

QSARs study & acute toxicity of pesticides

ESTUDIO QSAR DE LA TOXICIDAD AGUDA DE PESTICIDAS EN LA LOMBRIZ *EISENIA FOETIDA*

Pablo R. Duchowicz*^a, Daniel O. Bennardi^b y José F. Aranda^a

^a Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas (INIFTA), CONICET, Universidad Nacional de La Plata (UNLP), Diag. 113 y 64, C.C. 16, Sucursal 4, 1900 La Plata, Argentina.

^b Facultad de Ciencias Agrarias y Forestales, Universidad Nacional de La Plata (UNLP), 60 y 119, B1904AAN La Plata, Argentina.

*Autor Corresponsal: prduchowicz@gmail.com

Recibido: 31 de mayo de 2022

Aceptado: 15 de julio de 2022

Resumen

El objetivo de este trabajo fue el desarrollo de relaciones cuantitativas estructura-actividad (QSAR) predictivas con el fin de predecir la toxicidad aguda de los plaguicidas en lombrices de tierra (*Eisenia foetida*). Los datos de toxicidad experimentales fueron obtenidos de la base de datos de propiedades de pesticidas (PPDB) desarrollada en la Universidad de Hertfordshire. Se dibujaron las estructuras moleculares con ACD/ChemSketch. Los descriptores moleculares no conformacionales se calcularon mediante los programas PaDEL, Mold², DataWarrior y QuBiLS-MAS. Tras la exclusión de los descriptores constantes, linealmente dependientes y de valores únicos se obtuvieron un total de 2055 variables estructurales. Se usó el Método de Subconjuntos Balanceados (BSM) para dividir el conjunto molecular (58 moléculas) en 3 grupos de calibración, validación y predicción. Se llevaron a cabo regresiones lineales QSAR de 1-5 descriptores, mediante un algoritmo matemático, el Método de Reemplazo. Se encontró un modelo de regresión lineal de 4 descriptores que poseía los mejores resultados de validación. Posteriormente se utilizó este modelo para predecir la toxicidad de 21 plaguicidas cuyos valores experimentales eran no verificados o menos confiables. Las predicciones fueron por lo general próximas a los valores reportados no verificados.

Palabras clave: Plaguicidas; *Eisenia foetida*; QSAR; regresión lineal multivariable; validación; Método de Reemplazo.

Abstract

The aim of this work was the development of predictive quantitative structure–activity relationships (QSARs) in order to predict the acute toxicity of pesticides in earthworms (*Eisenia foetida*). Experimental toxicity data were obtained from the Pesticide Properties Database (PPDB) developed at the University of Hertfordshire. Molecular structures were drawn with ACD ChemSketch. Non-conformational molecular descriptors were calculated using PaDEL, Mold², DataWarrior and QuBiLs-MAS programs. After excluding the constant, linearly dependent, and unique value descriptors, a total of 2,055 structural variables were obtained. The Balanced Subset Method (BSM) was used to divide the molecular set (58 molecules) into 3 groups for calibration, validation and prediction. QSAR linear regressions of 1-5 descriptors were carried out using a mathematical algorithm, the Replacement Method. A 4-descriptor linear regression model was found to have the best validation results. This model was then used to predict the toxicity of 21 pesticides whose experimental values were unverified or less reliable. The predictions were generally close to the unverified reported values.

Keywords: Pesticides; *Eisenia foetida*; QSAR; multivariable linear regression; validation; Replacement Method.

1. Introducción

Los plaguicidas se utilizan ampliamente en la agricultura para la protección de las plantas y para aumentar los rendimientos de producción y la calidad de los productos agrícolas. Sin embargo, tienen una gran desventaja, la toxicidad [1]. Como resultado de su uso excesivo, se encuentran como residuos en el medio ambiente, ocasionando un riesgo para los seres humanos, los animales y otros organismos [2,3]. Una de las formas de medir el riesgo para el medio ambiente y los seres humanos es determinar la toxicidad aguda de los plaguicidas. Estos se prueban contra una variedad de animales, incluidos mamíferos, aves, peces e invertebrados [4]. Las lombrices como *Eisenia foetida* pueden usarse con éxito como bioindicadores para la evaluación de la toxicidad aguda de los plaguicidas en los ecosistemas terrestres [5-7]. Se han desarrollado varios protocolos para evaluar los efectos de los plaguicidas en *Eisenia foetida*, entre los cuales el más conocido es la guía 207 de la OCDE, que consiste en una prueba de suelo artificial de 14 días [8,9].

Lamentablemente, la determinación experimental de la toxicidad, además de requerir un tiempo considerable, es costosa y genera un dilema ético (demandas para reducir o abolir las pruebas en animales) [4,10]. El uso de métodos predictivos basados en

herramientas computacionales es una opción rápida, económica y ética para evaluar la toxicidad de los plaguicidas en animales [11,12]. Estos métodos comprenden las relaciones cuantitativas estructura-actividad (QSAR).

