

INTRODUCCIÓN

Durante décadas el uso repetitivo de los precursores químicos de síntesis en el diseño de nuevos productos veterinarios y pesticidas ha promovido el desarrollo acelerado de la resistencia en insectos y los efectos adversos en los organismos no objetivos. Ante la necesidad de realizar una apropiada identificación temprana del riesgo de miles de sustancias comerciales, el ensayo de toxicidad en embriones del pez cebra (*Danio rerio*) representa una estrategia sostenible para determinar la toxicidad del desarrollo de los agentes químicos, durante los procedimientos regulatorios establecidos a nivel mundial [1].

No obstante, diferentes autoridades gubernamentales han acelerado la incorporación de los principios de las 3Rs (reducir, reemplazar y refinar) con el objetivo de limitar la experimentación con animales al nivel mínimo absolutamente necesario, fomentando el desarrollo y validación de alternativas a la experimentación con animales, como los métodos de modelado asistido por computadora, donde se destaca la teoría de las Relaciones Cuantitativas Estructura – Actividad (QSAR) [2].

METODOLOGÍA

Se realizó el estudio QSAR de 188 compuestos comerciales con datos de toxicidad del desarrollo en embriones del pez cebra (*Danio rerio*) expresados como Concentración Activa media (AC_{50}) en la librería ToxCast™ fase I de la Agencia de Protección Ambiental (EPA) [3]. Las estructuras químicas fueron representadas bidimensionalmente con el programa gratuito ACDLabs Chemsketch [4] en formato MDL mol. (v2000) y SMILES.

Para codificar las características estructurales de los agentes químicos se calcularon 28.038 descriptores no conformacionales por medio de los programas de acceso libre PaDEL [5], Mold2 [6], QuBiLS-MAS y FRAGMENTOR [7, 8]. La relación descriptores-actividad se determinó con el Método del Reemplazo (RM) [9], un eficiente algoritmo basado en la técnica supervisada de Regresión Lineal Multivariable (MLR).

Para desarrollar el modelo QSAR se dividió el conjunto molecular total mediante el Método de Subconjuntos Balanceados (BSM) [10], basado en el análisis de agrupamientos k -medias. Así se obtuvieron tres subconjuntos representativos en su relación Estructura-Actividad, denominados Calibración (*Cal*), Validación (*Val*) y Predicción (*Pred*), el cual determina la capacidad predictiva del modelo.

La estabilidad y confiabilidad del modelo obtenido se evaluó a través de diferentes técnicas de validación interna, como la validación cruzada ‘dejar-uno-afuera’ (*loo*), ‘dejar-más-afuera’ (*lmo*) con 15% afuera, la prueba de Aleatorización-Y (*rand*), el Factor de Inflación de la Varianza (*VIF*) y el Dominio de Aplicación (AD) [11]. Todos los algoritmos utilizados en este trabajo fueron programados por nuestro grupo en GNU Octave 5.2.0. 2020 [12], y están a disposición del público.

RESULTADOS Y DISCUSIÓN

El método BSM generó tres subconjuntos equilibrados, *Cal* representa el 70% de la información de la base de datos ($N_{cal}=132$ moléculas), y tanto *Val* ($N_{val}=29$ moléculas) y *Pred* ($N_{pred}=27$ moléculas) consideran el 30% restante. Las mejores regresiones lineales fueron obtenidas a través del algoritmo RM basado en MLR, con capacidad de explorar 12.465 descriptores teóricos linealmente independientes. Se analizaron los modelos más representativos de $d=1-8$ descriptores moleculares. La selección del mejor modelo se basó en los parámetros estadísticos: coeficiente de determinación (R^2), Raíz Cuadrada del Error Cuadrático Medio (*RMSE*) y coeficiente de correlación máximo entre pares de descriptores ($R_{ij\max}^2$). Se logró establecer un modelo lineal basado en 7 descriptores no conformacionales, expresado en la Ecuación 1, con apropiada calidad estadística en los conjuntos *Cal* y *Val*, como se muestra a continuación:

$$\log_{10} CA_{50} = 2.2 + 0.7d_1 - 0.1d_2 - 5.7 \times 10^{-4}d_3 + 1.1d_4 - 0.01d_5 + 0.03d_6 - 1.6d_7 \quad \text{Ec.1}$$

