1.958	106.86	0.012
38 *		3.97	4.07	42509-80-8	3.97	2.877	58.4	−0.074
39		3.58	4.21	25311-71-1	3.88	3.176	56.79	−0.558
40		3.24	2.44	121-75-5	2.84	0.065	71.06	−0.251
41		4.39	4.48	950-10-7	4.11	5.281	47.89	−0.04
42		4.00	4.27	10265-92-6	4.29	1.373	52.32	0.133
43		4.08	4.18	950-37-8	4.10	1.539	62.58	−0.26
44 *		4.17	4.64	298-00-0	4.57	1.592	73.51	−0.142
45		3.45	3.45	953-17-3	3.30	1.83	18.46	−0.227
46		4.75	4.87	7786-34-7	4.75	3.042	71.06	0.115

47 *		4.17	4.45	6923-22-4	4.40	3.444	73.86	0.151
48		3.23	3.62	300-76-5	3.43	2.549	44.76	−0.431
49 *		4.05	3.85	1113-02-6	4.03	1.321	64.63	0.187
50		4.39	3.91	301-12-2	4.00	2.645	52.6	0.327
51		5.56	5.18	311-45-5	4.98	3.204	90.58	−0.132
52 **		3.13	5.16	56-38-2		1.716	73.51	−0.178
53 *		3.37	3.65	2597-03-7	3.63	0.859	44.76	−0.27
54		5.06	5.42	298-02-2	4.98	1.123	18.46	−0.092
55		3.70	3.64	2310-17-0	3.51	0.394	53.6	−0.728
56		2.89	3.31	115-78-6	2.94	5.51	0	−0.023
57 *		4.09	3.54	732-11-6	3.70	−0.344	57.53	−0.204

58		4.70	4.57	13171-21-6	4.45	2.695	65.07	−0.006
59		2.45	3.00	14816-18-3	3.08	1.994	63.84	−0.376
60		3.03	3.04	24151-93-7	3.06	2.041	38.77	−0.191
61 *		3.50	3.38	23505-41-1	3.44	2.436	56.71	−0.074
62		2.41	2.39	29232-93-7	2.67	2.257	56.71	−0.034
63		3.36	3.02	41198-08-7	2.99	3.894	35.53	0.002
64		3.76	3.65	31218-83-4	3.51	2.805	56.79	−0.401
65		3.16	2.90	119-12-0	3.02	2.433	62.58	−0.244
66		3.60	4.06	13593-03-8	3.98	2.744	53.47	−0.003

67		2.21	2.71	299-84-3	2.45	3.929	27.69	−0.645
68		4.17	4.11	3689-24-5	3.90	3.596	46.15	−0.094
69		2.82	3.70	35400-43-2	3.52	1.671	18.46	−0.186
70		3.32	2.67	3383-96-8	2.80	4.163	55.38	0.058
71		4.99	5.76	107-49-3	5.52	4.48	80.29	0.365
72		4.92	5.26	13071-79-9	4.83	1.284	18.46	−0.124
73		2.42	2.88	22248-79-9	2.65	6.699	44.76	−0.215
74		3.82	3.79	640-15-3	3.69	1.159	18.46	−0.027
75*		2.93	2.76	52-68-6	2.79	5.269	55.76	0.034
76		3.92	4.35	327-98-0	3.63	4.043	18.46	−0.859
77 ***		-	2.64	2.591-66-4	2.60	3.29	44.48	−0.41

78 ***		-	3.03	2633-54-7	2.81	3.767	27.69	−0.393
79 ***		-	3.50	5745-14-2	3.17	3.196	9.23	−0.325
80 ***		-	4.20	7260-35-7	3.80	3.435	18.46	−0.219
81 ***		-	4.08	1593-27-7	3.69	3.065	18.46	−0.303

* Test compounds included in the MLR models. ** Outliers detected by the MLR1 model. *** External set.

2. Methods

2.1. Definition of Target Property and Structural Descriptors

The experimental mouse, respectively rat oral acute toxicity (LD_{50}) (mg/kg body weight), molar converted to pLD_{50} values, were taken from the ChemIDplus web search system (https://chem.nlm.nih.gov/chemidplus/, accessed on) and were considered as the dependent, respectively independent variables for 76 organophosphorus compounds (Table 1).