La teoría QSAR establece una relación cuantitativa entre la estructura molecular y el efecto final como la actividad biológica, lo que permite predecir la bioactividad de compuestos con datos experimentales desconocidos. Estos cálculos basados en la estructura molecular consideran el empleo de descriptores moleculares, que son variables numéricas que contienen información específica y relevante sobre las características constitucionales, topológicas, geométricas, hidrofóbicas y/o electrónicas de los compuestos investigados [13-16].

Hasta ahora, se ha desarrollado una gran cantidad de modelos QSAR para predecir la toxicidad aguda de los plaguicidas [10, 17-21]. En un reciente estudio QSAR informado por Ghosh *et al.* [22], los autores desarrollaron un modelo de regresión lineal para predecir la toxicidad de los plaguicidas en las lombrices de tierra *Eisenia foetida*, empleando el método de Mínimos Cuadrados Parciales (PLS). El conjunto de plaguicidas que utilizaron fue dividido en dos grupos: conjunto de calibración o training set (43 compuestos) para el desarrollo del modelo, y conjunto de prueba o test set (14 compuestos) para validación del mismo. Calcularon un total de 501 descriptores moleculares empleando los programas Dragon 6 y PaDEL-Descriptor. El mejor modelo que obtuvieron contiene ocho descriptores, es robusto y predictivo.

En este trabajo proponemos un modelo alternativo QSAR de regresión lineal multivariable (MLR) para la predicción de la toxicidad aguda de los plaguicidas en *Eisenia foetida*. Para el desarrollo del mismo empleamos descriptores moleculares independientes de la conformación. Estos tienen la ventaja de que se basan únicamente en los aspectos constitucionales y topológicos de las moléculas. El propósito de nuestro estudio es contribuir al cuidado del medio ambiente brindando una guía racional para el diseño de plaguicidas con menor toxicidad.

2. Metodología

Los datos de toxicidad experimentales fueron obtenidos de la base de datos de propiedades de pesticidas (PPDB) desarrollada en la Universidad de Hertfordshire [23]. La actividad fue expresada como concentración letal media (CL_{50}): corresponde a la concentración (mg.kg^{-1}) de pesticida que conduce a la muerte del 50% de la lombriz *Eisenia foetida*. Los valores de toxicidad aguda fueron convertidos en $\log CL_{50}$. Se

extrajeron de la base de datos PPDB 79 pesticidas pertenecientes a diferentes familias químicas y con distintos usos, dividiéndoselos en 2 grupos. El primero estuvo constituido por 58 compuestos que incluían datos verificados por la base PPDB y por lo tanto tenían mayor grado de confianza. Este conjunto se utilizó para establecer el modelo QSAR. El segundo grupo de pesticidas estuvo formado por 21 compuestos con datos no verificados o menos confiables y fue utilizado como conjunto de ensayo. En la **Tabla 1X** del apéndice se incluyen los detalles de los compuestos estudiados, algunos de los cuales se muestran en la **Fig. 1**.

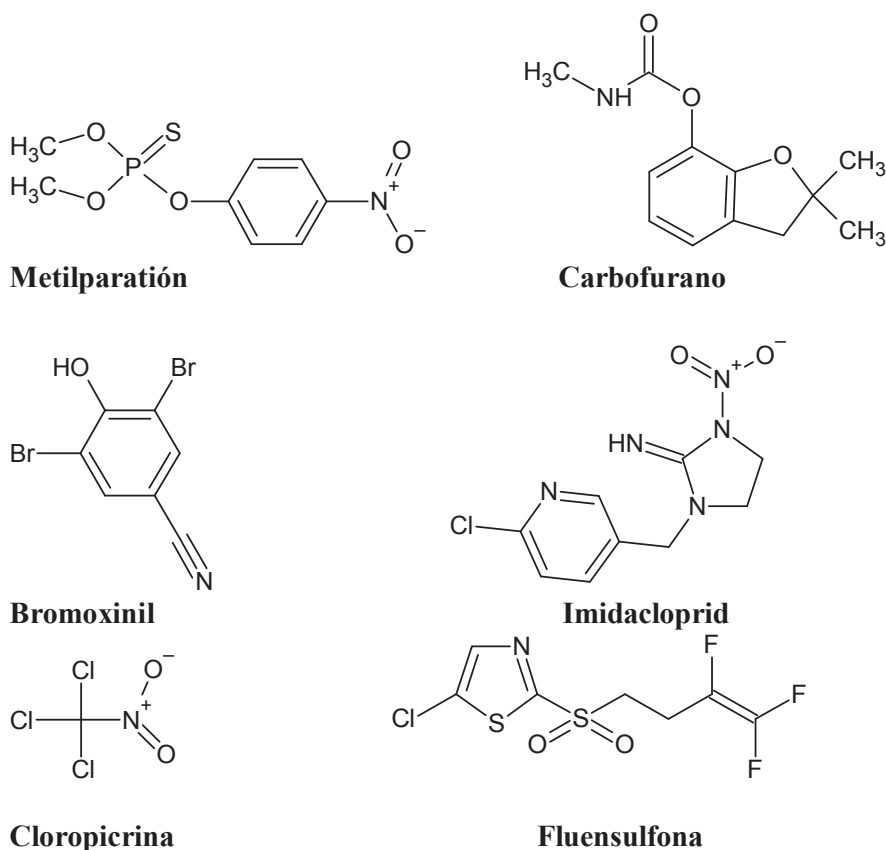


Figura 1. Estructuras moleculares pertenecientes a algunos de los pesticidas estudiados.