$$N_{cal}=132, R_{cal}^2=0.65, RMSE_{cal}=0.44, R_{ij\max}^2=0.92, R_{loo}^2=0.61, RMS_{loo}=0.52$$

$$R_{lmo}^2=0.58, RMS_{lmo}=0.56, RMS^{rand}=0.27$$

$$N_{val}=29, R_{val}^2=0.50, RMSE_{val}=0.54, VIF_{max}=7.70$$

$$N_{pred}=26(96\%), R_{pred}^2=0.76, RMS_{pred}=0.41$$

Con los resultados obtenidos se determinó que el modelo presenta adecuada estabilidad según los parámetros $R_{loo}=0.61$, $RMSE_{loo}=0.52$ y $R_{lmo}=0.58$, $RMSE_{lmo}=0.56$. La prueba de Aleatorización-Y reveló que el modelo no es producto de una correlación fortuita con un valor de $RMSE_{cal} < RMSE_{rand}$ (0.27). La Ec. 1 cumple con los criterios de validación de Golbraikh et al. 2003. Por otra parte, se evidenció que la Ec. 1 predice apropiadamente la toxicidad del desarrollo (CA_{50}) del conjunto *Pred*, como se exhibe en la Figura 1; adicionalmente cumplió con las distintas pruebas de validación internas y externas expresadas en la Tabla 1.

Tabla 1. Valores de los parámetros de validación interna y externa del modelo obtenido.

d	R_{pred}^2	RMS_{pred}	R_{loo}^2	R_{lmo}^2	R_m^2	Q_{F1}^2	Q_{F2}^2	Q_{F3}^2	CCC
7	0,76	0,41	0,60	0,58	0,53	0,67	0,67	0,76	0,85

CONCLUSIONES

En el presente trabajo se propone un modelo lineal QSAR para predecir la toxicidad del desarrollo en embriones del pez cebra (*Danio rerio*) de 188 agentes de uso comercial como antiparasitarios, antibacteriales y antimicrobianos, los cuales fueron codificados por medio de 28.038 descriptores no conformacionales. A través de la estrategia de selección de variables RM-MLR se logró definir el mejor modelo QSAR con $d=7$ descriptores, apropiada calidad estadística y notable poder predictivo de la AC_{50} . De este modo, se estableció una herramienta computacional alternativa para reducir el número de pruebas de toxicidad *in vivo* a las mínimas necesarias.

AGRADECIMIENTOS: L.M.S. agradece a CONICET por la Beca Postdoctoral. P.R.D. y L.M.S. agradecen al proyecto de CONICET PIP0311.

REFERENCIAS: 1. Chakraborty, C., et al. *Current Drug Metabolism*, **2009**, 10,116–124. 2. Hansch, C.; et al. *Exploring QSAR, Fundamentals and Applications in Chemistry and Biology*, American Chemical Society., Ed.; Washington, D. C., **1995**. 3. Padilla, S. et al. *Reprod. Toxicol.* **2012**, 33, 174–187. 4. ACD/ChemSketch. www.acdlabs.com, 2016. 5. PaDEL-Descriptor V 2.20, <http://www.yapcwsoft.com> 6. Hong, H., et al. *J. Chem. Info. Model*, **2008**, 48, 1337–1344. 7. Valdes-Martini, J.R., et al. *J. Chemoinformatics*, **2017**, 9, 35–61. 8. Varnek, A., et al. *Curr. Computer-Aided Drug Des.* **2008**, 4, 191–198. 9. Duchowicz, P.R.; et al. *MATCH Commun. Math. Comput. Chem.* **2006**, 55,179. 10. Rojas, C.; et al. *Chemometrics Intell. Lab. Syst.* **2015**, 140, 126. 11. Gramatica, P., et al. *J. Chem. Info. Model*, 2016, 56, 1127–1131. 12. GNU Octave 5.2.0, 2020, <https://www.gnu.org/software/octave>.