The OP structures were pre-optimized using the MMFF94 molecular mechanics force field included in the Omega (Omega v.2.5.1.4, OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com, accessed on) software [17,18] after curation of salts. Following parameters were used during the conformer ensemble generation: the maximum number of conformers per compound set of 400 and an RMSD value of 0.5 Å.

Structural parameters were further calculated using the minimum energy conformers by the DRAGON (Dragon Professional 5.5, 2007, Talete S.R.L., Milano, Italy) and InstantJChem (Instant JChem (2020) version 20.15.0, Chemaxon, http://www.chemaxon.com, accessed on) software.

An external set of 5 chemicals without experimental oral mouse acute toxicity data (Table 1) were collected from the PubChem Database (https://pubchem.ncbi.nlm.nih.gov/, accessed on). These compounds were chosen based on their structural similarity with the lowest toxic organophosphorous compound included in the above series of 76 OPs.

2.2. Multiple Linear Regression Approach and Model Validation

The multiple linear regression (MLR) approach [19] was employed to relate the mouse oral pLD_{50} values with the rat oral pLD_{50} values and calculated structural descriptors, using the QSARINS v. 2.2.4 program [20,21]. The genetic algorithm with leave-one-out cross-validation correlation coefficient was used for variable selection, as constrained function to be optimized, a mutation rate of 20%, the population size of 10 and 500 iterations.

The dataset was divided randomly into training and test (25% of the total number of compounds) sets. Following compounds: **1**, **3**, **5**, **11**, **12**, **14**, **23**, **33**, **35**, **38**, **44**, **47**, **49**, **53**, **57**, **61,** and **75** were included in the test set (Table 1).

For internal validation several measures of robustness were employed: Y-scrambling [22], adjusted correlation coefficient (r^2_{adj}), and q^2 (leave-one-out, q^2_{LMO}, and leave-more-out, q^2_{LMO}) cross-validation coefficient. In the Y-scrambling test, the dependent variable is arbitrarily mixed and a model is built using the same X matrix of molecular descriptors. The obtained MLR models (after 2000 randomizations) must have minimal r^2 (correlation coefficient) and q^2 (cross-validation coefficient) values [23].

The model overfit was checked using the Y-randomization test [23] and by comparing the root-mean-square errors (RMSE) and the mean absolute error (MAE) of the training and validation sets [24].

The applicability domain was checked using the Williams plot (hat diagonal values versus standardized residuals) for the training and prediction chemicals to find out the outliers and leverage compounds and the Insubria graph for chemicals without experimental data [25].

Several criteria were used to test the predictive model power: Q^2_{F1} [26], Q^2_{F2} [27], Q^2_{F3} [28], the concordance correlation coefficient (CCC) [29] (having the thresholds values higher than 0.85, [30]), and the predictive parameter r^2_m (with the lowest threshold value of 0.5) [31].

The Multi-Criteria Decision Making (MCDM) validation criterion [20,32] is used to summarize the performance of MLR models. To every validation criteria, a desirability function is associated, and MCDM has values between 0 (the worst) and 1 (the best).

3. Results and Discussion

The autoscaling method was employed for normalizing the data:

$$XT_{mj} = \frac{X_{mj} - \overline{X}_m}{S_m} \quad (1)$$

where for each variable m, XT_{mj} and X_{mj} are the j values for the m variable after and before scaling, respectively, $\overline{X}_m$ is the mean, and S_m is the standard deviation of the variable.

The variables contained in the MLR models were selected using the genetic algorithm. The statistical (fitting and predictivity) results are included in Tables 2–4. Two compounds (**18** and **52**) were detected as outliers, having standardized residual values greater than 2.5 standard deviation units, and were not included in the final MLR models.

The ‚MCDM all' scores, based on the fitting, cross-validated and external criteria were considered for choosing the best MLR models.

Table 2. Fitting and cross-validation statistical results of the MLR models.*.

