Facultad de Ciencia y Tecnología

Escuela de Ingeniería en Alimentos

Aplicación del aprendizaje automático para la predicción de los tiempos de retención de pesticidas medidos en cromatografía líquida de fluidos supercríticos en muestras de espinacas

Trabajo de graduación previo a la obtención del título de:
INGENIERO EN ALIMENTOS

Autor:
Alan Mateo Mendoza Zambrano

Director:
Dr. Cristian Rojas Villa

Cuenca – Ecuador
2022

TABLA DE CONTENIDOS

INDICE DE FIGURAS

INDICE DE TABLAS

DEDICATORIA

¿Como no dedicar este trabajo a mis padres? Quienes nunca descansaron por darme a mi y a mis hermanos una vida plena y llena de alegrías. Sus enseñanzas me guiaran siempre por el camino correcto. En honor a todos sus sacrificios dedicare este trabajo y mi vida a mis padres, Rodolfo Mendoza y Shirley Zambrano, así como ellos dedicaron su vida por sus hijos. Este triunfo es de ustedes, los amo.

AGRADECIMIENTOS

A mi novia Jessica Guerrero quien me acompaño en cada paso de mi formación académica y me brindo su apoyo sin esperar nada a cambio. A mis compañeros con quienes crecimos juntos en esta etapa, espero que nuestra amistad perdure en el tiempo. A mis profesores por sus enseñanzas y cocimientos impartidos que hoy forjan el profesional que aspiro a ser.

Aplicación del aprendizaje automático para la predicción de los tiempos de retención de pesticidas medidos en cromatografía líquida de fluidos supercríticos en muestras de espinacas

RESUMEN

En este estudio se ha desarrollado un modelo basado en las relaciones cuantitativas estructura-propiedad (QSPR) para una base de datos de 375 pesticidas identificados en muestras de espinacas. Las moléculas en la base de datos fueron curadas bajo diversos criterios. La propiedad experimental es el tiempo de retención (t_R) medido mediante cromatografía líquida de fluidos supercríticos (SFC). Los compuestos fueron representados por descriptores moleculares independientes de la conformación, los cuales fueron analizados mediante el método no supervisado W-VSP. Los descriptores restantes se usaron para calibrar diversos modelos de mínimos cuadrados ordinarios (OLS) acoplados con el método de remplazo (RM) para la selección supervisada de variables. Se obtuvo un modelo QSPR predictivo con 6 descriptores, el cual fue validado mediante técnicas internas y externas. Adicionalmente, se estableció el dominio de aplicabilidad del mismo y se proporciono los mecanismos de acción para cada descriptor molecular.

Palabras clave: *pesticidas, espinaca, QSPR, descriptores moleculares*

Ing. María Fernanda Rosales, MSc
Directora de Escuela

Dr. Cristian Rojas Villa
Director del Trabajo de Titulación

Alan Mateo Mendoza Zambrano
Autor

ABSTRACT

A quantitative structure-property relationship (QSPR) model was developed from a database of 375 pesticide residues identified in spinach samples. Molecules within the database were curated following diverse criteria. The key experimental property was the retention time (t_R) obtained by means of supercritical fluid chromatography (SFC). Molecules were represented by means of conformation- independent molecular descriptors, which were analyzed using W-VSP unsupervised machine learning methods to reduce the number of variables. The remaining descriptors were used to calibrate diverse multiple regression models based on ordinary least squares (OLS) coupled with the replacement method (RM) supervised variable subset selection. A predictive QSPR model with six descriptors was selected as the optimal one, which was appropriately validated using internal and external procedures. In addition, the applicability domain of the model was established and a mechanistic interpretation of each descriptor in predicting the t_R is provided.

Keywords: *spinach, pesticide residues, SFC, QSPR*

Ing. María Fernanda Rosales, MSc
Faculty Director

Dr. Cristian Rojas Villa
Thesis Supervisor

Translated by:

Alan Mateo Mendoza Zambrano
Author

CAPITULO I
MARCO TEÓRICO

1.1 Espinaca

La espinaca es una planta frondosa de hojas verdes originaria de Asia central y occidental. Sus hojas son una verdura comestible que se consume fresca o después del almacenamiento mediante técnicas de conservación tales como el enlatado, congelación o deshidratación. En el 2018, la producción mundial de espinaca fue de 26,3 millones de toneladas y solo China representó el 90% del total mundial producido (FAO, 2020). Al igual que otros cultivos, la espinaca es atacada por plagas y enfermedades durante la producción y el almacenamiento, los cuales provocan daños que reducen su calidad y rendimiento. Para minimizar la pérdida y mantener la calidad de la cosecha se utilizan pesticidas junto con otras técnicas de manejo de plagas para prevenir enfermedades. Los pesticidas más comúnmente encontrados en las espinacas son la *permetrina*, *imidacloprid* y *spinosad* A y D (Punzi et al., 2005). En la Figura 1 se muestra la estructura química de estos compuetos. El uso de pesticidas ha aumentado debido a que tienen una acción rápida, disminuyen las toxinas producidas por los organismos que infectan los alimentos y requieren menos mano de obra que otros métodos de control de plagas. Sin embargo, la presencia de residuos de pesticidas es una preocupación para los consumidores por sus potenciales efectos tóxicos, tales como interferir con los sistemas reproductivos y el desarrollo fetal, cáncer y asma, entre otros (Gilden et al., 2010). Es por esta razón que es importante estudiar las propiedades fisicoquímicas de los pesticidas o sus actividades biológicas que se pueden realizar mediante la teoría de las Relaciones Cuantitativas Estructura-Propiedad (QSPR, *Quantitative Structure/Property Relationships*).