Se dibujaron las estructuras químicas con ACD ChemSketch [24] con el formato molecular en MDL mol (V2000). La conversión de las moléculas se realizó con el programa para Windows Open Babel [25].

Los descriptores moleculares se calcularon mediante los programas gratuitos y de libre acceso: PaDEL [26,27], Recon [28], Mold² [29,30], DataWarrior [31] y QuBiLS-MAS [32,33]. Se obtuvieron un total de 23971 descriptores independientes de la

conformación. Esta cantidad se redujo a 2055 tras la exclusión de los descriptores constantes, casi constantes, y linealmente dependientes.

Se efectuó la partición del conjunto molecular ($N = 58$ moléculas) en tres subconjuntos. Uno de estos, calibración ($N_{cal} = 41$), se empleó para seleccionar los mejores descriptores del modelo MLR. El subconjunto validación ($N_{val} = 8$) se usó para evitar la presencia de sobreajuste del modelo durante la selección de descriptores efectuada. Mientras que el de predicción ($N_{pred} = 9$) se utilizó para determinar la capacidad predictiva del modelo (validación externa). Para llevar a cabo la partición mencionada se utilizó la técnica “Método de Subconjuntos Balanceados” (BSM) [34-36].

Se llevaron a cabo regresiones lineales QSAR de 1-5 descriptores, mediante un algoritmo matemático basado en la técnica MLR, el “Método del Reemplazo” (RM) [37,38], que permite seleccionar los mejores descriptores a partir un gran número de variables, de manera que se minimice la raíz cuadrada de la desviación cuadrática media de calibración ($RMSE_{cal}$). Aparte de utilizar un conjunto de validación externo, el mejor modelo obtenido fue validado internamente por medio de las técnicas validación cruzada ‘dejar-uno-afuera’ (*loo*) y aleatorización-y (*rand*) [39-42].

Para estimar la confiabilidad de las predicciones QSAR, se determinó el dominio de aplicabilidad (AD) del modelo a través del enfoque de influencia [43-45], por el cual el conjunto de predicción debe tener un valor de influencia calculado (h) menor que el valor de influencia de control (h^*). De esta manera la propiedad predicha no es el resultado de una extrapolación sustancial (predicción poco confiable).

Todos los algoritmos de Octave [46] se desarrollaron en nuestro grupo y están disponibles.

3. Resultados y discusión

Las mejores regresiones lineales QSAR de $d = 1-5$ descriptores, obtenidas mediante RM, involucran a los descriptores de la **Tabla 1**.

Tabla 1. Modelos lineales para la toxicidad aguda en *Eisenia foetida*.

D	Descriptores	R^2_{cal}	$RMSE_{cal}$	R^2_{val}	$RMSE_{val}$	R^2_{pred}	$RMSE_{pred}$
1	<i>D017</i>	0,21	0,69	0,14	0,72	0,54	0,44
2	<i>GATS4m; APC8NX</i>	0,48	0,56	0,09	0,65	0,60	0,48

3	<i>GATS4m; SHBint4; APC8NX</i>	0,61	0,48	0,67	0,40	0,60	0,42
4	<i>GATS4m; nBondsD; SHBint4; APC8NX</i>	0,72	0,41	0,70	0,38	0,87	0,31
5	<i>HybRatio; VE1D; MCS133; PC590; APC8NX</i>	0,82	0,33	0,07	0,88	0,76	0,31

Al tratarse de pocas moléculas estudiadas, el mejor modelo elegido de la **Tabla 1** debe cumplir con la relación práctica $N_{cal}/d \geq 6$ [47]; una relación lineal de 4 descriptores cumple con este requisito y posee además los mejores resultados de validación (mayor R^2_{val} y menor $RMSE_{val}$) evitando la presencia de sobreajuste del modelo:

$$\log CL_{50} = -0,56GATS4m + 0,15 nBondsD - 0,17 SHBint4 - 0,88APC8NX + 2,36 \quad (1)$$

$$N_{cal} = 41, \quad R^2_{cal} = 0,72, \quad RMSE_{cal} = 0,41$$

$$R^2_{ij\ max} = 0,03, \quad o_{2,5} = 0, \quad R^2_{rand} = 0,52, \quad RMSE_{rand} = 0,54 \text{ (100000 casos)}$$

$$R^2_{loo} = 0,66, \quad RMSE_{loo} = 0,46$$

$$N_{val} = 8, \quad R^2_{val} = 0,70 \quad RMSE_{val} = 0,38$$

$$N_{pred} = 9, \quad R^2_{pred} = 0,87, \quad RMSE_{pred} = 0,31$$

R^2_{cal} , R^2_{val} , R^2_{pred} son los coeficientes de determinación correspondientes a los conjuntos de calibración, validación y predicción. R^2_{loo} , R^2_{rand} son los coeficientes de determinación obtenidos mediante las técnicas de validación cruzada ‘dejar-uno-afuera’ (*loo*) y aleatorización-y (*rand*). $RMSE_{cal}$, $RMSE_{val}$, $RMSE_{pred}$ son las raíces cuadradas de la desviación cuadrática media correspondientes a los conjuntos de calibración, validación y predicción. $RMSE_{loo}$, $RMSE_{rand}$ son las raíces cuadradas de la desviación cuadrática media obtenidas mediante las técnicas de validación cruzada ‘dejar-uno-afuera’ (*loo*) y aleatorización-y (*rand*).

