

## 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]. Las lombrices de tierra (*Eisenia foetida*) han sido utilizadas como bioindicadores para la evaluación de la toxicidad aguda de los plaguicidas en los ecosistemas terrestres [2]. Lamentablemente, la determinación experimental de la toxicidad requiere un tiempo considerable, es costosa y genera un dilema ético (demandas para reducir o abolir las pruebas en animales). 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 [3]. Estos métodos comprenden las relaciones cuantitativas estructura-propiedad (QSAR). En este trabajo realizamos un estudio QSAR con el fin de predecir la toxicidad aguda de los plaguicidas en *Eisenia foetida*.

## 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 [4]. La actividad fue expresada como concentración letal media ( $LC_{50}$ ): corresponde a la concentración ( $mg \cdot kg^{-1}$ ) de pesticida que conduce a la muerte del 50% de la lombriz *Eisenia foetida*. Los valores de toxicidad aguda fueron convertidos en  $\log LC_{50}$ . Se extrajeron de la base de datos PPDB 79 pesticidas divididos 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.

Se dibujaron las estructuras químicas con ACD ChemSketch [5] con el formato molecular en MDL mol (V2000). La conversión de las moléculas se realizó con el programa para Windows Open Babel [6]. Los descriptores moleculares se calcularon mediante los programas gratuitos y de libre acceso: PaDEL [7], Recon [8], Mold2 [9], DataWarrior [10] y QuBiLS-MAS [11]. Se obtuvieron un total de 23971 descriptores independientes de la conformación. Esta cantidad se redujo a 2055 tras la exclusión de los descriptores vacíos, linealmente dependientes y de valores constantes o casi constantes. Se efectuó la partición del conjunto molecular (58 moléculas) en tres subconjuntos. Uno de estos, calibración ( $N_{cal} = 41$ ), se empleó para seleccionar los mejores descriptores del modelo de regresión lineal multivariable (MLR). Los otros dos subconjuntos, validación ( $N_{val} = 8$ ) y predicción ( $N_{pred} = 9$ ), se utilizaron para la validación del modelo MLR obtenido y determinar su capacidad predictiva. Para llevar a cabo la partición mencionada se utilizó la técnica "Método de Subconjuntos Balanceados" (BSM) [12].

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) [13], 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 ( $RMS_{cal}$ ). Aparte de utilizar un conjunto de validación externo, el mejor modelo obtenido fue validado por medio de validación cruzada 'dejar-uno-afuera' (*loo*) y aleatorización-y (*rand*) [14]. Todos los algoritmos de Octave [15] se desarrollaron en nuestro grupo y están disponibles.

## 4. CONCLUSIONES

A partir de datos verificados y confiables de la base de datos PPDB, se consiguió establecer un modelo QSAR predictivo de 4 descriptores no conformacionales para la toxicidad aguda de pesticidas evaluada en la lombriz *Eisenia foetida*. Esto constituye una guía QSAR valiosa para la búsqueda de estructuras con valores determinados de toxicidad. Además, la aplicación del modelo para la predicción de 21 plaguicidas con datos no verificados o menos confiables puede considerarse aceptable.

## AGRADECIMIENTOS

Se agradece al subsidio de CONICET (PIP0311) y al Ministerio de Ciencia, Tecnología e Innovación Productiva por las facilidades de la biblioteca electrónica.