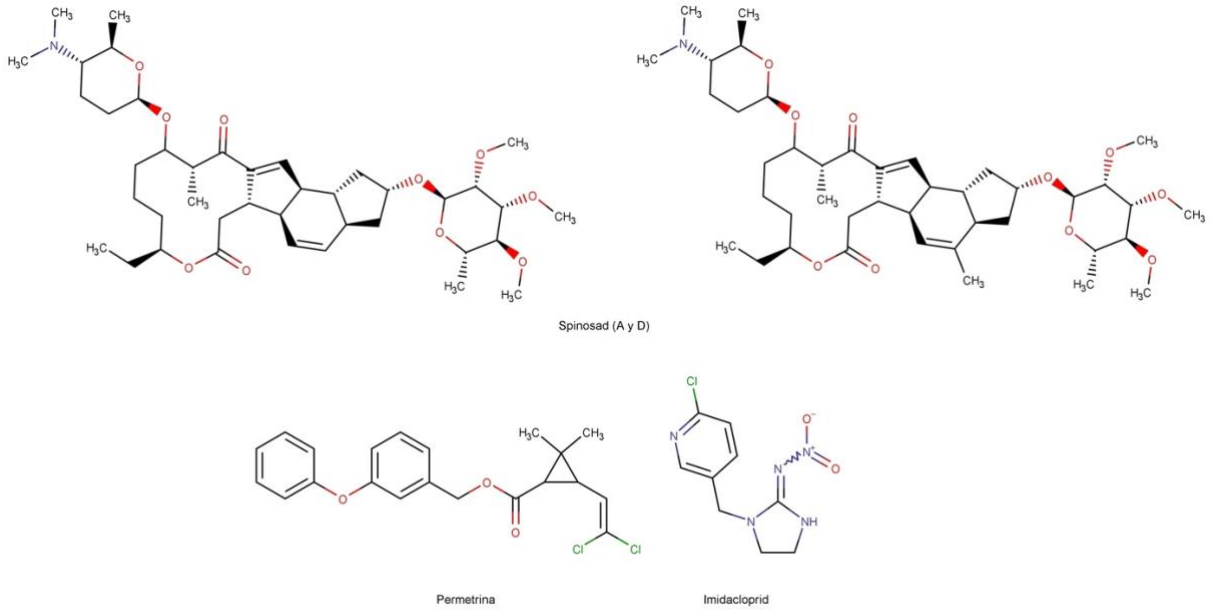

Figura 1. Estructura química de los pesticidas mayormente encontrados en muestras de espinacas. Fuente: Autor

1.2 Pesticidas

Un pesticida es cualquier sustancia que se usa para matar, repeler o controlar ciertas formas de vida vegetal o animal que se consideran plagas o pestes. Los pesticidas incluyen herbicidas para destruir malas hierbas y otra vegetación no deseada, insecticidas para controlar una amplia variedad de insectos, fungicidas utilizados para prevenir el crecimiento de moho y hongos, desinfectantes para prevenir la propagación de bacterias y compuestos utilizados para controlar ratones y ratas (Thayer et al., 2012). El uso de pesticidas ha brindado gran ayuda a la producción agrícola, aumentando la protección y el rendimiento de los cultivos. Sin embargo, el descubrimiento de residuos de pesticidas en varias secciones del medio ambiente ha generado serias preocupaciones con respecto a su uso; estas preocupaciones parecen tener más peso que los beneficios que brindan (Ali et al., 2014). Según un informe conjunto producido por la OMS (Organización Mundial de la Salud) y el PNUMA (Programa de las Naciones Unidas para el Medio Ambiente), aproximadamente 200000 personas mueren y alrededor de tres millones se envenenan cada año con pesticidas en todo el mundo, aunque la gran mayoría (95 %) de los casos proviene de países en desarrollo (WHO/UNEP, 1990).

Los residuos de pesticidas persistentes suelen permanecer más tiempo en los cultivos y llegar a los humanos a través de la cadena alimentaria. Estos residuos no deben exceder los límites permitidos ya que esto puede causar una amenaza para la salud humana. Por lo tanto, se han propuesto y desarrollado los siguientes parámetros (Bhushan et al., 2013): límites máximos de residuos (LMR),

ingesta diaria admisible (IDA) e ingesta diaria máxima teórica (IDMT) para monitorear los residuos de estos plaguicidas en la cadena alimentaria. El LMR es el límite máximo de residuos de pesticidas que pueden considerarse normales en un producto tratado con ellos, siempre que se hayan seguido las buenas prácticas agrícolas. Una IDA es la ingesta máxima aceptable de pesticidas de todas las fuentes dietéticas en un día, sin que representen ningún riesgo crónico para la salud. El IDTM es una estimación de la ingesta máxima de pesticidas con los LMR existentes para una persona, como resultado de una práctica dietética particular. Para minimizar el riesgo de los pesticidas a un costo razonable, la mayoría de las naciones han impuesto límites reglamentarios, ya sea en forma de estándares (que se deben hacer cumplir) o pautas (que son niveles deseables) dentro de una normativa técnica. Por ejemplo, en el Ecuador se rigen a la norma NTE INEN 1834 para espinacas frescas (INEN, 2013).

1.3 Cromatografía de fluidos supercríticos y tiempos de retención

Las separaciones cromatográficas con un fluido supercrítico como fase móvil surgieron hace más de 50 años. La disponibilidad comercial de este tipo de fluidos y de equipos más modernos, hace que esta técnica sea una de las más utilizadas en la actualidad. Muchas separaciones por este método se llevan a cabo con dióxido de carbono supercrítico ($SCCO_2$*, Supercritical Carbon Dioxide*) como fase móvil y columnas de cromatografía líquida empaquetadas como fase estacionaria. Aunque el dióxido de carbono tiene muchas ventajas prácticas, incluida su temperatura crítica cercana a la ambiental y una interferencia mínima con la detección espectrofotométrica, el uso de otros fluidos supercríticos o la adición de modificadores al dióxido de carbono puede ampliar las aplicaciones de esta técnica (Gere, 1983). Los fluidos supercríticos (SF, *Supercritical Fluid*) tienen densidades y capacidades de disolución similares a la de ciertos líquidos, pero viscosidades más bajas y mejores propiedades de difusión. Por consiguiente, el SF utilizado como fase móvil en cromatografía debería actuar tanto como transporte de sustancias, tal como lo hacen las fases móviles en la cromatografía de gases, así como disolver estas sustancias de la misma forma que lo hacen los disolventes en la cromatografía líquida (Taylor, 2009).

El método de cromatografía de fluidos supercríticos (SFC, *Supercritical Fluid Chromatography*) acoplado con espectrómetro de masas resulta ser un enfoque prometedor para el análisis de multiresiduos de pesticidas en muestras de espinacas. En el análisis de pesticidas, el sistema SFC debe contar con las siguientes condiciones experimentales: una bomba binaria, un componente de columna de temperatura controlada, un muestreador automático y un sistema de refrigeración. El espectrómetro de masas debe estar equipado con ionización por electrospray. Adicionalmente, las condiciones de SFC incluyen a la fase móvil ($SCCO_2$ (disolvente A) y metanol con formato de amonio al 0,1% (disolvente B)); 5 µL de volumen de inyección; caudal de 3 mL/min; temperatura de la columna de 35 °C; y el uso de la columna Inertsil ODS-EP (250 × 4,6 mm, tamaño de partícula de 5

µm). A esta columna de fase inversa se le agrega un grupo no polar a la pared de gel de sílice (Ishibashi et al., 2015).

