

## Modelización de la estructura-toxicidad de diversos pesticidas utilizando métodos QSAR

---

Luis F. Cárdenas C. y Piero Tripaldi

### Introducción

La calidad de los recursos hídricos es quizás actualmente el tema más discutido en lo que respecta a la preservación del ambiente, ya que los ecosistemas acuáticos han sufrido cambios en todo el mundo en la mayoría de los casos irreversibles. Estos cambios se asocian a menudo con las actividades humanas como la deforestación, la liberación de efluentes industriales y domésticos, además el uso de pesticidas en los campos agrícolas, lo cual contribuye a la caída de la calidad de los recursos hídricos (Zvinavash, 2009).

Los pesticidas son importantes para el sistema agrícola. Sin embargo, es crucial que se usen de manera adecuada a fin de preservar la calidad del producto final y los recursos naturales que apoyan su producción como lo son el suelo y el agua. Los pesticidas son productos cuya función es eliminar los daños que causan los organismos-plaga a los cultivos agrícolas garantizando así una alta productividad (Repetto, 2004).

A pesar de que estas moléculas, cuando se aplican, tienen los organismos objetivo, como su destino final, el 99% de los pesticidas que se aplican van al aire, agua y suelo, es decir, solo el 1% llega a su destino. Este hallazgo es inquietante debido a que la población mundial crece lo que significa que el uso de pesticidas aumentará para satisfacer las necesidades mundiales de alimentos por lo cual los recursos naturales se mantendrán bajo una intensa amenaza de estas moléculas (Toropov, 2010).

Su clasificación está dada en relación a diferentes parámetros como: origen del pesticida, plaga a controlar, persistencia, toxicidad y la estructura química. En general se prefiere clasificar a los pesticidas desde el punto de vista químico, ya que esto permite agruparlos con un criterio uniforme y establecer diversas correlaciones como: estructura-actividad, estructura-toxicidad, estructura-mecanismos de degradación. Dentro de esta categoría los principales pesticidas y los que serán parte del estudio son: los organofosforados, organoclorados, carbamatos y piretroides. Para determinar la toxicidad de los pesticidas se han estudiado y realizado diversos ensayos en animales, la ventaja fundamental de estos ensayos es su realización en condiciones rigurosas y controladas de laboratorio, de esta forma se puede establecer de manera cuantitativa la relación entre la exposición (dosis) y su efecto (respuesta) (Moreno, 2008).

Dentro de las desventajas de realizar estos ensayos es la detección de los efectos de dosis bajas en los animales, esto exige el uso de un número elevado de ejemplares lo que convierte al proceso en económicamente inviable; debido a las condiciones que requieren

los animales en el laboratorio. Además, el tiempo que necesita cada ensayo hace que el análisis de las muestras se realice en un lapso de tiempo demasiado largo (Toropov, 2010).

El costo de las pruebas en animales es prohibitivo es decir que este costo se asigna al precio final de los productos químicos lo cual los encarece. En un estudio de la Comisión Europea se estima que las pruebas toxicológicas de alrededor de 30.000 sustancias existentes se traducen en un costo total de 2 billones de euros en el mercado hasta el año 2013. En estas pruebas toxicológicas se estima que es necesario utilizar más de 12 millones de animales vertebrados. Dentro de las limitaciones de estos ensayos tenemos restricciones económicas, consideraciones éticas y la presión pública para reducir las pruebas en animales (Mazzatorta, 2010).

Debido a las limitaciones de los ensayos de toxicidad se han desarrollado nuevos métodos que nos ayudan a reducir el número de animales de experimentación, ahorrar tiempo y reducir el costo de las pruebas. Dentro de estos métodos tenemos los modelos matemáticos de predicción dentro de los cuales se encuentran los siguientes:

- Cinética ambiental de compuestos químicos
- Fármaco-toxicocinética
- Relación Cuantitativa Estructura Actividad (QSAR)

En la presente investigación utilizamos la técnica QSAR la cual tiene como objetivo principal el construir modelos matemáticos que relacionen la estructura química y la actividad biológica de una serie de compuestos análogos; debido a que los complejos fenómenos bioquímicos que afectan a la toxicidad puede ser descrito en términos matemáticos y químicos (Repetto, 2004).

