EVALUATION OF THE MODEL PREDICTION TOXICITY (LD50) FOR SERIES OF 42 ORGANOPHOSPHORUS PESTICIDES

HANANE FIKRI[*1], TAOUFIQ FECHTALI[1], MOHAMED MAMOUMI[1]

[1]Laboratory of Neurosciences, Integrated Physiopathology and Natural Substances - F.S.T. Mohammedia; BP:146 Mohammedia 20650, Morroco

Abstract: Structure-Toxicity Relationships have been studied for a set of 42 organophosphorous pesticides (OPs) through multiple linear regression (MLR) and artificial neural networks (ANN). A model with three descriptors, including: total lipophilicity [log (P)], widths radicals R_1 [(LR_1)] and R_2 [(LR_2)] has achieved good results in phase Training and phase prediction of toxicity [log LD50 (lethal dose 50, Oral rat)]. The linear model (MLR: n=40, r²=0.86, s=40 and $q^2 = 0.66$) and non-linear model with a configuration [3-6-1] (ANN: r²=0.95, s=0.73 and $q^2 = 0.17$) have proved very successful and complementary. The selected descriptors indicate the importance of lipophilicity and widths radicals R_1 and R_2 in the contribution of the toxicity of pesticides derived from OPs used in this study. This information is relevant for the design of a new model of non-toxic pesticides OPs.

Keywords: multiple linear regression (MLR), artificial neural networks (ANN), organophosphorous pesticides (OPS), LD50, descriptors

1. INTRODUCTION

The pesticide is a limit generic for a variety of classes such as chemical herbicides, fungicides and nematicides. The techniques of computer simulation are most effective means potentially offering alternatives to probe structure-toxicity relations. The objectives of our study are to provide additional information on the behaviour of organophosphorus compounds and set in the future the necessary criteria for designing a model for a new generation of organophosphorous pesticides.

The use of quantitative relationships structures QSAR activities currently has considerable attention [1] for pharmaceutical needs [2], as well as the study of the toxicological mechanisms of chemical environmental pollutants (products Endocrine disrupting phytochemicals) [3]. Today, a large number of families of compounds have already been the subject of such research. Among these families are organophosphorus pesticides.

Pesticides especially the organophosphorous (OPs) are the most used in the world of agricultural production. Of these, 70% are highly toxic, although they have a low persistence and are easily biodegradable, they are subject to ecological concerns because they are toxic to non-target species even at low concentrations. [4].

In recent years, chemists and biologists have paid great attention to quantitative structure-activity relationships, known as QSAR. The development of such relationships successfully predicts certain properties and activities of chemical structures without recourse to synthesizing or testing them [5].

Indeed, one of the major concerns of scientists is to have a pesticide that is more effective and less toxic. Only the achievement of these objectives will guarantee the development of the pesticide market. It is within the framework of this perspective that our approach, which allows not only to distinguish between the variations of the toxicity between the different pesticide molecules, but also to establish an effective model on a data bank of 42 organophosphorus pesticides.

2. EXPERIMENTAL SETUP

2.1. Experimental data
In order to establish our model of quantitative structure-toxicity relationship, we collected a sample, as wide possible, of 42 organophosphorous compounds described in FOOTPRINT's Pesicides Properties Database [6], with activity values Specific ecotoxicologicals of acute oral toxicity for rats (mammalian test organisms).

All these compounds have the motif [P (o) (o)] as a common chemical structure (Figure 1).

The chemical structures of the series of the organophosphorous compounds studied are grouped in Table 1 with the logLD50 values used as a dependent variable.
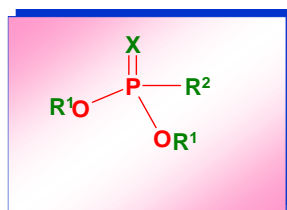

Fig. 1. General Structure of the organophosphorus pesticides:
R^1 = methyl or ethyl; R^2 = alkyl, heterocyclic or aromatic groupment; X = O or S.

Table 1. Chemical structures of the compounds studied and experimental values of logLD50 (lethal dose 50, oral, rat).