En esta técnica la respuesta experimental que se genera es el tiempo de retención cromatográfico (t_R). El t_R es el parámetro primario obtenido en un sistema de cromatografía para la identificación de los picos que generan los analitos, y mide el tiempo requerido desde la inyección de la muestra en la fase estacionaria hasta la elución completa del compuesto. Este parámetro considera el valor máximo del pico que pertenece a un pesticida en particular. El t_R para un compuesto dado no es fijo, ya que diversos factores afectan su determinación; por ejemplo, el caudal de la fase móvil, las diferencias de temperatura en el horno y la columna, así como la longitud y la degradación de la columna (Vu-Duc et al., 2019).

1.4 Relaciones cuantitativas estructura-propiedad

En la actualidad tiene gran interés el estudio de las relaciones entre la estructura molecular de un compuesto químico y sus propiedades fisicoquímicas y/o biológicas, debido a que las medidas experimentales de tales propiedades aun son desconocidas para algunos compuestos por ser nuevos, tóxicos, poco accesibles o demandan demasiado tiempo para la medición experimental (Gangwal et al., 2016). La metodología de las relaciones cuantitativas estructura-propiedad (QSPR) tiene como fin predecir las propiedades fisicoquímicas de las moléculas, las cuales sirven para realizar inferencias sobre la información del fenómeno y los mecanismos involucrados. La información que proveen los modelos QSPR sirven como complemento a otros estudios teóricos o experimentales que buscan elucidar de forma racional las interrogantes de tipo químico involucrada en dichos problemas (Kubinyi, 2008).

El modelado QSPR/QSAR se basa en establecer una correlación matemática entre una respuesta experimental (actividad/propiedad) y descriptores moleculares (atributos químicos que representan una estructura molecular) de las moléculas analizadas (Roy et al., 2015a). Dicha correlación puede derivarse de un enfoque basado en regresión (la propiedad o respuesta es cuantitativa y está disponible en una escala continua) o en un enfoque basado en clasificación (la propiedad o respuesta es categórica) (Roy et al., 2015b). A lo largo de los años ha habido un mayor interés entre los investigadores por utilizar este enfoque para modelar los fenómenos de retención en cromatografía, ya que es útil para predecir el t_R o de compuestos no evaluados o no sintetizados. Asimismo, es útil para preparar y optimizar experimentos cromatográficos con el fin de separar los compuestos de mezclas complejas e identificar posibles fármacos candidatos a partir de productos químicos sintetizados o diseñados por computadora (Rojas et al., 2015b, 2015a, 2018, 2019, 2022).

Para el desarrollo de los modelos QSPR, la estructura química de los compuestos debe ser legible por la computadora, de tal forma que retenga la mayor cantidad de información estructural posible. Para este propósito, se utiliza la especificación de introducción lineal molecular simplificada (SMILES, *Simplified Molecular Input Line Entry System*) para calcular diversos descriptores

moleculares. La topología molecular, que representa la estructura de la molécula a partir de un grafo molecular, se puede utilizar para definir varios tipos de descriptores moleculares independientes de la conformación (Katritzky et al., 2010; Rojas et al., 2018, 2019, 2021; Rojas & Duchowicz, 2021). Existen 4 pasos básicos para llevar acabo un estudio QSPR (Roy et al., 2015b):

- Preparación de los datos: se basa en generar una base de datos de un conjunto de compuestos químicos junto con una respuesta de interés (por ejemplo, pesticidas y tiempos de retención) y se obtienen una representación de las estructuras químicas (necesaria para el cálculo de descriptores).
- Procesamiento de los datos: donde se realiza un curado de la base de datos, se seleccionan los mejores descriptores y se aplica el aprendizaje automático.
- Validación interna y externa del modelo.
- Interpretación de los datos: definición del dominio de aplicabilidad e interpretación del mecanismo de acción.

1.5 Descriptores moleculares

"*El descriptor molecular es el resultado final de un procedimiento lógico y matemático que transforma la información química codificada dentro de una representación simbólica de una molécula en un número útil o el resultado de algún experimento estandarizado*" (Todeschini & Consonni, 2009). Los descriptores moleculares son un conjunto de parámetros que tienen la capacidad de codificar una estructura molecular de forma cuantitativa y se pueden extraer a partir de distintas formas de representación de una molécula. En consecuencia, se pueden distinguir dos grandes clases de descriptores: los experimentales, obtenidos mediante experimentos estandarizados, y los teóricos que se obtienen aplicando algoritmos matemáticos bien definidos a una representación inequívoca de la estructura molecular (Todeschini & Consonni, 2009). La cantidad de los descriptores depende del tipo de algoritmo empleado y define la naturaleza del análisis QSAR (Roy & Das, 2014). De forma general, los descriptores moleculares se pueden categorizar en tres grupos 1D, 2D y 3D (Lipkowitz & Boyd, 2007; Todeschini et al., 1994; Xue & Bajorath, 2012). Los descriptores moleculares 1D presentan las propiedades generales de los compuestos, como el número de átomos específicos, el peso molecular, entre otros; y se pueden calcular únicamente en función de una fórmula molecular. Los descriptores moleculares 2D presentan información estructural que se puede calcular a partir de un grafo molecular (representación 2D de una molécula) mediante la aplicación de diversos operadores invariantes. Los descriptores moleculares 3D presentan información estructural que se deriva de una representación 3D de una molécula, como el área superficial accesible al solvente con carga parcial positiva en la estructura.

Los descriptores moleculares independientes de la geometría han surgido como una herramienta alternativa para el desarrollo de modelos, no requieren grandes recursos computacionales y un período de tiempo prolongado para experimentos computacionales en comparación con los que

dependen de la geometría (Katritzky et al., 2010; Rojas et al., 2018, 2019, 2021; Rojas & Duchowicz, 2021). Actualmente se dispone de miles de descriptores y la principal dificultad a resolver en el modelado QSPR es la correcta selección de un conjunto reducido y representativo de descriptores moleculares, de tal forma de desarrollar un modelo que sea apto para explicar y predecir de mejor manera la propiedad bajo estudio (Todeschini & Consonni, 2008). Se ha argumentado que los descriptores moleculares 3D funcionan mejor que los descriptores moleculares 2D especialmente en aplicaciones en donde la estructura 3D de un compuesto, incluida la estereoquímica absoluta, es fundamental para la interacción con un receptor (acoplamiento molecular) (Hong et al., 2008). Sin embargo, en algunos estudios comparativos los descriptores 2D funcionan de manera equivalente a los descriptores 3D (Matter & Pötter, 1999; McGregor & Pallai, 1997), de manera particular cuando se estudian propiedades cromatográficas (Rojas et al., 2015a, 2015b, 2019, 2021).