Model	$r^2_{training}$	q^2_{LOO}	q^2_{LMO}	r^2_{adj}	**RMSEtr**	**MAEtr**	**CCCtr**	r^2_{scr}	q^2_{scr}	**SEE**	**F**
MLR1	0.850	0.819	0.810	0.839	0.316	0.260	0.919	0.072	−0.119	0.33	73.93
MLR2	0.833	0.811	0.805	0.823	0.334	0.272	0.909	0.053	−0.098	0.35	87.86
MLR3	0.820	0.801	0.795	0.814	0.346	0.284	0.901	0.035	−0.076	0.36	123.20
MLR4	0.800	0.786	0.782	0.796	0.365	0.301	0.889	0.017	−0.056	0.37	219.85

* $r^2_{training}$ —correlation coefficient; q^2_{LOO} —leave-one-out correlation coefficient; q^2_{LMO} —leave-more-out correlation coefficient; r^2_{adj} —adjusted correlation coefficient; RMSEtr-root-mean-square errors; MAEtr-mean absolute error; CCCtr-the concordance correlation coefficient; r^2_{scr} and q^2_{scr} —Y-scrambling parameters; SEE-standard error of estimates; F-Fischer test.

Table 3. The model predictivity results*.

Model	Q^2_{F1}	Q^2_{F2}	Q^2_{F3}	RMSE$_{ext}$	MAE$_{ext}$	CCC$_{ext}$
MLR1	0.826	0.822	0.857	0.309	0.223	0.910
MLR2	0.814	0.810	0.847	0.319	0.238	0.904
MLR3	0.780	0.775	0.819	0.347	0.273	0.875
MLR4	0.795	0.790	0.831	0.336	0.262	0.878

* Q^2_{F1}; Q^2_{F2}; Q^2_{F3}—external validation parameters; RMSE$_{ext}$—root-mean-square errors; MAE$_{ext}$—mean absolute error; CCC$_{ext}$-the concordance correlation coefficient.

Table 4. The 'MCDM all' score values, r^2_m predictivity parameter, and descriptors included in the MLR models.

Model	r^2_m	MCDM All	Descriptors Included in the MLR Models *
MLR1	0.827	0.851	pLD$_{50\ mouse}$, Mor06m, TPSA(NO), Mor26m
MLR2	0.851	0.842	pLD$_{50\ mouse}$, Mor06m, TPSA(NO)
MLR3	0.749	0.825	pLD$_{50\ mouse}$, R4v+
MLR4	0.758	0.822	pLD$_{50\ mouse}$

* pLD$_{50\ mouse}$—experimental oral mouse acute toxicity (mole/kg); Mor06m—3D-MoRSE—signal 06/weighted by atomic masses (3D-MoRSE descriptor); Mor26m—3D-MoRSE—signal 26/weighted by atomic masses (3D-MoRSE descriptor); TPSA(NO)—topological polar surface area using N,O polar contributions (molecular properties); R4v+—R maximal autocorrelation of lag 4/weighted by atomic van der Waals volumes (GETAWAY descriptors).

For the reliability of the best MLR1 model, the experimental versus predicted pLD$_{50}$ values, and Y-scramble plots are presented in Figures 1 and 2, respectively.