QSAR permite encontrar modelos teóricos que relacionan la estructura de los productos químicos y sus respectivas actividades biológicas. QSAR se utiliza para predecir las propiedades fisicoquímicas o propiedades biológicas (toxicológicas) de nuevas estructuras químicas. Específicamente QSAR es una relación cualitativa entre la estructura molecular y la presencia o ausencia de una determinada actividad biológica, o la capacidad para modelar una actividad biológica en base a la estructura molecular de un compuesto (Todeschini, 2009). Las aplicaciones de QSAR en la actualidad son varias, ya que facilita encontrar modelos matemáticos basados desde el punto de vista de la estadística multivariante, por lo cual QSAR se ha convertido en una herramienta básica de:

- Diseño de modernos fármacos
- Predicción de la toxicidad química
- Predicción de la actividad de compuestos en el ambiente
- Predicción de propiedades moleculares

Esta investigación tiene como objetivo principal encontrar un modelo matemático que nos permita determinar la toxicidad de nuevas estructuras moleculares de pesticidas mediante QSAR y apoyadas de técnicas estadísticas multivariantes. Las técnicas de análisis multivariante es el conjunto de métodos estadísticos cuya finalidad es analizar simultáneamente conjuntos de datos en donde existe una serie de variables medidas para cada objeto

estudiado. Por esta razón, este tipo de análisis da un mejor entendimiento de la problemática del fenómeno estudiado obteniendo toda la información posible de los datos estudiados, ya que con técnicas univariantes o bivariantes no se lograría este propósito (Todeschini, 2009).

Dentro de estas técnicas de análisis multivariante se tienen los métodos de clasificación estos permiten categorizar a los objetos por medio de clases. En la presente investigación se han establecido tres categorías: pesticidas altamente tóxicos clase 3, medianamente tóxicos clases 2 y finalmente los de baja toxicidad clase 1. Para encontrar el modelo matemático se utiliza el método KNN este es un método de clasificación no paramétrico, que para realizar la clasificación se basa sobre el concepto de analogía. El método identifica mediante una distancia oportuna los K objetos más cercanos a la molécula en estudio. Luego cuenta cuántos de estos K objetos caen en cada clase; la molécula en estudio es atribuida a la clase más numerosa; esta evaluación se extiende a todos los objetos hasta determinar a qué clases corresponde cada uno de ellos (Todeschini, 2003). El algoritmo KNN tiene las siguientes etapas:

1. Escalado de los datos
2. Selección de la distancia a utilizar
3. Elección del número de K útil para la clasificación
4. Cálculo de la matriz de las distancias
  - a. Para cada objeto se consideran K objetos más cercanos
  - b. El objeto se asigna a la clase más representada en los K vecinos

Debido a la gran cantidad de datos que se tiene en la investigación (alrededor de 200.000 datos) se necesita utilizar técnicas que permitan disminuir el tamaño de la base de datos. Dentro de estas técnicas se ha utilizado el método de selección de variables con los algoritmos genéticos (AG).

Los AG son una potente estrategia de investigación quimiométrica, este en un campo en rápido desarrollo y que tiene un gran potencial en aplicaciones de análisis de datos. Este método fue propuesto por Holland en 1975 y en los últimos tiempos han sido muy utilizados para la solución de problemas de optimización en muchos campos (Todeschini, 2009).

Los AG son métodos adaptativos que se basan en el proceso genético evolutivo de los organismos vivos. Un AG consiste en un ciclo interactivo de creaciones de conjuntos (cromosomas) de variables originales que generan poblaciones las cuales son las más óptimas para la construcción del modelo predictor. Este modelo predictor permitirá determinar la toxicidad de nuevas estructuras moleculares de pesticidas y sus consecuencias en los organismos y el ambiente (Todeschini, 2009).