1.6 Regresión lineal múltiple

La regresión lineal múltiple (MLR, *Multiple Linear Regression*) es un enfoque matemático que permite establecer relaciones lineales entre un grupo de variables explicativas (independientes) y una variable dependiente o respuesta cuantitativa (Rencher & Schaalje, 2001). Estos métodos ayudan a realizar predicciones de las *n* observaciones de una matriz **X**, mediante el uso de un vector de respuestas experimentales **Y**, en donde cada objeto n_i representa un vector descrito por una variable *p* en la matriz **X,** al cual se le asigna un valor de salida y_i en el vector **Y** (Rojas & Duchowicz, 2021)**.** En el modelado de fenómenos de retención los objetos n_i representan los compuestos químicos, las *p* variables los descriptores moleculares y el valor de salida y_i son los tiempos o índices de retención. El método de mínimos cuadrados ordinario (OLS, *Ordinary Least Squares*) es el método más simple y el que se aplica con mayor frecuencia cuando se busca realizar una regresión lineal. El método OLS brinda una estimación de los coeficientes de una regresión mediante la minimización de la suma de los residuos al cuadrado entre el vector de respuestas experimentales y el vector de respuestas calculadas (Rojas & Duchowicz, 2021).

CAPITULO II
MATERIALES Y MÉTODOS

2.1 Generación de la base de datos

Para generar la base de datos que se utilizó durante el modelado se recopiló la información experimental del artículo científico titulado *"High-throughput simultaneous analysis of pesticides by supercritical fluid chromatography coupled with high-resolution mass spectrometry"* (Ishibashi et al., 2015). Aquí se encuentra disponible la información de 508 pesticidas identificados en muestras de espinacas con sus respectivos tiempos de retención (t_R). El sistema cromatográfico usa el $SCCO_2$ como fase móvil y la columna de fase inversa Inertsil ODS-EP (250 x 4,6 mm, tamaño de partícula = 5 µm). En una primera etapa de filtración de la base de datos, se excluyeron los pesticidas cuyos t_R no habían sido calculados debido a que no se cumplían con los criterios de selección de picos e identificación de compuestos determinados por los autores "*desviación de masa permitida de 5 ppm; desviación de intensidad permitida de todos los iones isotópicos esperados 10%; puntuación de coincidencia de patrones isotópicos > 70 %; ancho del tiempo de retención (t_R) <30 s; y umbral de respuesta de 5000*" (Ishibashi et al., 2015). A continuación, se optó por seguir las recomendaciones de los autores en donde indican que los compuestos detectados a una resolución de masa de 70000 m/Δm y precisión de masa > 5ppm presentan una mayor exactitud al ser analizados por los criterios mencionados anteriormente, por lo que se excluyeron los compuestos que no encontrados a esa resolución. Para los compuestos filtrados se obtuvo la información química adicional de la quimioteca de acceso libre PubChem (Kim et al., 2021), aquí se verificó el nombre del compuesto y su fórmula molecular. Además, para cada pesticida se extrajo el número de registro CAS (*Chemical Abstracts Service*), notación SMILES (*Simplified Molecular Input Line Entry System*) (canónico) y el PubChem CID.

2.2 Curado de las estructuras moleculares

Con la base de datos recopilada, se utilizó el programa alvaMolecule (Alvascience, 2020) para el curado de las estructuras moleculares. Este programa es capaz de identificar errores presentes en las estructuras moleculares, además de que aplica los siguientes criterios de curado: estandarizar los anillos de benceno en forma aromática, convertir enlaces covalentes inusuales en formas iónicas, agregar carga al átomo de nitrógeno cuaternario, eliminar/agregar hidrógenos excedentes/faltantes y estandarizar grupos nitro, azida y diazo. Como el objetivo es desarrollar un modelo QSPR independiente de la conformación, se fusionaron aquellos compuestos que presentan la misma representación SMILES (estereoisómeros) y se usó el tiempo la retención promedio para los mismos.

2.3 Representación molecular y cálculo de descriptores moleculares

Las estructuras moleculares de los pesticidas fueron diseñadas en el programa MarvinSketch (ChemAxon Ltd., 2022). El programa proporciona una representación de la molécula determinada a partir de la notación lineal de cadena SMILES buscando usar de forma particular una representación independiente de la geometría molecular. Dicha representación es la base para el cálculo de los diversos tipos de descriptores moleculares. Con la base de datos curada y las estructuras moleculares diseñadas se calcularon 4146 descriptores moleculares independientes de la conformación usando el programa alvaDesc (Alvascience, 2022). En este mismo programa, se realizó una reducción de la dimensionalidad de los datos mediante la eliminación de descriptores con valores constantes, casi constantes, así como los descriptores con valores faltantes.

2.4 Reducción no supervisada de descriptores moleculares

Los métodos de reducción no supervisados de variables (no consideran la respuesta experimental), se usan para excluir aquellas que presentan ruido, redundancia y multicolinealidad; es decir, selecciona las variables más representativas de tal forma que se preserve la mayor cantidad de información. Una alternativa para reducir variables es el método V-WSP (Ballabio et al., 2014), el cual es una modificación del algoritmo propuesto por *Wootton, Sergent y Phan-Tan-Luu* (WSP) para diseño de experimentos. El método V-WSP selecciona un subconjunto representativo de variables, de tal forma que se encuentre una mínima correlación entre las mismas (por ejemplo, un umbral de 0.95) dentro de un espacio multidimensional definido (Ballabio et al., 2014).

2.5 Selección supervisada de descriptores moleculares

Con la finalidad de establecer la relación QSPR se utilizó la regresión de mínimos cuadrados ordinarios (OLS). Para encontrar un modelo parsimonioso se recurrirá al Método de Reemplazo (RM, *Replacement Method*) (Duchowicz et al., 2006) destinado a la selección supervisada de descriptores. Este algoritmo del aprendizaje supervisado optimiza (minimiza) la desviación estándar residual (*s)* durante el reemplazo de los descriptores. En el RM se calcula el error relativo de cada coeficiente de regresión dentro de un modelo de dimensión *d* (número de variables presentes en la ecuación), para posteriormente reemplazar los descriptores que presenten el mayor error relativo por aquellos presentes en el conjunto total *D* (número total de descriptores disponibles).

2.6 Validación del modelo

Para la validación del modelo QSPR para los t_R de los pesticidas, la base de datos se dividió en tres grupos: calibración, validación y predicción; de acuerdo al método de subconjuntos balanceados (BSM, *Balanced Subsets Method*) (Rojas et al., 2015a; Rojas & Duchowicz, 2021). Este método se basa en el análisis de conglomerados *k*-medias (*k*-MCA, *k-Means Cluster Analysis*) para crear *k*-

conglomerados o grupos de compuestos, de tal manera que los compuestos en el mismo conglomerado son muy similares en términos de una medida de distancia (en este caso la euclidiana), y los compuestos en diferentes conglomerados serán distintos entre sí. La partición BSM considera la propiedad experimental y los descriptores independientes de la geometría después reducción de aquellos altamente correlacionados. El grupo de calibración se usó para ajustar el modelo, mientras que el grupo de validación se consideró para controlar el sobreajuste del modelo. Finalmente, el grupo de predicción se empleó para evaluar la capacidad predictiva del modelo QSPR desarrollado.