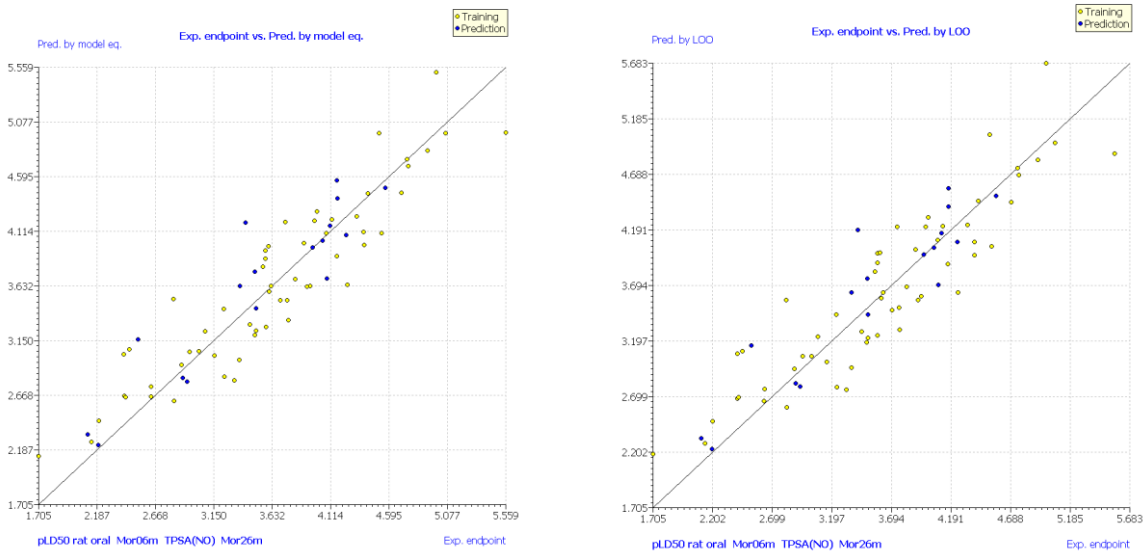


Figure 1. Plots of experimental versus predicted pLD$_{50}$ values for the MLR1 model predicted by the model (**left**) and by the leave-one-out (**right**) cross-validation approach (yellow circles-training compounds, blue circles-test compounds).

In the y-scrambling test performed for the MLR models, a significantly low scrambled r^2 (r^2_{scr}) and cross-validated q^2 (q^2_{scr}) values were obtained for 2000 trials. Figure 2 shows that in the case of all the randomized models, the values of r^2_{scr} and q^2_{scr} for the MLR1 model were < 0.5 (r^2_{scr} / q^2_{scr} of 0.072/−0.119). The low calculated r^2_{scr} and q^2_{scr} values indicate no chance correlation for all MLR chosen models (Table 2).

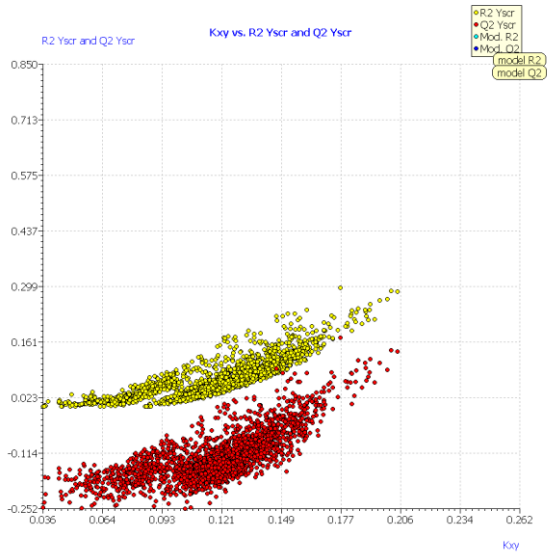


Figure 2. Y-scramble plots for the MLR1 model.

The Williams plot (standardized residuals versus leverages, with the leverage threshold h* = 0.263 for the MLR1 model), in the range of ±2.5σ, was used to verify the domain applicability. All compounds in the dataset are within the applicability domain of the MLR1 model, as presented in Figure 3.