## Metodología

Para lograr el objetivo principal de la investigación contamos con una base de datos de pesticidas de la Universidad de Hertfordshire de Inglaterra, esta base de datos tiene alrededor de 1250 sustancias químicas que son utilizadas en el agro. La base de datos está disponible en la red para desarrollar diversas investigaciones y es actualizada constantemente; el nombre de la base de datos es PPDB (Pesticide Properties DataBase) la dirección de

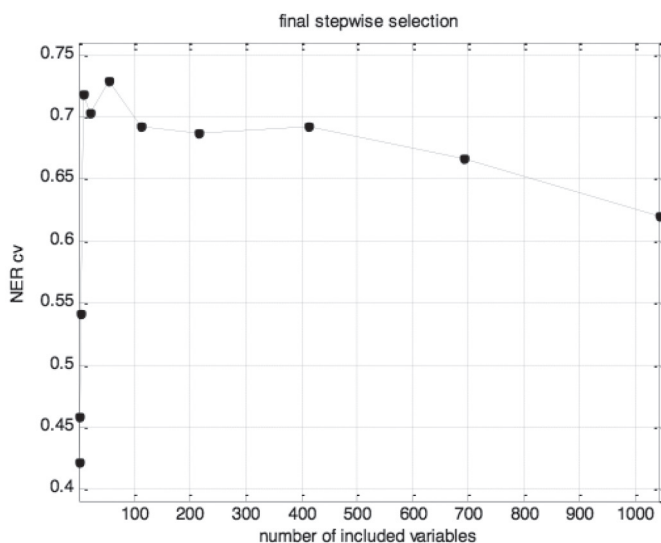
búsqueda es: <http://sitem.herts.ac.uk/aeru/footprint/es/index.htm>. Esta base de datos cuenta con siete secciones las cuales son las siguientes: descripción general del producto, propiedades físico-químicas, destino ambiental, degradación, absorción y movilidad en el suelo, ecotoxicología y salud humana. De esta base de datos se han seleccionado 192 pesticidas por su estructura química los cuales son: piretroides, organofosforados, organoclorados y carbamatos. De las secciones mencionadas se obtuvo el gráfico molecular del pesticida y el  $EC_{50}$  (concentración efectiva 50) de *Daphnia magna*. Mediante el software Hyperchem, se han digitalizado y optimizado las estructuras de las moléculas. Luego con el programa Dragón se han calculado 4800 descriptores moleculares para cada molécula. Luego el software inmediatamente reduce la dimensionalidad de la base de datos eliminando descriptores constantes o casi constantes ( $desv.est. < 0.00001$ ), los que presentan "missing values" y los que están correlacionados por más del 95%. De manera que la matriz de la base de datos resulta reducida a las siguientes dimensiones: 192 moléculas por 1.126 descriptores (216.192 datos).

Para normalizar la  $EC_{50}$  de *Daphnia magna* se calculó  $\ln(1/EC_{50})$ , luego los valores se dividieron en tres clases: la clase 1 que tienen una toxicidad baja (0-33% de los datos), la clase 2 posee una toxicidad intermedia (33%-66% de los datos) y finalmente la clase 3 con toxicidad alta (66%-100% de los datos). En el anexo 1 se muestra una tabla de datos en donde se encuentran los pesticidas y sus valores de clases y  $EC_{50}$  para *Daphnia magna*.

## Resultados

Una vez aplicado el algoritmo genético, se ha obtenido la relación entre los descriptores y la calidad del modelo en predicción la cual se muestra ahora:

Figura 1  
Relación entre número de descriptores considerados en la clasificación KNN  
y la calidad del modelo en predicción



Se puede notar que la calidad de predicción del modelo, expresada como No Error Rate cross, validada al 20% de las moléculas (5 grupos de cancelación), aumenta hasta alcanzar el máximo de 0.72 es decir una capacidad predictiva del 72%. En la tabla 1 se presenta la matriz de confusión de la clasificación.

