Journal of Agricultural, Food and Environmental Sciences Vol 77 No 2 (2023) 49-56
DOI: https://doi.org/10.55302/JAFES23772049k

Original scientific paper

QSAR MODELING OF pK_a VALUES OF SULFONYLUREA HERBICIDES

Kristina Krstevska, Igor Jordanov, Dejan Dimitrovski, Vesna Dimova*

Faculty of Technology and Metallurgy, Ss. Cyril and Methodius University in Skopje,
Ruger Boskovic 16, 1000 Skopje, Republic of Macedonia
*corresponding author: vdimova@tmf.ukim.edu.mk

ABSTRACT

Sulfonylureas are herbicides primarily used for control of weeds in early growth stages of cultivations. Structurally sulfonylureas contain a sulfonyl group with sulphur atom bonded to nitrogen atom of an ureylene group. According side chains there are pyrimidinyl-sulfonylureas and the triazinyl-sulfonylureas. Swiss ADME descriptors have been used to develop QSAR models for predicting the pK_a values of selected 27 sulfonlyurea herbicides: 17 pyrimidinyl-sulfonylurea herbicides and 10 triazinyl-sulfonylurea herbicides. Variable selection methods including stepwise, forward, and best model were employed. Two different approaches were performed to develop a predictive QSAR model: a set with all selected herbicides and a divided set according structure (pyrimidinyl/ triazinyl). QSAR models were analyzed using following statistical parameters: coefficient of correlation, adjusted coefficient of correlation, mean squared error, root mean square error, and Fischer test. Models with four descriptors in both sets of herbicides were statistically better, based on the values of these parameters.

Key words: herbicides, sulfonylureas, SwissADME descriptors, QSAR models.

INTRODUCTION

Pesticides are group of substances, exhibiting notably herbicide, insecticide, rodenticide, fungicide, or nematicide properties, used to control and repel pests in different fields (Marchand, 2023). Today pesticides are widely distributed environmental pollutants, because of their large use for agricultural, industrial, or domestic purposes (Pathak, et al., 2022).

Humans and animals are for that reason highly exposed to these chemicals, by an oral way through the consumption of food or water contaminated by pesticides. This is a major health issue, because today the toxicity of pesticides is well studied, meaning that they can cause various diseases, such as cancer and neurodegenerative, metabolic, reproductive or developmental pathologies. For toxicodynamic effect, exposed humans or animals, particularly at the gastro-intestinal level in response to oral exposure, must absorb pesticides.

In the field of drug discovery very important fact is that the passage of drugs across the intestinal barrier and the blood-brain barrier is usually extensively characterized using in vitro cellular models, animal experimentation, and clinical human pharmacokinetics studies before marketing authorization (Sjogren, et al., 2014).

On the other hand, human intestinal absorption and human brain distribution correspond to key steps of pesticide toxicokinetics, but remain poorly characterized (Chedik, et al., 2017). Problem with pesticides is that the prediction of brain permeation unfortunately could not be validated by human experimental pharmacokinetics data. That is the reasons for absence of experimental data in the scientific literature and pesticide database.

Today the intestinal absorption and brain disposition of drugs can moreover be *in silico* predicted with good accuracy, using various computer models (Toropov, et al., 2017).

The substituted urea herbicides are a large group of non-selective herbicidal agents, introduced in the 1960s, applied for broadleaf weed and grass control in such noncrop areas.

Sulfonylureas (SU) are group of substituted urea herbicides, primarily used for the control of weeds (annual and perennial) in early growth stages of cultivations (Bempelou, et al., 2019). Structurally all sulfonylureas consist a sulfonyl group ($-S(=O)_2$) with its sulphur atom bonded to nitrogen atom of an ureylene group (functional group derived from urea) (Fig. 1). The side chains (R_1 and R_2) distinguish various sulfonylureas, such as pyrimidinyl-sulfonylureas and the triazinyl-sulfonylureas. The first sulfonylurea herbicides were introduced in 1982 (CDC, 2023).