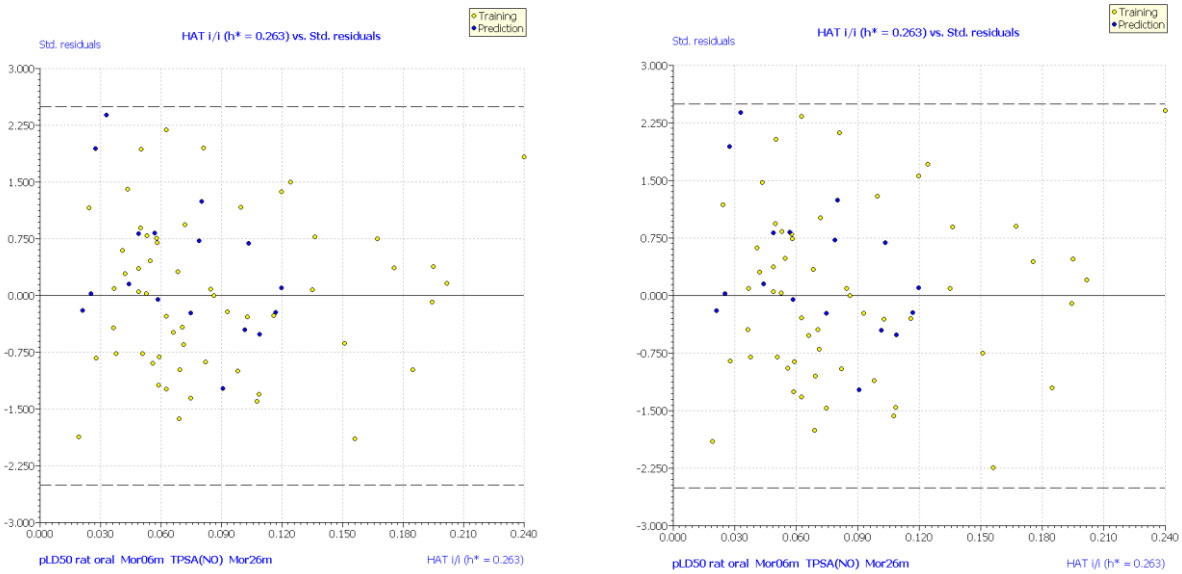


Figure 3. Williams plot predicted by the MLR1 model (**left**) and by the leave-one-out (**right**) cross-validation approach (yellow circles-training compounds, blue circles-test compounds).

The selected descriptors included in the MLR1 best model are not intercorrelated, as presented in the correlation matrix from Table 5.

Table 5. Correlation matrix of the descriptors included in the best MLR1 model, and their standardized coefficients (Std. coeff.).

	pLD50 Rat Oral	Mor06m	TPSA(NO)	Mor26m	Std. coeff.
pLD50 rat oral	1.0000				0.931
Mor06m	0.1866	1.0000			−0.121
TPSA(NO)	−0.2652	−0.0508	1.0000		0.126

Mor26m	−0.2929	−0.3333	0.1990	1.0000	0.134

Good correlations with the acute toxicity and predictive model power were notices for all MLR models. Closer values of the root-mean-square errors (RMSE) and the mean absolute error (MAE) of the training and validation sets were observed for the MLR2, MLR3, and MLR4 models. MLR1 model was considered being the best one according to several other statistical parameters of fitting and the 'MCDM all' score values.

The best MLR1 model has three descriptors: two 3D-MorSE descriptors (Mor06m, which represents 3D-MoRSE-signal 06/weighted by atomic masses and Mor26m, which represents 3D-MoRSE-signal 26/weighted by atomic masses); and one molecular property: TPSA(NO), which represents the topological polar surface area using N, O polar contributions. The increase of the Mor06m descriptor values would lead to lower acute toxicity. Higher values of Mor26m and TPSA(NO) descriptor values raise the OP toxicity.

To predict the mouse oral acute toxicity for OP chemicals without experimental data the best MLR1 model was applied to five external test compounds, found in the PubChem database, based on their structural similarity with the lowest known experimental OP mouse oral acute toxicity data of the 76 OPs.

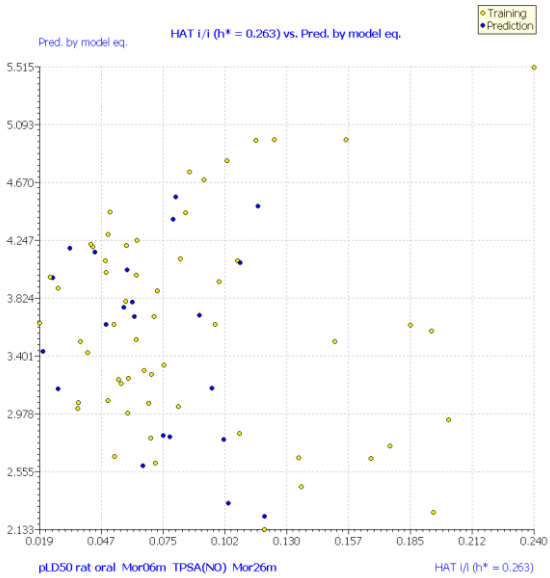


Figure 4. Insubria plot predicted by the MLR1 model (yellow circles-training compounds, blue circles-test compounds).