$R^2_{ij\ max}$ es el cuadrado del coeficiente de correlación máximo entre pares de descriptores del modelo [48]. El bajo valor obtenido ($R^2_{ij\ max} = 0,03$) revela que no existe superposición de información estructural entre ellos.

La relación cuantitativa estructura-actividad representada por la ecuación 1 (**Ec. 1**) cumple con la técnica aleatorización-y, ya que después de 100000 casos de aleatorización de la actividad experimental, los parámetros alcanzados son $R^2_{cal} > R^2_{rand}$

y $RMSE_{cal} < RMSE_{rand}$ para el peor caso posible de aleatorización, indicando la ausencia de correlación casual. El modelo es estable y predictivo puesto que aprueba los parámetros estadísticos de validación cruzada ($R^2_{loo} > 0,5$) [49] y externa (valor mínimo de $R^2_{val} = 0,7$ y $R^2_{pred} > 0,7$) [50-52], con valores ligeramente mejores a los obtenidos en el estudio realizado por Ghosh *et al.* [22] ($R^2_{loo} = 0,614$ y $R^2_{pred} = 0,734$). Cabe destacar que la **Ec. 1** es un modelo más simple ya que involucra 4 descriptores en lugar de los 8 del citado trabajo. Asimismo se cumplen las siguientes condiciones establecidas por Golbraikh y Tropsha [49] para establecer el poder predictivo del modelo:

$$1 - R^2_0 / R^2_{pred} < 0,1 \text{ (0,024)}; \quad 1 - R^2_0 / R^2_{pred} < 0,1 \text{ (3,63.10}^{-5}\text{)};$$

$$0,85 \leq k \leq 1,15 \text{ (1.11)}; \quad 0,85 \leq k' \leq 1,15 \text{ (0,89)}$$

Donde k y k' son las pendientes de las rectas de regresión a través del origen de las actividades predichas *versus* experimentales y viceversa, respectivamente y R^2_0 y R^2_0 son los coeficientes de correlación elevados al cuadrado asociados.

En las **Figs. 2 y 3** se grafican los resultados de las predicciones de $\log CL_{50}$ y dispersión de residuos. Se observa que con estos 4 descriptores moleculares se describe de manera aceptable a los conjuntos de validación y predicción. Según sugieren estas figuras, ningún compuesto del conjunto de calibración posee un residuo mayor a 2,5 S_{cal} .

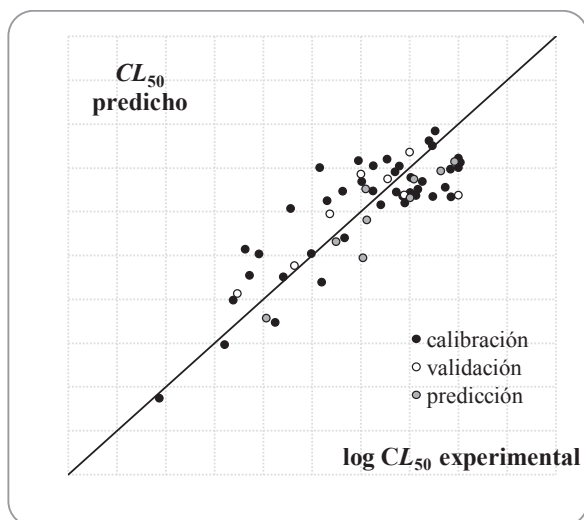


Figura 2. Valores de $\log CL_{50}$ predichos *versus* experimentales según el modelo QSAR.

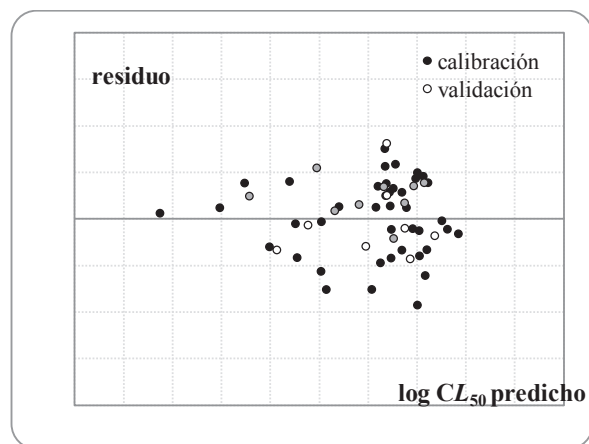


Figura 3. Gráfico de dispersión de residuos para el modelo QSAR.