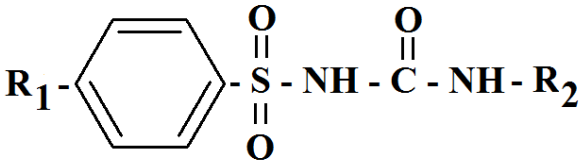

Figure 1. Structural formula of sulfonylurea compounds

SU are characterized by a broad-spectrum weed control at low application rates, good crop selectivity, and low acute and chronic animal toxicity (LD50 > 4000 mg/kg). Since they are widely applied as pre- and post-emergence herbicides, there are high possibility that their residues may contaminate water, soil, and air and accumulate in plant products (Losito, et al., 2006). They inhibit the biosynthesis of the essential amino acids valine and isoleucine, stopping the cell division and plant growth (Liu, et al., 2023); generally are slightly toxic to freshwater fish and invertebrates, and practically nontoxic to wildfowl and other mammals (CDC, 2023).

The pK_a constant is a measure of the acidity of a molecule or compound and its is an important parameter in many chemical and biological processes, and its value can have significant effects on the behavior of molecules and compounds (Pereira, et al., 2016).

The behavior of pesticides and their metabolites in the environment are largely dependent on their physicochemical properties. In case of ionisable pesticides, their pKa, values determine the degree of ionisation in water at the pH of the soil or biological system, and this in turn determines their effective lipophilicity. Accurate pKa values are therefore necessary to model pesticide behavior (Devlin, et al., 2008).

SwissADME is a web tool designed for predicting pharmacokinetics parameters (Daina, et al., 2017; Chedik, et al., 2017), which can be used for pesticides. The BOILED egg method, as part of SwissADME tool, can predict the transfer of drugs across the blood-brain barrier with high accuracy (Daina & Zoete, 2016).

In this work, an attempt has been made to apply this method for other classes of compounds such as sulfonylurea herbicides: pyrimidinyl- sulfonylurea herbicides and triazinyl-sulfonylurea herbicides.

MATERIALS AND METHODS

Pesticide Set

The list of analyzed pesticides in the study and their symbols are shown in Table 1. A Simplified Molecular Input Line Entry System (SMILES) was collected online from for PubChem database (PubChem 2023).

Table 1. Name, symbol and pK_a values of investigated sulfonylurea herbicides H1-H27

Symbol	Pyrimidinyl-sulfonylurea herbicides	pK_a	Symbol	Triazinyl-sulfonylurea herbicides	pK_a
H1	Amidosulfuron	3.58	H18	Chlorsulfuron	3.40
H2	Azimsulfuron	3.60	H19	Cinosulfuron	4.72
H3	Cyclosulfamuron	5.04	H20	Ethametsulfuron methyl	4.20
H4	Ethoxysulfuron	5.28	H21	Metsulfuron methyl	3.75
H5	Flazasulfuron	4.37	H22	Prosulfuron	3.76
H6	Flucetosulfuron	3.50	H23	Thifensulfuron methyl	4.00
H7	Foramsulfuron	4.60	H24	Triasulfuron	4.64
H8	Halosulfuron methyl	3.44	H25	Tribenuron methyl	4.65
H9	Imazosulfuron	3.94	H26	Triflusulfuron methyl	4.40
H10	Nicosulfuron	4.78	H27	Tritosulfuron	4.69
H11	Oxasulfuron	5.10			
H12	Primisulfuron	3.47			
H13	Propyrisulfuron	4.89			
H14	Pyrazosulfuron ethyl	3.70			
H15	Rimsulfuron	4.00			
H16	Sulfometuron methyl	5.20			
H17	Sulfosulfuron	3.51			

pK_a values

All pK_a values, presented in Table 1, were taken from Pesticide Properties Database (PPDB) website, a comprehensive source of data on pesticide chemical, physical and biological properties (PPDB, 2023).