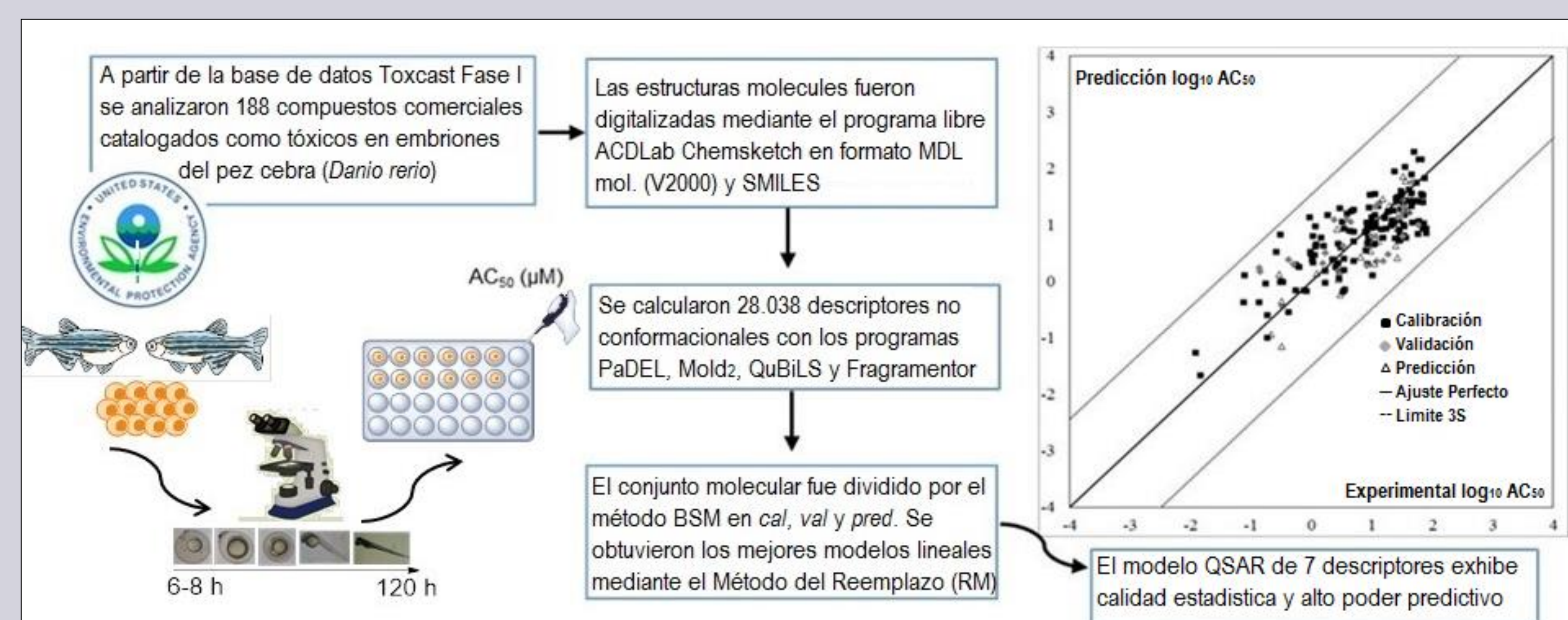


Figura 1. Procedimiento de diseño del Modelo QSAR con 7 descriptores no conformacionales, y la correlación del valor de AC_{50} experimental y calculado para cada compuesto en estudio.

Los 7 descriptores independientes de la conformación (0D-2D) fueron generados por cuatro programas diferentes: d_1 y d_2 son descriptores indicadores de PaDEL, d_3 y d_4 representan descriptores topológicos Mold2, d_5 es un índice algebraico QuBiLS-MAS y d_6 y d_7 codifican descriptores de conteo de fragmentos generados por el programa FRAGMENTOR. Los resultados obtenidos para la Ec. 1 son comparables a un modelo QSAR publicado previamente en la literatura por Lavado et al., 2020, para estimar la toxicidad del desarrollo de los embriones del pez cebra (*Danio rerio*).

Finalmente, el presente enfoque QSAR podría contribuir a la falta general de modelos disponibles para estimar la toxicidad del desarrollo de extensos conjuntos moleculares a partir de un modelo animal alternativo y sostenible.