N	X	R_1	R_2	$\log LD_{50}$ (Exp)	N°	X	R_1	R_2	$\log LD_{50}$ (Exp)
1	S	C_2H_5		**1.08**	11	S	CH_3		3.45
2		CH_3		**0.95**	12	S	C_2H_5		1
3	S	CH(CH₃)₂		**2.43**	13	O	CH_3		1.48
4	S	CH_3		**3.20**	14	S	C_2H_5		1.85
5	S	C_2H_5		**1.72**	15	S	C_2H_5		2.23
6	S	C_2H_5		**1**	16	O	CH_3		1.90

N	X	R_1	R_2	log LD_{50} (Exp)	N°	X	R_1	R_2	log LD_{50} (Exp)
7	S	C_2H_5	$Cl_3C-CH-O$	**1.64**	17	O	CH_3	(structure)	1.23
8	O	C_2H_5	$Cl-HC=C-O$ (dichlorophenyl)	**1.08**	18	O	CH_3	(structure)	2.40
9	S	C_2H_5	$Cl-CH_2-S$	**0.84**	19	S	C_2H_5	$CH_3-H_2C-S-CH_2-CH_2-S$	1.59
10	S	C_2H_5	(trichloropyridine)	**1.82**	20	S	C_2H_5	$-S-CH_2-S-P(=S)(OCH_2CH_3)_2$	2.32
21	O	CH_3	(ethylpyrimidinone)	**3.25**	32	O	CH_3	$-S-CH_2-C(=O)-HN-CH_3$	1.81
22	S	CH_3	(nitro-methylphenyl)	**2.52**	33	O	CH_3	$-S-CH_2-CH_2-S(=O)-C_2H_5$	1.68
23	S	CH_3	(methylthio-methylphenyl)	**2.30**	34	S	CH_3	$-S-CH(C_6H_5)-C(=O)-O-C_2H_5$	2.40
24	S	CH_3	$-S-CH_2-C(=O)-N(CH_3)-CHO$	**2.56**	35	S	C_2H_5	$-S-CH_2-S-C_2H_5$	0.66
25	O	CH_3	(chloro bicyclic structure)	**1.98**	36	S	CH_3	$-S-CH_2-$ (phthalimide)	2.05
26	S	C_2H_5	(triazole, $CH(CH_3)_2$, Cl)	**1.43**	37	S	C_2H_5	$-S-CH_2-C(=O)-HN-CH(CH_3)_2$	0.90
27	S	CH_3	(dichloro-iodophenyl)	**3,37**	38	S	C_2H_5	(pyrimidine, $N(C_2H_5)_2$, CH_3)	2,15
28	S	CH_3	$-S-CH(C(=O)-O-C_2H_5)-C(=O)-O-C_2H_5$	**3.07**	39	S	CH_3	(pyrimidine, $N(C_2H_5)_2$, CH_3)	3.15
29	S	C_2H_5	$-S-CH_2-C(=O)-N(CH_3)-C(=O)-O-C_2H_5$	**1.56**	40	S	C_2H_5	$-O-P(=S)(O-C_2H_5)-O-C_2H_5$	1
30	O	CH_3	$H_3C-O-C(=CH)-C(=O)-N(CH_3)$	**1.23**	41	S	CH_3	$S-CH_2-CH_2-S-C_2H_5$	1.60
31	O	CH_3	$-O-C(Cl)-C(Cl)(Br)_2$	1.96	42	S	C_2H_5	$-O-$ (triazole-phenyl)	1.82

2.2. Analysis by Multiple Linear Regression (MLR)

A total of 63 physicochemical descriptors were generated to code each molecule by the molecular modeling pro program [7].

The experimental data, containing all the values by n compounds (n = 42), was subjected to Multiple Linear Regression analysis [8]. We used two main approaches in the MLR: the stepwise regression approach [9]; and the "backward" approach, [10]. The contributions of the relevant descriptors are calculated using the Gore

method [11, 12], which excludes a single example of all the data for building the entire test. leave-one-aut .The predictive power of the model has been evaluated by the cross-validation method.