SwissADME descriptors

SwissADME gives access to physicochemical, lipophilicity, water solubility, pharmacokinetics and drug-likeness properties and descriptors (Daina, et al., 2017)**.**

Brain or intestinal estimated permeation

Using the SwissADME a web tool, the BOILED-Egg (Brain or Intestinal Estimated permeation) graph was applied for investigated pesticides (Islamoğlu & Hacifazlioğlu, 2022). The BOILED-Egg allows for intuitive evaluation of passive gastrointestinal absorption and blood-brain penetration (BBB) in function of the position of the molecules in the WLOGP (a purely atomistic method based on Wildman and Crippen's (Wildman & Crippen, 1999) piecewise system of the octanol-water distribution coefficient (log P) used as a measure of lipophilicity)-versus-TPSA (topological polar surface area) referential in the SwissADME a web tool (Daina, et al., 2017). In BOILED-Egg graph, the yellow area represents the transition to the blood-brain barrier (BBB), and the white area represents the absorption in the gastrointestinal system (AGS).

Multiple Linear Regression Models (MLR)

MLR is a method used for modeling linear relationship between a dependent variable Y, in this work pKa values and independent variable X - previously calculated 34 selected descriptors. Stepwise, forward and best model (with 2, 3 and 4 descriptors) variable selection methods were used.

To develop a predictive QSAR model using XLSTAT software (XLSTAT, 2014), two different approach were performed:

i) Set with all 27 selected herbicides H1-H27 labeled as **TS**.

ii) Divided set: pyrimidinyl-sulfonylurea herbicides H1-H17 labeled as **TS1** and triazinyl-sulfonylurea herbicides H18-H27 labeled as **TS2**.

RESULTS AND DISCUSSION

The present study intended to develop a mathematical QSAR model between the SweesADME descriptors of sulfonylurea herbicides and their pK_a values. For this reason, an experimental database was consist of 27 herbicides and 34 different numerical descriptors. The herbicide set was divided on 17 pyrimidinyl-sulfonylurea herbicides (H1-H17) and 10 triazinyl-sulfonylurea herbicides (H18-H27). Several variable selection methods were used such as: Stepwise (SW), Forward (FW) and Best model selection with 2, 3 and 4 descriptors (BM2, BM3 and BM4).

In this study, one of the tasks was using SwissADME web tool *in silico* investigation of the human intestine and brain permeation of 27 sulfonylurea herbicides. SwissADME web tool, primarily was developed and validated for drugs (Daina, et al., 2017).

In the BOILED-EGG chart, the white region indicates a high possibility of passive absorption from the gastrointestinal tract, while the yellow region is for a high possibility of brain diffusion. In the graph, pyrimidinyl-sulfonylurea herbicides (H1-H17) are represented as blue dots and triazinyl-sulfonylurea herbicides (H18-H27) as red dots. (Figure 2).

Our results showed that a large proportion of studded herbicides (more than 92.59%) is predicted to be low absorbed by the human gastro-intestinal tract (Figure 2– gray region) (Mostafalou & Abdollahi, 2017). Only two herbicides: sulfometuron methyl and chlorsulfuron (H16 and H18) are high possibility of passive absorption by the gastrointestinal tract (Figure 2 - white region).