The Insubria plot of the predicted pLD_{50} versus hat values indicates that the five external set compounds are included in the applicability domain of the set of 76 OP compounds. The lowest predicted acute toxicity pLD_{50} values of the external set compounds **77** and **78** were confirmed by all four MLR models (Tables 2–4). These compounds contain a thiophosphonate, respectively thiphosphate group attached to the 2,4,5-trichlorophenyl moiety. Their predicted LD_{50} values of 767.8 mg/kg, respectively 519.3 mg/kg, obtained by the MLR1 model, indicate a low oral mouse acute toxicity.

4. Conclusions

Interspecies quantitative structure-toxicity-toxicity relationships were developed using the multiple linear regression approach to model the oral mouse acute toxicity of a series of organophosphorous compounds. The OP structures were modeled using the MMFF94s force field. The experimental mouse oral acute toxicity data of OPs was related to the rat oral acute toxicity using the multiple linear regression (MLR) approach. Additionally calculated descriptors of the minimum conformers improved the fitting quality

of the MLR models. Good correlations and predictive models were obtained. Molecular properties and 3D-MorSE descriptors included in the best MLR model can be used for the prediction of missing mouse oral acute toxicity data, saving experimental time and money. Two OPs with known structure (which include three chlorine atoms attached to a phenyl group and a thiophosphonate/thiophosphate group), without mouse toxicity data, were found to have potential low oral acute toxicity for this species.