3. RESULTS AND DISCUSSION
3.1. Establishment of MLR models
Given the large number of 63 descriptors used to code each molecule, we subjected our data to Stepwise stepwise selection [14, 15, 16], in order to highlight the most relevant descriptors. The two steric parameters width of radical R1 [L(R1)], and radical R2 [L(R2)], and lipophilicity [Log P], represent the relevant descriptors in the determination of the toxicity of these organophosphorus pesticides studied. The best model obtained by the multiple linear regression is the following (Equation 1).

$$\text{Log (LD}_{50}) = (3.858 \pm 0.790) + (0.081 \pm 0.023)*\text{LogP} -(0.938 \pm 0.133)* [L(R_1)] + (0.257 \pm 0.062)* [L(R_2)]$$
$$n = 42; r = 0.84 \ (r^2 = 064); s = 0.45; F = 29.52$$

(1)

The contributions of the parameters [L(R1)], [L (R2)] and Log P are respectively 45.78%, 33.33 and 20.89%. The model is statistically significant and accounts for up to 63% of initial information.

The residual standard deviation is relatively small (s = 0.45), it is of the order of the error generally committed in biological tests of this type. The large contribution of the width of the substituent R1 reflects the great influence of the size of this substituent in particular on the toxicity of organophosphorus pesticides.

To get closer and closer to the experimental error that is generally of the order of 5% and to check the prediction limits in our sample, we obtained a new sample of 40 molecules. The application of multiple linear regression to this sample leads to equation (2) after elimination of molecules number 39 and number 40 (Equation 2).

$$\text{Log (LD}_{50}) = (1.941 \pm 0.064) + (0.360 \pm 0.082)*\text{LogP} - (0.581 \pm 0.077)* [L(R_1)] + 0.273(\pm 0.070)* [L (R_2)]$$
$$n = 40; r = 0.86 \ (r^2 = 0.728); s = 0.40; F = 32.18$$

(2)

The statistical quality of the equation is very good, it explains 73% of the total variance, and it's higher than the other models described in the literature if we take into account the number of descriptors used. It explains up to 73% of the total variance with a standard error "s" much lower than the average error made on the observed values of Log (LD50) which is of the order of 0.741 for an interval ranging from 0.66 to 3.45. To show that our model is not due to chance, we applied the experiment of changing the column of the dependent variable randomly so that each molecule does not find its true activity but the activity of a another molecule, without touching the columns of the independent variables.

The result of this test on the sample of 42 molecules shows that the statistical quality of our model decreases very remarkably, it goes from r = 0.86, s = 0.40 and F = 32.18 at r = 0.46, s = 0.60 and F = 0.07. This result clearly indicates that the descriptors selected for this study describe well the activity of the series of organophosphorus compounds.

3.2. Establishment Of non-linear models
In order to improve the linear model obtained and test the possibility of non-linear effects that may possibly exist between the activity and the descriptors of the linear model, we used the artificial neural network. The neural network employed therefore has the architecture [3-6-1], (tree neurons in the input layer, six neurons in the hidden layer and a neuron in the output layer).

This preliminary study (Table 2) allowed us to conclude that the neural network with the architecture [3-6-1] schematized in Figure 2 is able to establish a satisfactory relationship between the relevant descriptors and the Log activity (LD50). We observe an improvement in the statistical parameters of the non-linear model (r=0.95 and s = 0.17) compared to the linear model (r = 0.86 and s = 0.40), suggests the possible existence of non-linear relations between the activity and the descriptors. The evaluation of the contribution of the relevant descriptors gives the following classification: LogP> L(R1)> L (R2).

Table 2. Variation of r^2 and s with number of neurons.