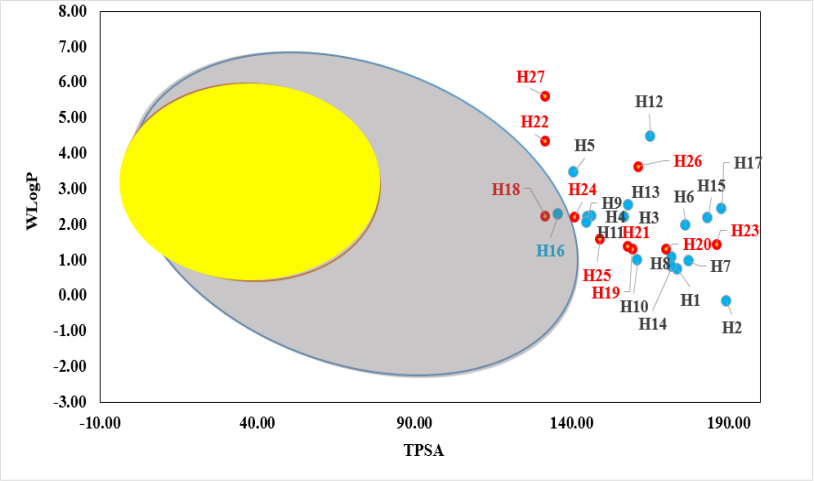

Figure 2. BOILED-Egg graph for all investigated pesticides

The next step in the research, QSAR models were constructed using all 27 herbicides (**TS**) and all calculated descriptors using multiple linear regression (MLR). MLR is a statistical method capable to evaluate the linear relationship between the molecular descriptors and the pKa values. The correlation coefficients (R^2) for those models were analyzed (Table 2).

Since unsatisfactory statistical results were obtained $R^2 < 0.7$ (Table 2), next step was constructed QSAR model for separated sets of herbicides: pyrimidinyl-sulfonylurea herbicides H1-H17 (**TS1)** and triazinyl-sulfonylurea herbicides H18-H27 (**TS2)**.

The results showed that QSAR models with a larger number of descriptors (3 and 4) are statistically more significant ($R^2 > 0.8$) (Table 2).

Table 2. R^2 values for stepwise (SW), forward (FW) and best model selection with two (BM2), three (BM3) and four (BM4) descriptors for TS, TS1 and TS2

R^2	SW	FW	BM2	BM3	BM4
TS	0.675	0.675	0.459	0.602	0.675
TS1	0.808	0.808	0.718	**0.827**	**0.936**
TS2	0.371	0.371	0.691	**0.888**	**0.974**

Further analysis was made on statistically adequate models presented in Table 3. Models are labeled as **TS1-1** for set 1 and best model with 3 descriptors; **TS1-2** for set 1 and best model with 4 descriptors; **TS2-1** for set 2 and best model with 3 descriptors and **TS2-2** for set 2 and best model with 4 descriptors.

Table 3. QSAR equation for ModelsTS1-1, TS1-2, TS2-1 and TS2-2

Model	QSAR equation
TS1-1	pK_a = 2.881 + 0.878*MLOGP + 185.737*ESOL + 2.521*BioScore
TS1-2	pK_a = 7.317 - 0.032*MW + 0.069*MR - 0.514*SiLogSw + 0.809*Lv
TS2-1	pK_a = 4.776 + 0.296*Hba + 0.747*AliLog S + 0.313*Vv
TS2-2	pK_a = 1.659 + 0.407*Hba - 0.605*XLOGP3 + 0.407*Vv - 0.629*Mv

ESOL - ESOL Solubility (mol/l); BioScore - Bioavailability Score; MW – molecular weight; MR – molar refractivity; SiLogSw - Silicos-IT LogSw; Lv - Leadlikeness #violations; Hba - #H-bond acceptors; Vv - Veber #violations

The negative sign of the coefficient of MW and SiLogSw in Models TS1-2 and XLOGP3 and MV in TS2-2 reflects that pK_a values will be improved for low values of both descriptors. On the contrary, the positive sign of the coefficients of the other descriptors indicates that high values of these descriptors will increase pK_a values.

The variance inflation factor (VIF) test ensures that the modeling process is not accompanied with multicollinearity. To accept the model, the VIF value should be between 1 and 5, but in the case of VIF values higher than 10, there is significant multicollinearity; so the model must be corrected (Sadeghi, et al., 2022). In our case, descriptors in all models has VIF < 4 (Table 4), so these descriptors showed no intercorrelation.