Author Contributions: Conceptualization, S.F.T.; investigation, G.I. and A.B.; data curation, A.B.; validation, S.F.T.; writing—review and editing, S.F.T. and G.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This project was financially supported by Project 1.1 of the Coriolan Dragulescu Institute of Chemistry of the Romanian Academy. Access to the OpenEye Ltd., and Chemaxon Ltd. software are greatly acknowledged by the authors.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Katz, F.S.; Pecic, S.; Schneider, L.; Zhu, Z.; Hastings, A.; Luzac, M.; Macdonald, J.; Landry, D.W.; Stojanovic, M.N. New Therapeutic Approaches and Novel Alternatives for Organophosphate Toxicity. *Toxicol. Lett*. **2018**, *291*, 1–10. https://doi.org/10.1016/j.toxlet.2018.03.028.
2. Lerro, C.C.; Koutros, S.; Andreotti, G.; Friesen, M.C.; Alavanja, M.C.; Blair, A.; Hoppin, J.A.; Sandler, D.P.; Lubin, J.H.; Ma, X.; et al. Organophosphate insecticide use and cancer incidence among spouses of pesticide applicators in the Agricultural Health Study. *Occup. Environ. Med.* **2015**, *72*, 736–744. https://doi.org/10.1136/oemed-2014-102798.
3. Sultatos, L.G. Mammalian toxicology of organophosphorus pesticides, *J. Toxicol. Environ. Health Part A* **1994**, *43*, 271–289. https://doi.org/10.1080/15287399409531921.
4. Fukuto, T.R. Mechanism of action of organophosphorus and carbamate insecticides. *Environ. Health Persp*. **1990**, *87*, 245–254. https://doi.org/10.1289/ehp.9087245.
5. Strickland, J.; Clippinger, A.J.; Brown, J.; Allen, D.; Jacobs, A.; Matheson, J.; Lowit, A.; Reinke, E.N.; Johnson, M.S.; Quinn, M.J., Jr.; et al. Status of acute systemic toxicity testing requirements and data uses by U.S. regulatory agencies. *Regul. Toxicol. Pharmacol.* **2018**, *94*, 183–196.
6. Minerali, E.; Foil, D.; Zorn, Ki.; Ekins, S. Evaluation of Assay Central Machine Learning Models for Rat Acute Oral Toxicity Prediction. *ACS Sustain. Chem. Eng*. **2020**, *8*, 16020–16027. https://doi.org/10.1021/acssuschemeng.0c06348.
7. Kar, S.; Das, R.N.; Roy, K.; Leszczynski, J. Can Toxicity for Different Species be Correlated? The Concept and Emerging Applications of Interspecies Quantitative Structure-Toxicity Relationship (i-QSTR) Modeling. *Int. J. Quant. Struct. -Prop. Relatsh.* **2016**, *1*, 23–51. https://doi.org/10.4018/IJQSPR.2016070102.
8. Cronin, M.T. (Q)SARs to predict environmental toxicities: Current status and future needs. *Environ. Sci. Processes Impacts* **2017**, *19*, 213–220. https://doi.org/10.1039/C6EM00687F.
9. Cronin, M.T.D.; Dearden, J.C. QSAR in Toxicology. 2. Prediction of Acute Mammalian Toxicity and Interspecies Correlation. *Quant. Struct. -Act. Relat*. **1995**, *14*, 117-120. https://doi.org/10.1002/qsar.19950140202.
10. Cronin, M.T.D. Biological read-across: Mechanistically based species-species and endpoint-endpoint extrapolations. In *In Silico Toxicology: Principles and Applications*, Cronin, M.T.D., Madden, J.C., Eds.; Royal Society of Chemistry: Cambridge, UK, 2010, pp. 446–477.
11. Vermeire, T.G.; Baars, A.J.; Bessems, J.G.M.; Blaauboer, B.J.; Slob, W.; Muller, J.J.A. Toxicity Testing For Human Health Risk Assessment. In *Risk Assessment of Chemicals*, 2nd ed.; *An Introduction*; van Leeuwen, C.J., Vermeire, T.G., Eds.; Springer: Dordrecht, The Netherlands, 2007; pp. 227–280.
12. Traas, T.P.; van Leeuwen, C.J. Ecotoxicological Effects. In *Risk Assessment of Chemicals. An Introduction*, 2nd ed.; van Leeuwen, C.J., Vermeire, T.G., Eds.; Springer: Dordrecht, The Netherlands, 2007; pp. 281–356.
13. Das, R.N.; Roy, K.; Popelier, P.L.A. Interspecies quantitative structure–toxicity–toxicity (QSTTR) relationship modeling of ionic liquids. Toxicity of ionic liquids to *V. fischeri*, *D. magna* and *S. vacuolatus*. *Ecotoxol. Environ. Saf.* **2015**, *122*, 497–520. https://doi.org/10.1016/j.ecoenv.2015.09.014.
14. Cassani, S.; Kovarich, S.; Papa, E.; Roy, P.P.; van der Wal, L.; Gramatica, P. Daphnia and fish toxicity of (benzo)-triazoles: Validated QSAR models, and interspecies quantitative activity–activity modelling, *J. Hazard. Mater.* **2013**, *258–259*, 50–60. https://doi.org/10.1016/j.jhazmat.2013.04.025.