Los 4 descriptores no-conformacionales se clasifican de la siguiente manera:

- un descriptor de autocorrelación 2D: $GATS4m$, autocorrelación de Geary-distancia 4 / ponderada por la masa
- un descriptor constitucional: $nBondsD$, el número de enlaces dobles
- un descriptor electrotopológico: $SHBint4$, suma de descriptores de fuerza del estado-E para enlaces hidrógeno potenciales de camino de longitud 4
- un descriptor indicador de pares de átomos: $APC8NX$, la cuenta de enlaces N-X a la distancia topológica 8

Los valores numéricos de los descriptores de los pesticidas se incluyen en la **Tabla 2X**.

Dado que un pesticida es más tóxico en la lombriz *Eisenia foetida* cuanto menor sea su valor de concentración letal media, es posible proponer la siguiente guía QSAR para la búsqueda de estructuras con valores determinados de esta propiedad. Según el signo de los coeficientes de regresión de la **Ec. 1**, cuanto mayores sean simultáneamente los valores numéricos de $GATS4m$, $SHBint4$, $APC8NX$ y menor el valor de $nBondsD$, mayor toxicidad aguda presentará el pesticida.

Con respecto al dominio de aplicación de la **Ec. 1**, en la **Fig. 4** están graficados los residuos estandarizados en función del valor de influencia. Puede verse que solo una única molécula de predicción posee un valor de influencia ligeramente superior al límite h^* : imidacloprid, aunque puede considerarse que su predicción es confiable por situarse cerca de ese límite.

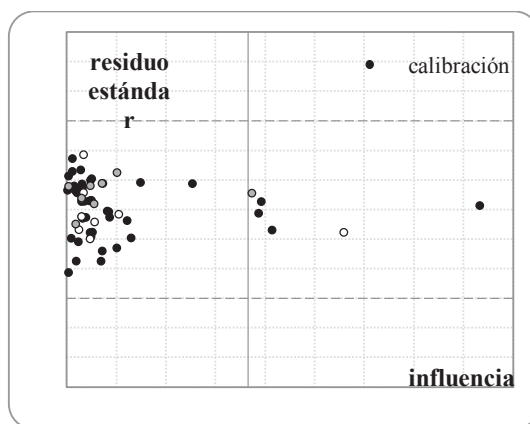


Figura 4. Gráfico de Williams para la Ec. 1. Valor de influencia límite $h^* = 0,366$.

El siguiente paso del estudio fue utilizar el modelo QSAR para predecir el conjunto de ensayo de 21 plaguicidas con datos no verificados o menos confiables. Según la **Tabla 1X**, las predicciones de las moléculas **59-79** son por lo general próximas a los valores reportados no verificados. Solo un plaguicida está fuera del dominio de aplicación de la **Ec.1** (aclonifeno), mientras que dos moléculas poseen predicciones que se diferencian más del valor reportado no verificado (1,2-dicloropropano y tiobencarb).

4. Conclusiones

A partir de datos verificados y confiables de la base de datos PPDB, se consiguió establecer un modelo QSAR de regresión lineal multivariable para la predicción de la toxicidad aguda de plaguicidas (pertenecientes a diferentes familias químicas y aplicaciones diversas) en *Eisenia foetida*. Para ello se emplearon descriptores no conformacionales calculados a través de programas gratuitos y de libre acceso. El modelo obtenido supera los resultados publicados respecto a tamaño, robustez y capacidad predictiva. Además, la aplicación del modelo para la predicción de 21 plaguicidas con datos no verificados o menos confiables puede considerarse aceptable. Este estudio constituye una guía QSAR valiosa para la búsqueda de plaguicidas con menor toxicidad, lo cual ayudará a mantener la biodiversidad del suelo a través de la supervivencia de las lombrices.

Agradecimientos

Al Ministerio de Ciencia, Tecnología e Innovación por el acceso a la Biblioteca Electrónica.

Descripción de la información de apoyo

La información de apoyo para este trabajo se presenta en un Apéndice que contiene tablas complementarias. La **Tabla 1X** muestra los valores de log CL₅₀ experimentales y predichos (**Ec. 1**) para los 79 pesticidas. La **Tabla 2X** presenta los valores numéricos de los descriptores del modelo QSAR.