Adicionalmente, el modelo se validó mediante protocolos de validación interna, por ejemplo dejar-uno-fuera (loo, *leave-one-out*) y dejar-varios-fuera (lmo, *leave-many-out*). En loo se excluye una molécula del modelo a la vez, luego se recalibra el modelo y se utiliza para predecir su propiedad. De forma análoga, en lmo se excluye un porcentaje definido de moléculas (por ejemplo, el 20%) y se usan las restantes (80%) para recalibrar el modelo y seguidamente predecir la propiedad de las moléculas excluidas. Finalmente, se aplicará la aleatorización-Y (Rücker et al., 2007) para descartar la presencia de correlación casual en el modelo. El desarrollo y validación del modelo se realizó en el programa alvaModel (Alvascience, 2021).

2.7 Grado de contribución de descriptores moleculares

Para medir el grado de contribución de los descriptores presentes en el modelo se estandarizarán los coeficientes de regresión lineal de los mismos. De tal forma que mientras más alto sea el valor absoluto del coeficiente para un descriptor determinado, mayor será la importancia de tal descriptor para predecir el tiempo de retención.

2.8 Mecanismo de acción (interpretación de los descriptores)

El mecanismo de acción de un modelo QSPR se basa en establecer una ruta por la cual la propiedad experimental se describe mediante los descriptores moleculares del modelo (siempre que sea posible). Para este propósito, se realiza una explicación de la definición de cada descriptor molecular y la forma en que se relaciona con el t_R, ya sea de forma sinérgica o antagónica.

2.9 Análisis del dominio de aplicabilidad

El dominio de aplicabilidad (AD, *Applicability Domain*) es una región teórica del espacio químico definida por los descriptores del modelo y la respuesta modelada. El AD se describe en términos del conjunto de calibración del modelo, el cual es aplicable para realizar predicciones de nuevos compuestos que se encuentran dentro de este espacio químico teórico (Rojas & Duchowicz, 2021; Roy et al., 2015b) . Los modelos QSPR son modelos locales (no universales), por lo que el AD del modelo juega un papel decisivo en la estimación de la incertidumbre durante la predicción de la

propiedad de un nuevo compuesto. Por lo tanto, la predicción de una respuesta usando el modelo QSPR es aplicable solo si el compuesto que se predice cae dentro del dominio del modelo. El AD está definido por el valor de influencia (*leverages*) el cual mide la distancia de cada objeto del grupo de predicción con respecto al centro del modelo (Rencher & Schaalje, 2001; Rojas & Duchowicz, 2021). Para determinar el AD se debe definir un valor de umbral crítico calculado como tres veces el valor promedio de los valores de influencia del conjunto de calibración, es decir, $h^* = 3p/n$ (p es el número de parámetros del modelo y n es el número de moléculas del conjunto de calibración). De esta forma, moléculas que superen este umbral son las que se encuentran distantes del centro del modelo y, por lo tanto, se las considera predicciones poco confiables o extrapolaciones del modelo. El análisis del dominio de aplicabilidad se realizó en el software alvaModel (Alvascience, 2021).

CAPITULO III
RESULTADOS Y DISCUSIONES

3.1 Generación de la base de datos y curado

Los datos experimentales se obtuvieron del artículo de Ishibashi y colaboradores (Ishibashi et al., 2015), donde se reportan el tiempo de retención de cada uno de los 508 pesticidas identificados en muestras de espinacas. Inicialmente se excluyeron los pesticidas etiquetados como “no encontrados” (n.f., *not found*) en una resolución de masa especificada de 70000. Posteriormente, para cada pesticida se realizó una verificación del nombre y se obtuvo el número de registro CAS y SMILES canónico en la quimioteca digital PubChem (Kim et al., 2021).

Con el propósito de obtener una base de datos estandarizada se procedió a analizar cada molécula con el software alvaMolecule (Alvascience, 2020) para que, después de implementar los criterios de curado, se le otorgue a cada molécula una cadena SMILES canónico estandarizada. Para filtrar la base de datos, se usó el SMILES canónico para identificar compuestos con la misma representación química, por ejemplo, *Bitertanol isómero 1 y Bitertanol isómero 2*. Para estos compuestos se calculó el promedio del t_R como propiedad experimental. De esta manera, se obtuvieron 375 pesticidas para el desarrollo de la relación cuantitativa estructura-propiedad.

3.2 Representación de la estructura molecular y cálculo de descriptores moleculares

Se calcularon 4146 descriptores moleculares independientes de la conformación en el programa alvaDesc (Alvascience, 2022). Para reducir los descriptores calculados, se excluyeron 1316 descriptores con valores constantes, 162 con valores casi constantes y 288 descriptores con al menos un valor faltante. De esta manera se obtuvieron 2380 descriptores moleculares para el desarrollo del modelo QSPR.

3.3 Reducción no supervisada de descriptores moleculares

Con los 2380 descriptores moleculares se aplicó la reducción de variables V-WSP implementado en MATLAB, el cual reduce la presencia de ruido, redundancia y multicolinealidad en los datos. Se eligió un umbral de correlación de 0.95 (*thr* = 0.95). La reducción consiste en comparar el coeficiente absoluto de correlación de Pearson (R_{ij}) y el valor de umbral preestablecido (*thr* = 0.95). Es decir, si se cumple que $R_{ij} \geq thr$, se eliminará el descriptor que presenta una mayor correlación promedio con las demás variables. De esta manera se excluyeron 1565 descriptores moleculares para la etapa posterior de desarrollo del modelo QSPR.

3.4 Reducción supervisada de descriptores moleculares

Para el desarrollo del modelo, la base de datos fue divida en grupos de calibración y predicción con 262 moléculas y 113 compuestos, respectivamente. Para la selección supervisada de descriptores mediante el método de reemplazo se utilizó el grupo de calibración para ajustar los modelos y se utilizó la validación cruzada de dejar-uno-fuera para controlar la presencia de sobreajuste en el modelo. La selección mediante el RM se aplicó en dos etapas: 1) se corrió el RM de forma separada en cada bloque de descriptores moleculares y 2) los mejores descriptores de la primera selección se fusionaron para aplicar nuevamente el RM y encontrar la mejor solución de descriptores para el tiempo de retención. En la Tabla 1 se presenta el número de descriptores seleccionados para cada bloque.

Tabla 1. Cantidad de descriptores moleculares para cada bloque obtenidos mediante el método de reemplazo.