## REFERENCIAS

- [1] Wan, M. *J. Environ. Sci. Heal. Part B*, 2013, 48, 344-363.
- [2] Zou, X., Xiao, X., Zhou, H., Chen, F., Zeng, J., Wang, W., Feng, G., Huang, X., *J. Hazard. Mater.*, 2018, 359, 365-372.
- [3] Sullivan, K., Manuppello, J., Willett, C., *SAR QSAR Environ. Res.*, 2014, 25, 357-365.
- [4] <http://sitem.herts.ac.uk/aeru/ppdb/en/index.htm> (último acceso 23/2/2021).
- [5] <https://www.aclilabs.com/> (último acceso 23/2/2021).
- [6] [http://openbabel.org/wiki/Main\\_Page](http://openbabel.org/wiki/Main_Page) (último acceso 23/2/2021).
- [7] <http://www.yapcwsoft.com> (último acceso 23/2/2021).
- [8] RECON, version 5.5/5.3. C. Breneman, W. Katt, M. Martinov, M. Rhen, N. Sukumar, T. Thompson, C. Whitehead and D. Zhuang. (Rensselaer Polytechnic Institute, 2003).
- [9] Hong, H., Xie, Q., Ge, W., Qian, F., Fang, H., Shi, L., Su, Z., Perkins, R., Tong, W., *J. Chem. Inf. Model.*, 2008, 48, 1337-1344.
- [10] T. Sander, Idorsia Pharmaceuticals Ltd. OSIRIS DataWarrior. Versión 4.7.3. (2014).
- [11] QuBiLS-MAS, Valdés-Martini, J., Marrero-Ponce, Y., García-Jacas, C., Martínez-Mayorga, K., Barigye, S., Silveira Vaz d'Almeida, Y., Pham-The, H., Pérez-Giménez, F., Morell, C., *J. Cheminform.*, 2017, 9, 35-61.
- [12] Rojas, C., Duchowicz, P., Trípaldi, P., Pis Diez, R., *J. Chromatogr. A*, 2015, 1422, 277-288.
- [13] Duchowicz, P., Castro, E., Fernández, F., *MATCH Commun. Math. Comput. Chem.*, 2006, 55, 179-192.
- [14] Gramática, P., *QSAR Comb. Sci.*, 2007, 26, 694-701.
- [15] GNU Octave, Scientific Programming Language <https://www.gnu.org/software/octave/index> (último acceso 23/2/2021)

## 3. RESULTADOS

La mejor relación cuantitativa estructura-actividad encontrada a través de RM involucra 4 descriptores:

$$\log LC_{50} = -0.56 \text{ GATS4m} + 0.15 \text{ nBondsD} - 0.17 \text{ SHBint4} - 0.88 \text{ APC8NX} + 2.36$$

$$N_{cal} = 41, R_{cal}^2 = 0.72, RMS_{cal} = 0.41,$$

$$R_{val}^2 = 0.03, \sigma_{2.5} = 0, R_{pred}^2 = 0.52, RMS_{pred} = 0.54 \text{ (100000 casos)}$$

$$R_{tot}^2 = 0.66, RMS_{tot} = 0.46, N_{val} = 8, R_{val}^2 = 0.70, RMS_{val} = 0.38,$$

$$N_{pred} = 9, R_{pred}^2 = 0.87, RMS_{pred} = 0.31.$$

$R^2$  es el coeficiente de determinación. La correlación máxima entre pares de descriptores no es significativa ( $R_{ij}^2 \max = 0.03$ ).

La capacidad predictiva del modelo QSAR es satisfactoria: cumple con los parámetros estadísticos de validación cruzada y externa y aleatorización-y.

En las Figs. 1 y 2 se grafican los resultados de las predicciones de  $\log LC_{50}$  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  $\sigma_{cal}$ .

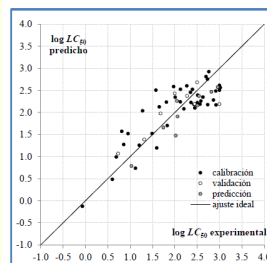


Fig. 1. Predicciones del modelo QSAR.

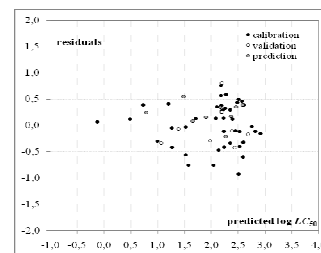


Fig. 2. Dispersión de residuos

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

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. En la mayoría de los casos las predicciones fueron próximas a los valores reportados no verificados. Solo 1 plaguicida estuvo fuera del dominio de aplicación del modelo QSAR (aclonifen), mientras que 2 moléculas poseían predicciones que se diferenciaban más del valor reportado no verificado (1,2-dicloropropano y tiobencarb).