Referencias

1. Wan M.T. *J. Environ. Sci. Health Part B* **2013**; 48, 344–363.
2. Müller K., Tiktak A., Dijkman T., et al. Advances in pesticide risk reduction. En: van Alfen N.K., editor. *Encyclopedia of Agriculture and Food Systems*. USA: Academic Press, **2014**, pp. 17–34.
3. Gopi R.A., Satyavani G., Shanmugasundaram R., et al. *Int. J. Environ. Sci.* **2014**; 4, 1121–1129.
4. Price N.R., Watkins R.W. *Pestic. Outlook* **2003**; 14, 127–129.
5. Zou X., Xiao X., Zhou H., et al. *J. Hazard. Mater.* **2018**; 359, 365–372.
6. Espinoza-Navarro O., Bustos-Obregón E. *Int. J. Morphol.* **2004**; 22, 297–302.
7. Lock K., Schampelaere K., Janssen C. *Arch. Environ. Contam. Toxicol.* **2002**; 42, 217–221.
8. Gao M., Song W., Zhang J., Guo J. *Environ. Toxicol. Pharmacol.* **2013**; 35, 427–433.
9. Pelosi C., Joimel S., Makowski D. *Chemosphere* **2013**; 90, 895–900.
10. Golbamaki A., Cassano A., Lombardo A., et al. *SAR QSAR Environ. Res.* **2014**; 25, 673–694.
11. Sullivan K.M., Manuppello J.R., Willett C.E. *SAR QSAR Environ. Res.* **2014**; 25, 357–365.
12. Gissi A., Gadaleta D., Floris M., et al. *Altex* **2014**; 31, 23–36.
13. Todeschini R., Consonni V. *Molecular Descriptors for Chemoinformatics (Methods and Principles in Medicinal Chemistry)*. Wiley-VCH, Weinheim, Alemania, **2009**.
14. Masand V.H., Rastija V. *Chemom. Intell. Lab. Syst.* **2017**; 169, 12–18.
15. Diudea M.V. *QSPR/QSAR Studies by Molecular Descriptors*. Nova Science Publishers, New York, **2001**.
16. Mauri A., Consonni V., Todeschini R. Molecular Descriptors. En: Leszczynski J., editor. *Handbook of Computational Chemistry*. Springer, Dordrecht, **2016**, pp. 1–29.
17. Mazzatorta P., Smiesko M., Lo Piparo E., Benfenati E. *J. Chem. Inf. Model.* **2005**; 45, 1767–1774.
18. Hamadache M., Benkortbi O., Hanini S., et al. *J. Hazard. Mater.* **2016**; 303, 28–40.
19. Martin T.M., Lilavois C.R., Barron M.G. *SAR QSAR Environ. Res.* **2017**; 28, 525–539.
20. Toropov A.A., Toropova A.P., Marzo M., et al. *Environ. Toxicol. Pharmacol.* **2017**; 53, 158–163.
21. Hamadache M., Benkortbi O., Hanini S., et al. *Environ. Sci. Pollut. Res.* **2018**; 25, 896–907.
22. Ghosh S., Ojha P.K., Carnesecchi E., et al. *Ecotoxicol. Environ. Saf.* **2020**; 190, 110067.
23. Agriculture & Environment Research Unit (AERU) at the University of Hertfordshire, <http://sitem.herts.ac.uk/aeru/ppdb/en/index.htm>. Acceso el 28/7/2022.
24. ACD/ChemSketch, <https://www.acdlabs.com>. Acceso el 28/7/2022.
25. Open Babel for Windows, http://openbabel.org/wiki/Main_Page. Acceso el 28/7/2022.
26. PaDEL, <http://www.yapcwsoft.com>. Acceso el 28/7/2022.
27. Yap C.W. *J. Comput. Chem.* **2011**; 32, 1466–1474.
28. Lavine B.K., Davidson C.E., Breneman C., et al. *J. Chem. Inf. Comput. Sci.* **2003**; 43, 1890–1905.
29. Hong H., Xie Q., Ge W., et al. *Chem. Inf. Model.* **2008**; 48, 1337–1344.
30. Mold², <https://www.fda.gov/science-research/mold2/access-mold2>. Acceso el 28/7/2022.
31. Datawarrior, <http://www.openmolecules.org/datawarrior/>. Acceso el 28/7/2022.
32. Valdes-Martini J.R., Jacas C.R.G., Marrero-Ponce Y., et al. *J. Cheminform.* **2017**; 9, 35–61.

33. QuBiLS-MAS, <http://tomocomd.com/software>. Acceso el 28/7/2022.
34. Rojas C., Duchowicz P.R., Tripaldi P., *et al.* *J. Chromatogr. A* **2015**; 1422, 277–288.
35. Rojas C., Duchowicz P.R., Tripaldi P., *et al.* *Chemom. Intell. Lab. Syst.* **2015**; 140, 126–132.
36. Rojas C., Tripaldi P., Duchowicz P.R. *IJQSPR* **2016**; 1, 76–90.
37. Duchowicz P.R., Castro E.A., Fernández F.M., *et al.* *Chem. Phys. Lett.* **2005**; 412, 376–380.
38. Duchowicz P.R., Castro E.A., Fernández F.M. *MATCH Commun. Math. Comput. Chem.* **2006**; 55, 179–192.
39. Tropsha A., Gramática P., Gombar V.K. *QSAR Comb. Sci.* **2003**; 22, 69–77.
40. Gramática P. *QSAR Comb. Sci.* **2007**; 26, 694–701.
41. Rücker C., Rücker G., Meringer M. *J. Chem. Inf. Model.* **2007**; 47, 2345–2357.
42. Arlot S., Celisse A. *Stat. Surv.* **2010**; 4, 40–79.
43. Eriksson L., Jaworska J., Worth A.P., *et al.* *Environ. Health Perspect.* **2003**; 111, 1361–1375.
44. Jaworska J., Nikolova-Jeliazkova N., Aldenberg T. *ATLA* **2005**; 33, 445–459.
45. Roy K., Kar S., Ambure P. *Chemom. Intell. Lab. Syst.* **2015**; 145, 22–29.
46. GNU Octave, Scientific Programming Language, <https://www.gnu.org/software/octave/index>. Acceso el 28/7/2022.
47. Tute M.S. *History and Objectives of Quantitative Drug Design in Advances in Drug Research*. Academic Press, London, **1971**.
48. Todeschini R., Consonni V., Gramática P. Chemometrics in QSAR. En: R. Tauler, Walczak B., Brown S.D., editors. *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*. Elsevier, **2009**. pp. 129–170.
49. Golbraikh A., Tropsha A. *J. Mol. Graph. Model.* **2002**; 20, 269–276.
50. Chirico N., Gramática P. *J. Chem. Inf. Model.* **2012**; 52, 2044–2058.
51. Gramática P., Cassani S., Roy P.P., *et al.* *Mol. Inf.* **2012**; 31, 817–835.
52. Gramática P. *IJQSPR* **2020**; 5, 1–37.