Bloque	Familia de Descriptor Molecular	Calculado	Seleccionado
B1	Descriptores constitucionales	55	16
B2	Descriptores de anillo	10	6
B3	Índices topológicos	55	22
B4	Número de trayectos moleculares	55	20
B6	Índices de información	55	26
B7	Descriptores basados en la matriz 2D	28	19
B8	Autocorrelaciones 2D	55	26
B9	Autovalores de Burden	55	22
B10	Descriptores tipo P_VSA	55	12
B11	Índices del átomo topoquímico ampliado	36	13
B12	Índices de adyacencia de arista	55	23
B21	Número de grupos funcionales	55	11
B22	Fragmentos centrados en el átomo	55	15
B23	Índices del estado electrotopológico atómico	6	3
B24	Descriptores farmacóforos	10	8
B25	Pares de átomos 2D	55	22

B28	Propiedades moleculares	10	7
B32	Descriptores MDE	55	14
B33	Descriptores de quiralidad	55	21
	TOTAL	**815**	**306**

A partir de los mejores 306 descriptores moleculares seleccionados para cada bloque, el RM encontró un modelo óptimo de seis descriptores. La selección del modelo óptimo se realizó considerando los parámetros de calidad del grupo de calibración y validación, es decir menor desviación estándar residual (s) y el mayor coeficiente de correlación (R^2) en validación. Asimismo, se buscó que el coeficiente de correlación máximo entre descriptores (R^2_{ijmax}) sea el más bajo posible. El modelo obtenido es:

$$t_R = -23.010 + 46.418 \times Mv + 0.131 \times nH - 0.250 \times X\% + 0.062 \times P\_VSA\_ppp\_D - 0.965 \times SM05\_EA(ri) + 1.383 \times nRNR2$$

Ecuación (1)

Los seis descriptores del modelo QSPR para la predicción del tiempo de retención (Ecuación 1) son:

- *Mv*: volumen atómico de van der Waals promedio (escalado con el átomo de carbono).
- *X%*: porcentaje de átomos de halógeno.
- *P_VSA_ppp_D*: *P_VSA* de potenciales puntos farmacólogos (donantes de enlaces de hidrogeno *D*).
- *nH*: número de átomos de hidrógeno.
- *SM05_EA(ri)*: momento espectral de quinto orden de la matriz de adyacencia ponderando por la integral de resonancia.
- *nRNR2*: número de aminas terciarias (alifáticas).

3.5 Validación del modelo

Los parámetros estadísticos para el grupo de calibración son R^2 = 0.713 y *RMSEC* = 1.235 y los de predicción son R^2 = 0.738 y *RMSEP* = 1.1792. El modelo de la ecuación (1) fue sometido a varios enfoques de validación cruzada: dejar-uno-fuera (R^2 = 0.695, RMSECV = 1.2731), ventanas venecianas con 5 grupos (R^2 = 0.691, RMSECV = 1.282), Monte Carlo, dejando un 20% fuera con 1000 iteraciones (R^2 = 0.697, RMSECV = 1.272), Bootstrap con 1000 iteraciones (R^2 = 0.686, RMSECV = 1.2957). La mínima diferencia entre los resultados de bondad de ajuste del modelo y la

estabilidad interna frente a perturbaciones del modelo (validación interna) reflejan que la relación cuantitativa estructura-propiedad no presenta sobreajuste. Complementariamente, la aleatorización-Y permitió verificar la estabilidad del modelo luego de 1000 iteraciones (R^2 = 0.024, RMSECV = 2.277).

En la Figura 2, se observa la relación entre el t_R experimental y t_R predicho obtenido con la Ecuación (1). Se observa que el t_R tiene una tendencia de relación lineal alrededor de la línea de ajuste perfecto.

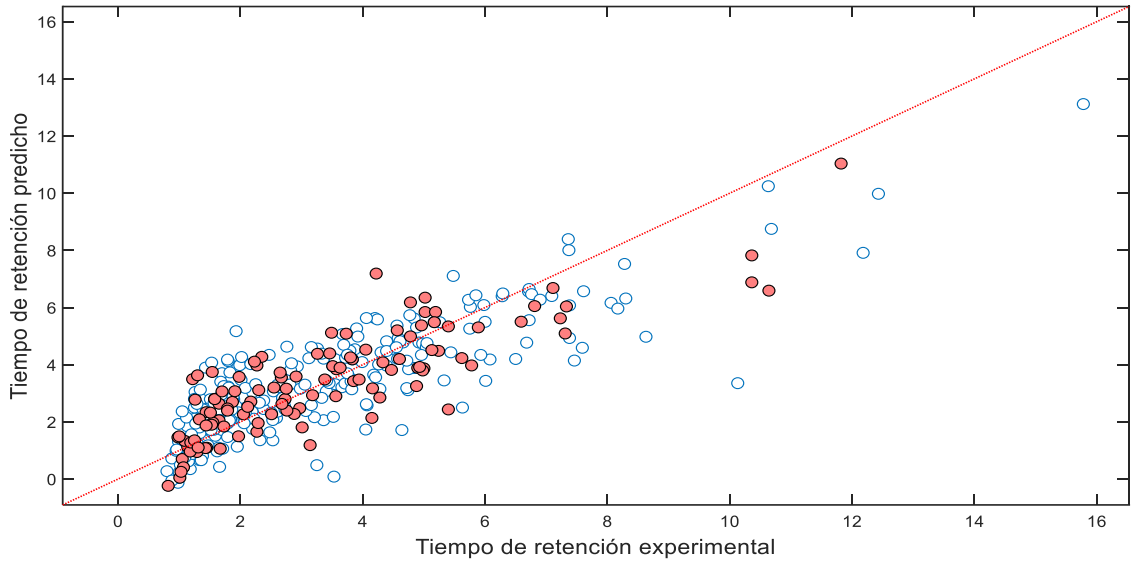

Figura 2. Tiempos de retención experimental vs. predicho de pesticidas medidos en la columna de fase inversa Inertsil ODS-EP. Las moléculas del grupo de calibración se encuentran en azul con fondo blanco y las del grupo de predicción se encuentran en rojo.

Por otra parte, en la Figura 3 se presenta la dispersión de los residuos estandarizados frente al tiempo de retención predicho. La distribución aleatoria de los residuos estandarizados alrededor de la línea cero indica que el modelo de mínimos cuadrados ordinarios es apropiado para modelar el t_R. En este modelo, las moléculas metil paratión y tidiazuron se encuentran fuera del límite de ± 3 RMSC.