Tabla 1  
Matriz de confusión

		Clase Asignada		
		Clase 1	Clase 2	Clase 3
Clase Verdadera	Clase 1	44	13	7
	Clase 2	9	42	12
	Clase 3	3	13	49

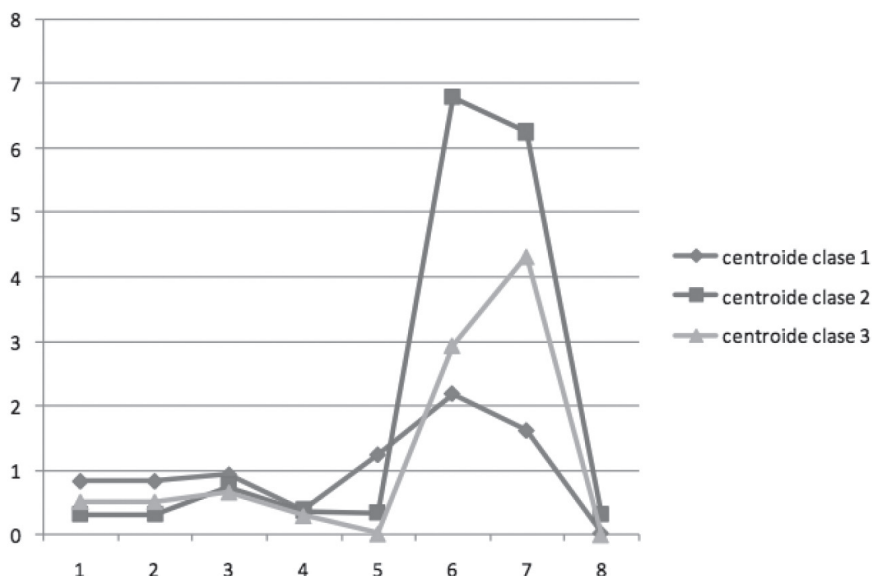
- En la clase 1 (baja actividad) el método presenta un No Error Rate del 69%
- En la clase 2 (media actividad) el método presenta un Error Rate del 66%
- En la clase 3 (actividad alta) el método presenta un Error Rate del 75%

El algoritmo es más sensible a las moléculas de pesticidas más activos que son los de la clase 3, en tanto que resulta menos sensible a las moléculas de clase 2 y 1. Los descriptores seleccionados (según el código de DRAGON) han sido:

1. B05[C-P] que indica la presencia o ausencia de un enlace C-P a una distancia topológica de 5
2. B02[C-N] que indica la presencia o ausencia de enlace C-N a una distancia topológica de 2
3. MATS5i que indica que existe una autocorrelación calculada según Moran pesando los potenciales de ionización de los átomos a una distancia de 5 enlaces
4. E2p es un descriptor de direccionalidad de la molécula perteneciente al grupo WHIM
5. CATS2D\_02\_DA es un descriptor básico que indica la existencia de un par donador-aceptor a una distancia de 2 enlaces
6. CATS2D\_04\_LL es un descriptor básico que indica la existencia de un par lipofílico-lipofílico cada 4 enlaces
7. CATS2D\_05\_LL es un descriptor básico que indica la existencia de un par lipofílico-lipofílico cada 5 enlaces
8. F09(O-F) frecuencia de un enlace O-F a una distancia topológica de 9

Los descriptores moleculares que mejor separan las clases son el descriptor 6 y 7 esto lo podemos observar en la figura 2.

Figura 2  
Centroides de las clases



Se puede ver que las variables 6 y 7 son la que más diferencian las tres clases de actividad. En la clase 3 de más actividad las variables 6 (CATS2D\_04\_LL) y 7 (CATS2D\_05\_LL) son las que tienen valores intermedios indicando así la presencia media de lipofílicas. En la clase 2 de mediana actividad las mismas variables tienen valores altos, mientras que en la clase 1 presentan valores bajos. Esto indica que para presentar una elevada actividad la molécula debe presentar una lipofilicidad intermedia.