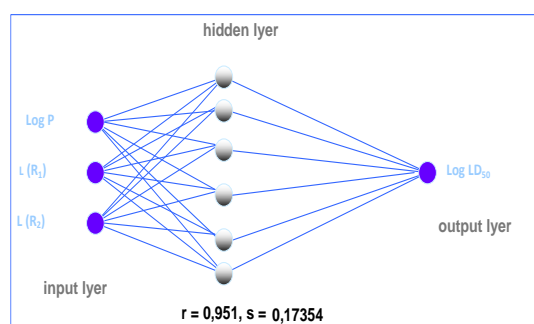| NNC	s	r
2	0.27564299	0.75056049
3	0.25787559	0.82837559
4	0.1948047	0.92550181
5	0.17503741	0.94251181
6	0.17354904	0.95147377
7	0.17417838	0.94901349

Fig. 2. Description of architectural neural network.

The comparison of this classification with that obtained in the linear model reveals a change of order between Log P, L (R1) and L (R2). This change can be explained by the possible existence of a non-linear relationship between activity and lipophilicity, which is not the case for widths.

4. CONCLUSION

Quantitative Structure Activity Relationship (QSAR) models are expected to play an important role in the risk assessment of chemicals on humans and the environment [17]. Because some organophosphorus compounds are known for their pesticidal properties [18]. We have established a study of the quantitative relationships between the molecular structure and toxicity of a series of organophosphorus pesticides, with the aim of predicting the activity of new products at high levels added value.

QSAR models are expected to play an important role in the risk assessment of chemicals on humans [19]. The most relevant descriptors were evidenced by stepwise linear multiple regression analysis. The analysis of the statistical parameters of models published in the literature shows that our model is more efficient [20, 21, 22]. It generated the steric aspect, expressed both by the widths of the radicals R1 and R2 and the however, the nature of the heteroatom (X) does not seem to have a decisive effect on the activity of organophosphorus pesticides.

The structure-toxicity relationship model produced during this work can be classified as a predictive model that assists in the selective design of new molecules with low toxicity, in the family of organophosphorus compounds.

Indeed, the analysis of the schematized model, based on a molecular description of the organophosphorus compounds, shows that a low toxicity (high LD50) is associated with a low value in [L(R1)], a global lipophilicity and a width [L(R2)] more important. In conclusion, this prediction model makes it possible to avoid experimental tests and provide an immediate result.

For future developments, lazy structure–activity relationships [23] will be established and tested algorithms, semantic web aware web services, and language bindings, which can serve as building blocks for new algorithms and applications. We hope that these facilities will speed up the development cycle of future predictive toxicology applications, and will ultimately lead to improved and more relevant applications in this area.

REFERENCES

[1] Weaver, S., Gleeson, M.P, The importance of the domain of applicability in QSAR modeling, Journal of Molecular Graphics and Modelling, vol. 26, no. 8, 2008, p. 1315–1326.

[2] United States Patent 6503938, 2-phenylindoles as antiestrogenic pharmaceutical agents, PatentStorm, 2003.

[3] McKinlay, R., Plant, J.A., Bell, J.N.B., Voulvoulis, N., Endocrine disrupting pesticides: Implications for risk assessment, Environment International, vol. 34, no. 2, 2008, p. 168–183.

[4] Sánchez-Peña, L.C., Reyes, B.E., López-Carrillo, L., Recio, R., Morán-Martínez, J., Cebrián, M.E., Quintanilla-Vega, B., Organophosphorous pesticide exposure alters sperm chromatin structure in Mexican agricultural workers, Toxicol Appl Pharmacol, vol. 196, no. 1, 2004, p. 108-113.

[5] xxx, United States Patent US042189, one-dimensional QSAR Models, 2005.

[6] xxx, Footprint, pesticide properties database PPDB by University Hertfordshire, project FOOTPRINT financed by UE (FP6-SSP-022704) 2007-2008.

[7] xxx, Molecular modeling pro-Demo TM (M.M.P.), Revision 301 demo published by ChemSW SoftwareTM.

[8] Livingstone, D.J, Manallck, D.T., Statistics using neural networks: chance effects, Journal of Medicinal Chemistry, vol. 36, no. 9, 1993, p. 1295–1297.

[9] xxx, Statgraphics Plus for Windows Version 42/15, 1999.

[10] xxx, MATLAB Version.7, the mat works, Inc, Natck, MA, 2002.