Table 4. VIF values for Models: TS1-1, TS1-2, TS2-1 and TS2-2

Model TS1-1	**Statistic**	**MLOGP**	**ESOL**	**BioScore**	
	Tolerance	0.8025	0.8089	0.9422	
	VIF	1.2460	1.2362	1.0613	
Model TS1-2	**Statistic**	**MW**	**MR**	**SiLogSw**	**Lv**
	Tolerance	0.2504	0.2879	0.6924	0.5384
	VIF	3.9936	3.4737	1.4442	1.8574
Model TS2-1	**Statistic**	**Hba**	**Ali Log S**	**Vv**	
	Tolerance	0.5045	0.4995	0.9712	
	VIF	1.9821	2.0020	1.0296	
Model TS2-2	**Statistic**	**Hba**	**XLOGP3**	**Vv**	**Mv**
	Tolerance	0.3175	0.4250	0.4717	0.3302
	VIF	3.1496	2.3527	2.1200	3.0285

Additional statistical parameters such as: adjusted coefficient of correlation ($R^2_{adj.}$), mean squared error (MSE), root mean square error (RMSE) and Fischer test (F-test), were also analyzed (Table 5). Adjusted coefficient of correlation ($R^2_{adj.}$) is a modified version of R^2 that has been adjusted for the number of descriptors in the QSAR model.

Table 5. $R^2_{adj.}$, MSE, RMSE and F-test values for Models: TS1-1, TS1-2, TS2-1 and TS2-2

	Model TS1-1	**Model TS1-2**	**Model TS2-1**	**Model TS2-2**
$R^2_{adj.}$	0.7866	0.9144	0.8315	0.9524
MSE	0.1046	0.0420	0.0378	0.0107
RMSE	0.3234	0.2049	0.1944	0.1033
F-test	20.653	43.711	15.801	46.040
$F_{crit.}$: *3,411 for TS1-1; 3.260 for TS1-2; 4.751 for TS2-1 and 5.192 for TS2-2*				

According the values of $R^2_{adj.}$, MSE, RMSE and F-test, models with 4 descriptor in both sets of herbicides were statistically better: the $R^2_{adj.}$ and F-test values are higher; MSE and RMSE values are lower in four-parametric models TS1-2 and TS2-2. The value of F-test compared to the critical value (F_{crit}) in all QSAR models is relatively high, meaning that the error committed is less than what the model explains (Table 5).

The plot of observed versus predicted activities for Models: TS1-1, TS1-2, TS2-1 and TS2-2 is presented in Figure 3. The all modes are statistically significant since all the point are very close to regression line. There is a perfect fit between experimental and predicted pK_a values in all analyzed models ($R^2_{pred.} > 0.8$).