15. Furuhama, A.; Hasunuma, K.; Aoki, Y. Interspecies quantitative structure–activity–activity relationships(QSAARs) for prediction of acute aquatic toxicity of aromatic amines and phenols. *SAR QSAR Environ. Res.* **2015**, *26*, 301–323. https://doi.org/10.1080/1062936X.2015.1032347.
16. Roy, K.; Das, R.N.; Popelier, P.A. Predictive QSAR modelling of algal toxicity of ionic liquids and its interspecies correlation with Daphnia toxicity. *Environ. Sci. Pollut. Res.* **2015**, *22*, 6634–6641. https://doi.org/10.1007/s11356-014-3845-0.
17. Hawkins, P.C.D.; Skillman, A.G.; Warren, G.L.; Ellingson, B.A.; Stahl, M.T. Conformer generation with OMEGA: Algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **2010**, *50*, 572–584. https://doi.org/10.1021/ci100031x.
18. Hawkins, P.C.D.; Nicholls, A. Conformer generation with OMEGA: Learning from the data set and the analysis of failures. *J. Chem. Inf. Model.* **2012**, *52*, 2919–2936. https://doi.org/10.1021/ci300314k.
19. Wold, S.; Dunn III, W.J. Multivariate quantitative structure-activity relationships (QSAR): Conditions for their applicability. *J. Chem. Inf. Comput. Sci*. **1983**, *23*, 6–13. https://doi.org/10.1021/ci00037a002.
20. Chirico, N.; Sangion, A.; Gramatica, P.; Bertato, L.; Casartelli, I.; Papa, E.. QSARINS-Chem standalone version: A new platform-independent software to profile chemicals for physico-chemical properties, fate, and toxicity. *J. Comput. Chem.* **2021**, *42*, 1452–1460. https://doi.org/10.1002/jcc.26551.
21. Gramatica, P.; Chirico, N.; Papa, E.; Cassani, S.; Kovarich, S. QSARINS: A new software for the development, analysis, and validation of QSAR MLR models. *J. Comput. Chem*. **2013**, *34*, 2121–2132. https://doi.org/10.1002/jcc.23361.
22. Todeschini, R.; Consonni, V.; Maiocchi, A. The K correlation index: Theory development and its application in chemometrics. *Chemom. Intell. Lab*. **1999**, *46*, 13–29. https://doi.org/10.1016/S0169-7439(98)00124-5.
23. Eriksson, L.; Johansson, E.; Kettaneh-Wold, N.; Wold, S. Multi and megavariate data analysis: Principles and applications. Umetrics AB: Umea, Sweden, 2001; pp. 92–97, 489–491.
24. Goodarzi, M.; Deshpande, S.; Murugesan, V.; Katti, S.B.; Prabhakar, Y.S. Is Feature Selection Essential for ANN Modeling? *QSAR Comb. Sci*. **2009**, *28*, 1487–1499. https://doi.org/10.1002/qsar.200960074.
25. Gramatica, P. Principles of QSAR Modeling: Comments and Suggestions from Personal Experience. *Int. J. Quant. Struct. -Prop. Relatsh.* **2020**, *5*, 1–37. https://doi.org/10.4018/IJQSPR.20200701.oa1.
26. Shi, L.M.; Fang, H.; Tong, W.; Wu, J.; Perkins, R.; Blair, R.M.; Branham, W.S.; Dial, S.L.; Moland, C.L.; Sheehan, D.M. QSAR models using a large diverse set of estrogens. *J. Chem. Inf. Model*. **2001**, *41*, 186–195. https://doi.org/10.1021/ci000066d.
27. Schüürmann, G.; Ebert, R.U.; Chen, J.; Wang, B.; Kühne, R. External validation and prediction employing the predictive squared correlation coefficient test set activity mean vs training set activity mean. *J. Chem. Inf. Model*. **2008**, *48*, 2140–2145. https://doi.org/10.1021/ci800253u.
28. Consonni, V.; Ballabio, D.; Todeschini, R. Comments on the definition of the Q2 parameter for QSAR validation. *J. Chem. Inf. Model*. **2009**, *49*, 1669–1678. https://doi.org/10.1021/ci900115y.
29. Chirico, N.; Gramatica, P. Real External Predictivity of QSAR Models: How To Evaluate It? Comparison of Different Validation Criteria and Proposal of Using the Concordance Correlation Coefficient. *J. Chem. Inf. Model*. **2011**, *51*, 2320–2335. https://doi.org/10.1021/ci200211n.
30. Chirico, N.; Gramatica, P. Real External Predictivity of QSAR Models. Part 2. New Intercomparable Thresholds for Different Validation Criteria and the Need for Scatter Plot Inspection. *J. Chem. Inf. Model*. **2012**, *52*, 2044–2058. https://doi.org/10.1021/ci300084j.
31. Roy, K.; Mitra, I. On the Use of the Metric r_m^2 as an Effective Tool for Validation of QSAR Models in Computational Drug Design and Predictive Toxicology. *Mini-Rev. Med. Chem*. **2012**, *12*, 491−504. https://doi.org/10.2174/138955712800493861.
32. Keller, H.R.; Massart, D.L.; Brans, J.P. Multicriteria decision making: A case study. *Chemom. Intell. Lab. Syst*. **1991**, *11*, 175–189. https://doi.org/10.24048/ams3.no1.2014-109.. https://doi.org/