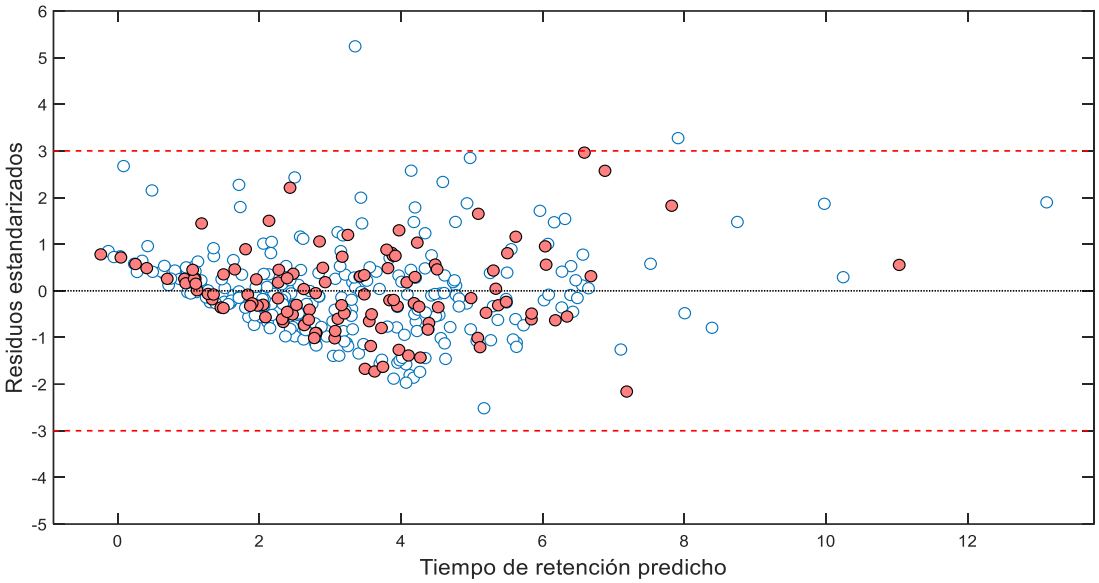

Figura 3. Gráfica de dispersión de residuos estandarizados vs. el tiempo de retención predicho. Las moléculas del grupo de calibración se encuentran en azul con fondo blanco y las del grupo de predicción se encuentran en rojo.

3.6 Mecanismo de acción

Mediante el cálculo de los coeficientes de regresión estandarizados se determinó el grado de contribución de cada descriptor para predecir el t_R: *Mv (1.098) > X% (-0.543) > P_VSA_ppp_D (0.452) > nH (0.317) > SM05_EA(ri) (-0.224) > nRNR2 (0.142)*. Cuatro descriptores presentan un efecto sinérgico (positivo) en la predicción de esta propiedad (*Mv, P_VSA_ppp_D, nH, nRNR2*); mientras que los descriptores *X%, SM05_EA(ri)* presentan un efecto antagónico (negativo).

El volumen atómico de van der Waals promedio (escalado con el átomo de carbono) (*Mv)* es un descriptor de orden cero (0D). Es un índice constitucional calculado como la suma de los volúmenes de van der Waals para el número de átomos, tomando en cuenta que los átomos de carbono deben ser escalados. *Mv* está directamente relacionado con el volumen de van der Waals, es decir, el tamaño de la molécula (Hanai & Hubert, 1984). Debido al grado de contribución de este descriptor (positivo), mientras más alto sea el valor de *Mv* mayor será el t_R.

El porcentaje de átomos de halógeno (*X%*) tiene un comportamiento antagónico. Este descriptor tiene en cuenta el número de átomos de halógenos en el esqueleto químico de un pesticida. Los derivados con este tipo de átomos tienen la propiedad de ser solubles en compuestos no polares como el CO_2 (fase móvil) (Jeschke, 2022). Además, la incorporación de átomos de halógeno en estructuras moleculares es bien reconocida como un medio eficaz para controlar la densidad de electrones de una molécula y la disposición molecular en estructuras de agregados moleculares, induciendo una polarización (Morita et al., 2022).

Los descriptores tipo *P_VSA*, inicialmente propuestos por (Labute, 2000), son descriptores definidos como la cantidad de área superficial de van der Waals (VSA, *van der Waals Surface Area*) calculada como la suma de VSA_i de cada átomo *i* (sin contar el área contenida entre los atomos), donde a cada átomo se le otorga una propiedad numerica P_i con el fin obtener descriptores específicos dentro de un cierto rango o intervalo. Estos descriptores corresponden a una partición del área superficial molecular condicionada por los valores atómicos de la propiedad *P*. Por consiguiente, el descriptor de naturaleza sinérgica *P_VSA_ppp_D, P_VSA* considera donantes de enlaces de hidrógeno como la característica farmacóloga de un pesticida.

El número de átomos de hidrógeno (*nH*) es un descriptor de conteo del total de este tipo de átomos presentes en una molécula. *nH* está directamente relacionado con el tamaño de la molécula por lo que también se le considera un descriptor de tamaño, es decir, la contribución de los átomos de hidrógeno al volumen de una molécula es significativo; por ejemplo, el aumento del volumen en porcentaje del átomo de carbono (*C*) al grupo metilo (CH_3) es de alrededor del 82% (Buchwald, 2000). Si se toma en cuenta la relación de tamaño de la molécula con el t_R se observa que tanto el *Mv* y *nH* afectan de forma positiva.

El momento espectral de quinto orden de la matriz de adyacencia de borde, *SM05_EA(ri)*, se calcula como la suma de todos los trayectos de longitud 5 en un grafo lineal, comenzando y terminando en el mismo vértice (autorretorno). La matriz de adyacencia de bordes de un grafo molecular es idéntica a la matriz de adyacencia de vértices del grafo lineal (Estrada, 1996). La matriz de adyacencia de borde ponderada se obtiene al reemplazar los elementos correspondientes a enlaces adyacentes con propiedades de enlace específicas (Todeschini & Consonni, 2009). En este caso las integrales de resonancia, que son parámetros que miden la probabilidad de absorción en la zona de resonancias.

Finalmente, el número de aminas terciarias (alifáticas), *nRNR2*, es un descriptor que cuenta el número de grupos funcionales aminas terciarias alifáticas (>N−) (Grisoni et al., 2018). Por lo que, existirá una fuerte relación con la reactividad (basicidad, características nucleofílicas y electrofílicas) (Speck-Planche et al., 2011), debido a esto el descriptor *nRNR2* podría estar relacionado con la basicidad de un pesticida.

3.7 Dominio de aplicabilidad

El dominio de aplicabilidad define el espacio químico teórico dentro del cual el modelo realiza predicciones confiables de los índices de retención (interpolaciones). El valor de influencia estableció un umbral de 0.0668 (Figura 4), lo que significa que cualquier pesticida por debajo de este umbral se encuentra dentro del dominio de aplicabilidad, por consiguiente, las predicciones son el resultado de la interpolación del modelo. En este estudio se identificaron 9 moléculas que encuentran fuera del dominio (valor de influencia mayor al umbral crítico). Estas extrapolaciones se encuentran descritas en la Figura 5.