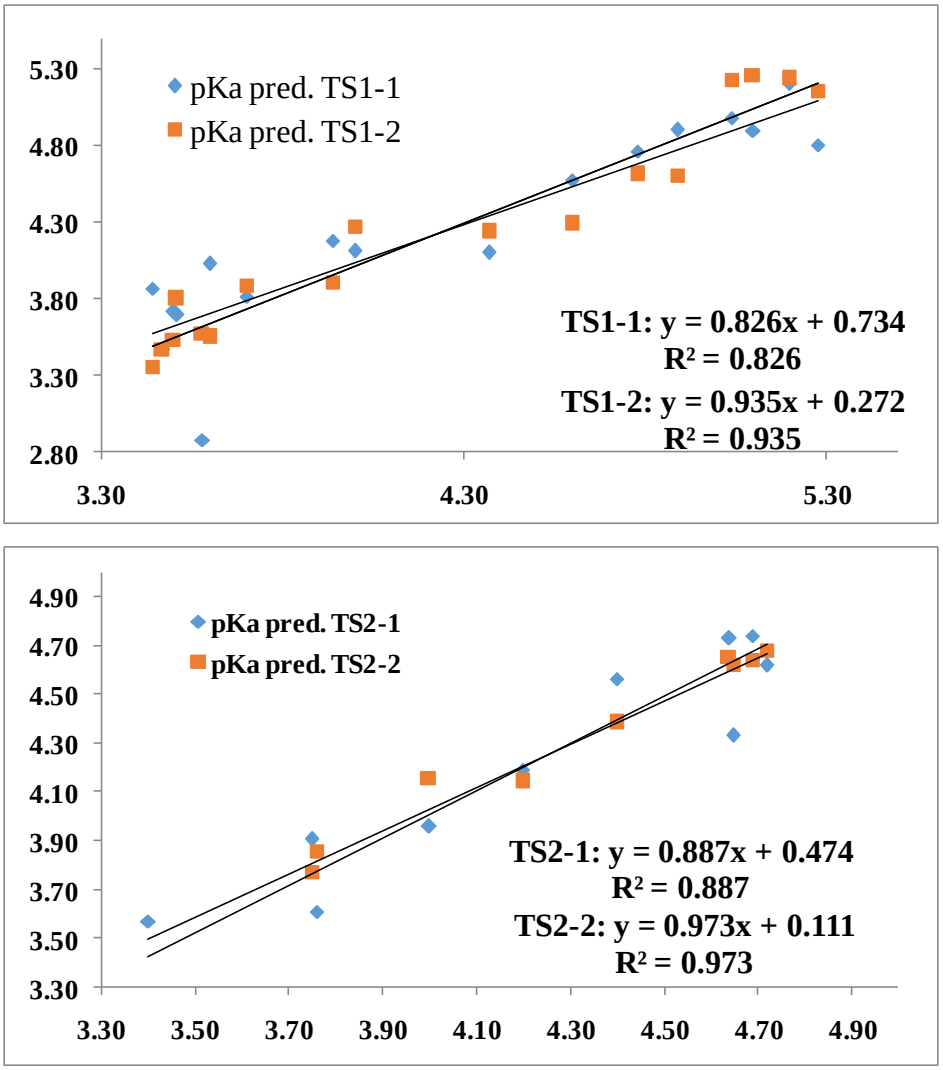

Figure 3. Plot of observed versus predicted activities for Models:
a) TS1-1 and TS1-2; b) TS2-1 and TS2-2

After summarizing all the results, it can be concluded that better models are obtained for structurally similar herbicides (set 1 and set 2) by using four parameter models: Model TS1-2 and TS2-2. These statistically significant QSAR models can be used to design new structurally similar herbicides, which will satisfy at the same time the requirements for the better herbicidal activity and reduced or no toxicity at all.

CONCLUSION

In this study, the QSAR method was used to predict the acidity of a set of sulfonylureas (SU) herbicides. Two different approaches were performed – set with all selected herbicides and set divided according to the structural similarity: pyrimidinyl- and triazinyl-sulfonylurea herbicides.

Using a SwissADME online tool *in silico* predicting human intestine and brain permeation indicated that a 92.59% of studded herbicides are predicted to be poorly absorbed by the human gastro-intestinal tract. Only sulfometuron methyl and chlorsulfuron have a high possibility of passive absorption by the gastrointestinal tract.

The QSAR models validity has been established through the selection of appropriate statistical parameters such as coefficient of correlation (R^2), adjusted coefficient of correlation (R^2adj.), mean squared error (MSE), root mean square error (RMSE) and Fischer test (F -test). According to the values of those parameters, models with 4 descriptors in both sets of herbicides were statistically better.

REFERENCES

Bempelou, E., Kappatos, P., & Liapis, K. (2019). Determination of Eight Sulfonylurea Herbicide Residues by LC/MS/MS Using a Sample Separation Technique with Ethyl Acetate. *J. AOAC Int.*, *102*, 239-245.

CDC - Centers for Disease Control and Prevention. Biomonitoring Summary -Substituted Urea Herbicides (2023, March). https://www.cdc.gov/biomonitoring/

Chedik, L., Mias - Lucquin, D., Bruyere, A., & Fardel, O. (2017). In Silico Prediction for Intestinal Absorption and Brain Penetration of Chemical Pesticides in Humans. *Int. J. Environ. Res. Public Health*, *14*, 708.

Daina, A., Michielin, O., & Zoete, V. (2017). SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.*, *7*, 42717.

Daina, A., & Zoete, V. (2016). BOILED-Egg to predict gastrointestinal absorption and brain penetration of small molecules. *Chem. Med. Chem.*, *11*, 1117-1121.

Devlin, D., Barden, C., Barnes, P., Bernards, M., Franti, T., Milner, M., Regassa, T., Shea, P., & Wortmann, C. (2008). *Pesticide Management for Water Quality Protection in the Midwest*, Regional Water Coordination Initiative, Kansas State University.