[11] Gore, W.L., Statistical methods for chemical experimentation, Inter-science, New York, U.S.A., 1952, p. 141.

[12] Ferandez, M., Caballero, J., Bayesian-regularized genetic neural networks applied to the modeling of non-peptide antagonists for the human luteinizing hormone-releasing hormone receptor, Journal of molecular graphics and modelling, no. 25, no. 4, 2006, p. 410-422.

[13] Cherquaoui, D., Esseffar, M., Villemin, D., Cence, J.M., Chastrette, M., Zakarya, D., Structure-musk odour relationship studies of tetralin and indan compounds using neural networks, New Journal of Chemistry, vol. 22, no. 8, 1998, p. 839-843.

[14] Zahouily, M., Rhihil, A., Bazoui, H., Sebti, S., Zakarya, D., Structure-toxicity relationships study of a series of organophosphorus insecticides, Journal of Molecular Modeling., vol. 8, no. 5, 2002, p. 168-172.

[15] Zahouily, M., Lzar, A., Elmakssoudi, A., Rakik, J., Elyachi, S., Rayadh, A., QSAR for anti-malarial activity of 2-aziridinyl and 2,3-bis(aziridinyl)-1,4-naphthoquinonyl sulfonate and acylate derivatives, Journal of Molecular Modeling, vol. 12, no. 4, 2005, p. 398-405.

[16] Zahouily, M., Rakik, J., Lazar, M., Bahloui, M.A., Rayadh, A., Komiha, N., Exploring QSAR of non-nucleoside reverse transcriptase inhibitors by artificial neural networks: HEPT derivatives, ARKIVOC, vol. 14, 2007, p. 245-256.

[17] Hamadache, M., Benkortbi, O., Hanin, S., Amrane, A., Khaouane, L., Cherif, M., A Quantitative structure activity relationship for acute oral toxicity of pesticides on rats: validation, domain of application and prediction, Journal of Hazardous Materials, vol. 303, 2016, p. 28-40.

[18] Hardy, B., Douglas, N., Helma, C., Rautenberg, M., Jeliazkova, N., Jeliazkov, V., Nikolova, I., Benigni, R., Tcheremenskaia, O., Kramer, S., Girschick, T., Buchwald, F., Wicker, J., Karwath, A., Gütlein, M., Maunz, A., Sarimveis, H., Melagraki, G., Afantitis, A., Sopasakis, P., Gallagher, D., Poroikov, V., Filimonov, D., Zakharov, A., Lagunin, A., Gloriozova, T., Novikov, S., Skvortsova, N., Druzhilovsky, D., Chawla, S., Ghosh, I., Ray, S., Patel, H., Escher, S., Collaborative development of predictive toxicology applications, Journal of Chemical Information and Modeling, vol. 2, no.1, 2010.

[19] Maunz, A., Helma, C., Prediction of chemical toxicity with local support vector regression and activity-specific kernels, SAR QSAR in Environmental Research, vol. 19, 2008, p. 413-431.

[20] Veselinović, J.B., Nikolić, G.M., Trutić, N.V., Monte Carlo QSAR models for predicting organophosphate inhibition of acetylcholinesterase, SAR and QSAR in Environmental Research, vol. 26, no. 6, 2015, p. 449-460.

[21] Galassi, M., Davies, J., Theiler, J., Gough, B., Jungman, G., Alken, P., Toxicological Sciences, no. 95, 2009, p. 5-12.

[22] Alves, V.M., Muratov, E.N., Zakharov, A., Muratov, N.N., Andrade, C.H., Tropsha, A. Chemical toxicity prediction for major classes of industrial chemicals: Is it possible to develop universal models covering cosmetics, drugs, and pesticides?, Food and Chemical Toxicology, vol. 112, 2018, p. 526-534.

[23] Maunz, A., Gütlein, M., Rautenberg, M., Vorgrimmler, D., Gebele, D., Helma, C., Lazar: a modular predictive toxicology framework, Frontiers in Pharmacology, vol. 4, no. 38, 2013.