## Discusión

El modelo obtenido presenta una buena capacidad predictiva ( $NER = 0.72$ ) a pesar de tener una gran cantidad de moléculas (192). Según Ferreira (2006), encuentra un modelo con un menor  $NER = 0.71$  y con un número mucho menor de moléculas (25). Lo cual nos permite predecir la actividad biológica de nuevas estructuras moleculares de pesticidas.

KNN al ser un método de clasificación nos permite agrupar los diferentes tipos de pesticidas ya que los modelos continuos solo describen de manera parcial la relación entre la estructura molecular de los pesticidas y su respectiva toxicidad (Benfenati, 2005). Por lo cual KNN nos permiten analizar de manera global la toxicidad y la estructura molecular de diversos tipos de pesticidas clasificados por su toxicidad y estructura química.

Según Benfenati (2005) los modelos continuos son más utilizados y estudiados debido a que su predicción devuelve un valor de toxicidad, en tanto que los modelos discretos como nuestra investigación devuelven el valor de su clase es decir muestra el valor de la escala de la clase asignada a cada pesticida analizado.

## Conclusiones

El algoritmo genético reduce la dimensionalidad de la base de datos ya que al iniciar el análisis se tenían 1.216 descriptores moleculares de los cuales solo 8 describen la actividad de las moléculas de pesticidas.

El  $NER = 0.72$  encontrado con el método KNN presenta una buena capacidad predictiva a pesar de tener una gran cantidad de moléculas (192) y sobre todo tener pesticidas de diversas características químicas.

Los modelos continuos son más utilizados y estudiados a pesar de que solo describen ciertas características moleculares y se basan en moléculas similares. En tanto que los modelos discretos describen de mejor manera la actividad y su respectiva toxicidad ya que podemos tener varias clases de pesticidas organizados por su estructura molecular.

## Referencias

- Benfenati  
2005 *Quantitative Structure-Activity Relationships (QSAR) for Pesticide Regulatory Purposes*. Amsterdam, The Netherlands
- Devillers, J.  
2010 "A General QSAR Model for Predicting the Acute Toxicity of Pesticides to *Lepomis Macrochirus*". *SAR y QSAR in Environmental Research*, 11:5-6, 397-417. France.
- Ferreira  
2005. "Multivariate QSAR". *J. Braz. Chem. Soc.* Vol. 13. Nº 6, 742-753.
- Mazzatorta, P.  
2010. "QSAR Model for Predicting Pesticide Aquatic Toxicity. Istituto di Ricerche Farmacologiche 'Mario Negri'". *Chem. Inf. Model.* 45, 1767-1774. Milán.
- Moreno, D.  
2008. *Toxicología ambiental*. Madrid: McGraw-Hill.
- Todeschini  
2003 *Introducción a la Quimiométrica*. Edises. Milán: Universidad Milano Bicocca.  
2009 *Descriptors from Molecular Geometry. DISAT*. Milán: Universidad Milano Bicocca.
- Toropov, A.  
2010 "QSAR models for *Daphnia* toxicity of pesticides based on combinations of topological parameters of molecular structures". *Bioorganic y Medicinal Chemistry*. 14: 2779-2788. Milán.
- Repetto  
2004 "Métodos Alternativos a la experimentación animal en la Evaluación de la Seguridad. Instituto Nacional de Toxicología y Ciencias Forenses". Jornadas de Toxicología Ambiental-Torre Guil. Murcia.
- Zvinavashe, *et al.*  
2009 "Quantitative structure-activity relationship modeling of the toxicity of organothiophosphate pesticides to *Daphnia magna* y *Cyprinus carpio*". *Chemosphere*, 75: 1531-1538.