İslamoğlu, F., & Hacifazlioğlu, E. (2022). Investigation of the Usability of Some Triazole Derivative Compounds as Drug Active Ingredients by ADME and Molecular Docking Properties. *Moroccan Journal of Chemistry, 10*, 861-880.

Liu, C., Zhang, T., Yang, X., Wang, L., Long, Y., Hasi, A., Pei, X. (2023). A Lu ALS Mutation with High Sulfonylurea Herbicide Resistance in Linum usitatissimum L. *Int. J. Mol. Sci.*, *24*, 2820.

Losito, I., Amorisco, A., Carbonara, T., Lofiego, S., & Palmisano, F. (2006). Simultaneous determination of phenyl- and sulfonyl-urea herbicides in river water at sub-parts-per-billion level by on-line preconcentration and liquid chromatography–tandem mass spectrometry. *Anal. Chim. Acta, 575*, 89-96.

Marchand, P. A. (2023). EU Chemical Plant Protection Products in 2023: Current State and Perspectives. *Agrochemicals*, *2*, 106-117.
Mostafalou, S., & Abdollahi, M. (2017). Pesticides: An update of human exposure and toxicity. *Arch. Toxicol*., *91*, 549–599.
Pathak, V. M., Verma, V. K., Rawat, B. S., Kaur, B., Babu, N., Sharma, A., Dewali, S., Yadav, M., Kumari, R., Singh, S., Mohapatra, A., Pandey, V., Rana, N., & Cunill, J. M. (2022). Current status of pesticide effects on environment, human health and it's eco-friendly management as bioremediation: A comprehensive review. *Front. Microbiol., 13*, 962619.
Pereira, V. J., Cunha, J. P. A. R., Morais, T. P., Ribeiro - Oliveira, J. P., & Morais, J. B. (2016). Physical-chemical properties of pesticides: concepts, applications, and interactions with the environment. *Bioscience Journal, 32*, 627-641.
PPDB: Pesticide Properties DataBase (2023, March) Agriculture and Environment Research Unit, School of Life & Medical Sciences, University of Hertfordshire, College Lane, Hatfield, Hertfordshire, UK. http://sitem.herts.ac.uk/aeru/ppdb
PubChem - Open chemistry database at the National Institutes of Health (2023, March). https://pubchem.ncbi.nlm.nih.gov
Sadeghi, F., Afhami, A., Madrakian, T., & Ghavami, R. (2022). QSAR analysis on a large and diverse set of potent phosphoinositide 3-kinase gamma (PI3Kγ) inhibitors using MLR andANN methods. *Scientifc Reports*, *12*, 6090.
Sjogren, E., Abrahamsson, B., Augustijns, P., Becker, D., Bolger, M. B., Brewster, M., Brouwers, J., Flanagan, T., Harwood, M., Heinen, C., Holm, R., Juretschke, H. P., Kubbinga, M., Lindahl A., Lukacova, V., Münster., U, Neuhoff., S, Nguyen, M. A., Peer, A., Reppas, C., Hodjegan, A. R., Tannergren, C., Weitschies, W., Wilson, C., Zane, P., Lennernäs, H., Langguth, P., & Heinen, C. (2014). In vivo methods for drug absorption - Comparative physiologies, model selection, correlations with in vitro methods (IVIVC), and applications for formulation/API/excipient characterization including food effects. *Eur. J. Pharm. Sci*., *57*, 99–151.
Toropov, A. A., Toropova, A. P., Beeg, M., Gobbi, M., & Salmona, M. (2017). QSAR model for blood-brain barrier permeation. *J. Pharmacol. Toxicol. Methods*, *88*, 7-18.
Wildman S. A., & Crippen, G. M. (1999). Prediction of Physicochemical Parameters by Atomic Contributions. *J. Chem. Inf. Comput. Sci.*, *39(5)*, 868-873.
XLSTAT – Version 5.03 - Statistical Software for Excel (2014).