APÉNDICE a Estudio QSAR de la toxicidad aguda en la lombriz *Eisenia foetida*, Duchowicz *et al.*

Tabla 1X. Valores de log CL₅₀ experimentales y predichos (**Ec. 1**) para 79 pesticidas.

^ conjunto de validación; * conjunto de predicción; ** conjunto de estimación.

N	CAS	Pesticida	Exp.	Ec. 1
1 [^]	2634-33-5	1,2-Benzisotiazolin-3-ona	2,4440	2,1922
2 [*]	542-75-6	1,3-Dicloropropeno	1,7451	1,6590
3 [*]	94-75-7	2,4-D (ácido 2,4-diclorofenoxiacético)	2,5441	2,3732
4	135410-20-7	Acetamiprid	0,9542	1,5186
5	65-85-0	Ácido benzoico	2,5843	2,2573
6	116-06-3	Aldicarb	1,8129	2,2347
7	858956-08-8	Aminociclopiraclor	2,5647	2,1860
8	3337-71-1	Asulam	3,0000	2,6123
9 [^]	22781-23-3	Bendiocarb	2,2742	2,3758
10	1689-84-5	Bromoxinil	1,6532	2,1262
11	95465-99-9	Cadusafôs	0,8573	1,2744
12 [^]	10605-21-7	Carbendazim	0,7324	1,0669
13 [*]	16118-49-3	Carbetamida	2,8195	2,4663
14	1563-66-2	Carbofurano	2,3502	2,4561
15	76-06-2	Cloropicrina	1,5775	2,5043
16	1897-45-6	Clotalonil	2,4289	2,1798

17	999-81-5	Cloruro de clormecuat	2,5051	2,2192
18*	24307-26-4	Cloruro de mepiquat (cloromecuato)	2,5045	2,1601
19	210880-92-5	Clotianidina	1,1209	0,7371
20	533-74-4	Dazomet	0,8129	1,5726
21	1014-69-3	Demetrina	2,2041	2,0803
22	2764-72-9	Dicuat	1,9745	2,5840
23	55290-64-7	Dimetipin	2,0086	2,3452
24	60-51-5	Dimetoato	1,4914	1,5226
25	165252-70-0	Dinotefurano	0,6902	0,9913
26	3347-22-6	Ditianona	2,7619	2,9231
27	534-52-1	DNOC	1,2041	1,2571
28	2439-10-3	Dodina	2,7380	2,1744
29	26225-79-6	Etofumesato	2,1271	2,5256
30	13194-48-4	Etoprofos	1,5977	1,1968
31	122-14-5	Fenitrotión	2,3636	2,2260
32^	74738-17-3	Fenpiclonil	2,0000	2,4305
33	951659-40-8	Flupiradifurona	2,2686	2,6008
34^	298-02-2	Forato	1,3181	1,3857
35	22259-30-9	Formetanato	3,0204	2,5640
36	10004-44-1	Himexazol	2,4501	2,1007
37*	138261-41-3	Imidacloprid	1,0294	0,7859
38	35554-44-0	Imizalil	2,7332	2,7545
39	556-61-6	Metilisotiocianato	0,6021	0,4834
40	58-89-9	Lindano	1,8325	1,7023
41	12427-38-2	Maneb	2,9243	2,1708
42	94-74-6	MCPA	2,5119	2,3918
43	70630-17-0	Mefenoxam	2,9191	2,4865
44*	41394-05-2	Metamitrona	2,9609	2,5734
45	67129-08-2	Metazacloro	2,6990	2,8110
46	16752-77-5	Metomilo	1,2788	2,0375
47	21087-64-9	Metribuzina	2,6304	2,3458
48*	23135-22-0	Oxamilo	2,0492	2,2608
49*	301-12-2	Metil-oxidemetón	2,0607	1,9073
50^	87-86-5	Pentaclorofenato	1,6812	1,9765
51^	106700-29-2	Petoxamida	2,4997	2,6815
52	709-98-8	Propanil	2,8657	2,2798
53	86763-47-5	Propisoclor	2,3945	2,5219
54	175217-20-6	Siltiofam	2,1239	2,2379
55	946578-00-3	Sulfoxaflor	-0,0680	-0,1282
56^	117-18-0	Tecnazén	3,0000	2,1899
57*	111988-49-9	Tiacloprid	2,0212	1,4746
58	62-56-6	Tiourea	3,0000	2,5043
59**	78-87-5	1,2-Dicloropropano	3,6274	1,7347

60**	34256-82-1	Acetoclor	2,0233	2,5037
61**	74070-46-5	Aclonifeno	2,1761	1,0307
62**	3060-89-7	Metobromuron	2,3674	2,4116
63**	94361-06-5	Ciproconazol	2,2253	1,6938
64**	81777-89-1	Clomazona	1,8921	2,2515
65**	1194-65-6	Diclobenil	2,1303	2,0682
66**	87674-68-8	Dimetenamida	2,1673	2,3394
67**	163515-14-8	Dimetenamida-P	2,4689	2,3394
68**	3766-81-2	Fenobucarb	1,0294	2,2901
69**	55-38-9	Fenti3n	2,5740	2,0266
70**	318290-98-1	Fluensulfona	2,1847	2,4915
71**	10265-92-6	Metamidof3s	1,5315	1,8432
72**	1746-81-2	Monolinur3n	3,0000	2,3075
73**	15299-99-7	Napropamida	2,4502	2,5252
74**	90-43-7	orto-Fenilfenol (OPP)	1,9961	2,6384
75**	298-00-0	Metilparati3n	1,6021	2,1009
76**	53112-28-0	Primetanil	2,4955	2,7269
77**	52888-80-9	Prosulfocarb	1,8561	2,0932
78**	28249-77-6	Tiobencarb	2,6405	1,2432
79**	185676-84-0	Verbutina	2,3572	2,1216

Tabla 2X. Valores num3ricos de los mejores descriptores de la toxicidad aguda.

N	GATS4m	nBondsD	SHBint4	APC8NX
1	1,33905289	4	0	0
2	1,49956736	1	0	0
3	1,01792681	4	0	0
4	0,97208074	4	0	1
5	1,22364160	4	0	0
6	0,73997572	2	0	0
7	1,27090547	4	0,26064119	0
8	1,11732428	6	0	0
9	1,01327549	4	0	0
10	1,19438509	3	0	0
11	2,18187034	1	0	0
12	1,15113969	5	8,05732566	0
13	1,11453406	5	0	0
14	0,87089232	4	0	0
15	0,00000000	1	0	0
16	1,09926728	3	0	0
17	0,24391021	0	0	0
18	0,34886079	0	0	0
19	0,90529743	4	4,7868067	1
20	1,65283038	1	0	0

21	0,72284939	3	1,82123827	0
22	1,16767258	6	0	0
23	1,32950262	5	0	0
24	2,00336991	2	0	0
25	0,85343210	2	6,89247398	0
26	0,82777680	7	0	0
27	1,22992630	5	6,68613932	0
28	0,84698552	2	0	0
29	1,00931763	5	0	0
30	2,31936225	1	0	0
31	1,54096460	5	0	0
32	1,17816345	5	0	0
33	0,87601542	5	0	0
34	1,98436898	1	0	0
35	0,94124567	5	0	0
36	0,97782713	2	0	0
37	0,97167965	5	0	2
38	0,86521782	6	0	0
39	3,84679392	2	0	0
40	1,16088869	0	0	0
41	0,85337600	2	0	0
42	0,98489242	4	0	0
43	1,07884441	5	0	0
44	1,18635075	6	0	0
45	0,76488416	6	0	0
46	1,08991277	2	0	0
47	0,80469082	3	0	0
48	0,95553174	3	0	0
49	1,32084535	2	0	0
50	1,17959330	3	0,92356253	0
51	0,73290317	5	0	0
52	1,06559477	4	0,38912993	0
53	0,75409970	4	0	0
54	1,36485978	4	-0,35188869	0
55	1,03140116	5	0	3
56	1,34306815	4	0	0
57	1,05017641	4	0	1
58	0,00000000	1	0	0
59	1,10336695	0	0	0
60	0,78645703	4	0	0
61	1,39592599	7	9,18692327	0
62	0,94979833	4	0	0
63	0,92299507	5	0	1

64	1,23387800	4	0	0
65	1,29723124	3	0	0
66	0,81611668	3	0	0
67	0,81611668	3	0	0
68	1,16539368	4	0	0
69	1,63289598	4	0	0
70	1,06982810	5	0	0
71	1,17282882	1	0	0
72	1,13448836	4	0	0
73	1,27187295	6	0	0
74	1,07117269	6	0	0
75	1,76280990	5	0	0
76	0,91403304	6	0	0
77	1,51458605	4	0	0
78	1,46061643	4	0	1
79	1,20243754	3	0	0