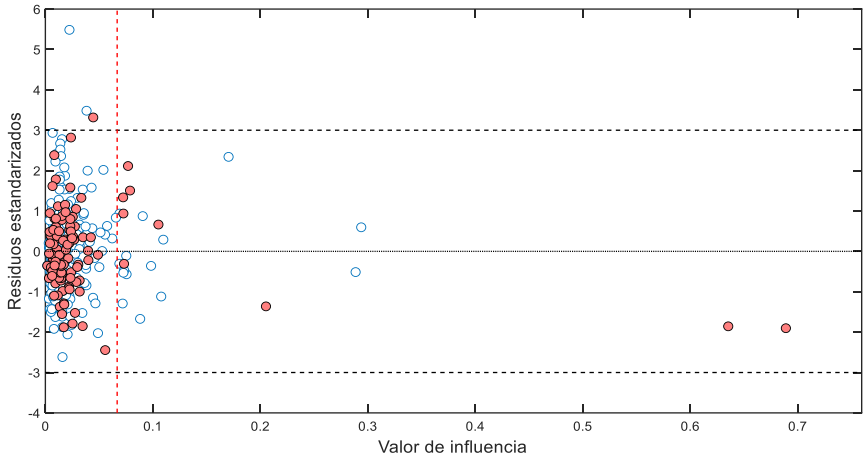

Figura 4. Diagrama de William: residuos estandarizados vs. valor de influencia, para definir el dominio de aplicabilidad del modelo QSPR. Las moléculas del grupo de calibración se encuentran de blanco y las del grupo de predicción se encuentran en rojo.

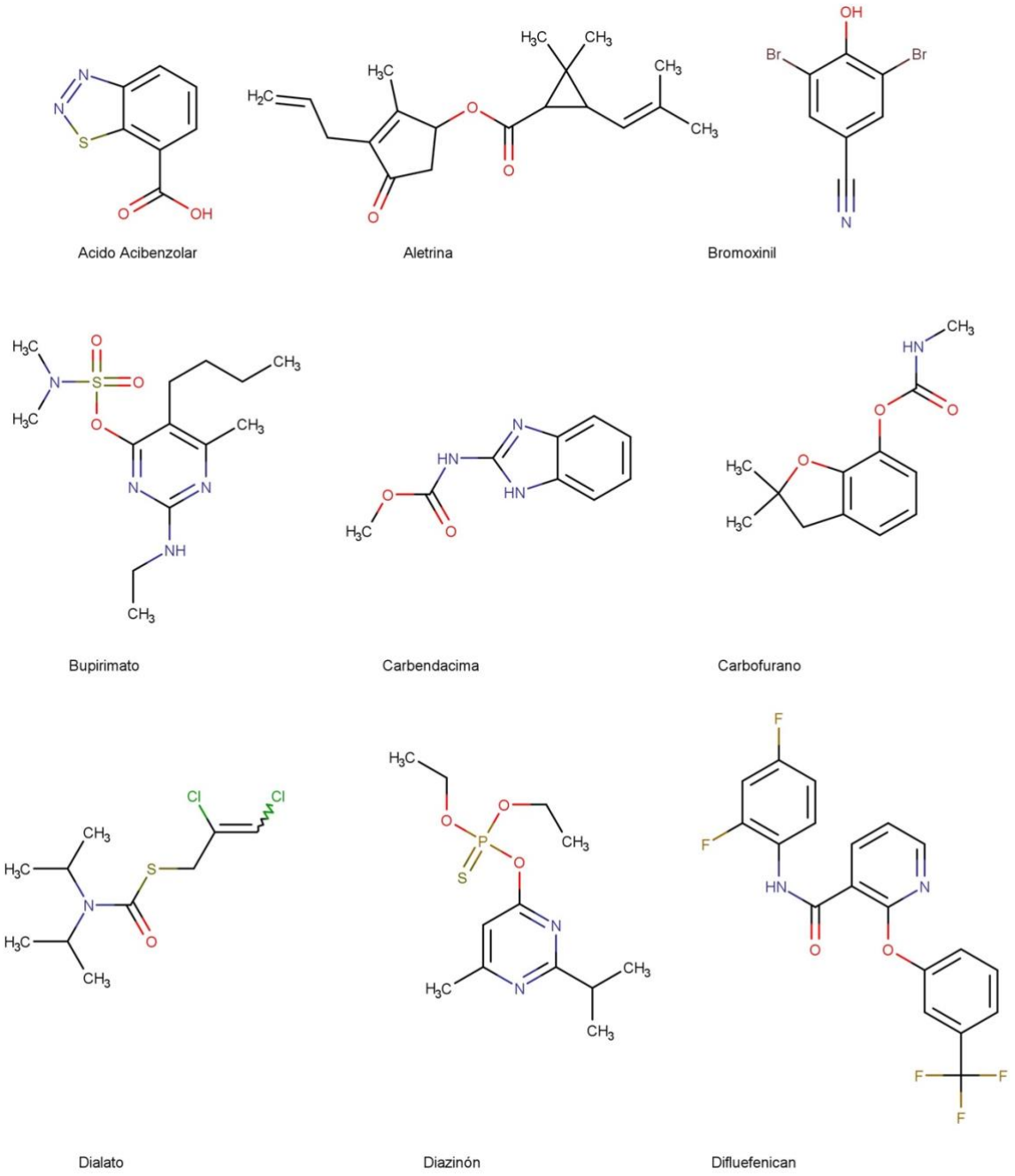

Figura 5. Estructura química de los pesticidas que se ubican fuera del dominio de aplicabilidad del modelo QSPR. Fuente: Autor

CONCLUSIONES

En este trabajo se desarrolló un modelo quimioinformático basado en el enfoque QSPR para los tiempos de retención de 375 pesticidas identificados en extractos de espinacas. Se utilizó el SMILES canónico para calcular diversos descriptores moleculares independientes de la conformación. Para hacer frente a la gran cantidad de descriptores, el uso de la técnica de reducción variable no supervisada V-WSP permitió la exclusión de los descriptores no informativos. Posteriormente, el método de reemplazo permitió la selección de seis descriptores óptimos. El modelo presenta una calidad descriptiva del 71.3 % y una capacidad predictiva del 73.8 %. Por lo tanto, este enfoque QSPR independiente de la conformación podría utilizarse por parte de los los investigadores en el área de química de los alimentos, en particular de quienes trabajan en la identificación de residuos de pesticidas en alimentos crudos o procesados, mediante cromatografía de fluidos supercríticos de alta resolución en la fase estacionaria Inertsil ODS-EP.

REFERENCIAS

Ali, U., Syed, J. H., Malik, R. N., Katsoyiannis, A., Li, J., Zhang, G., & Jones, K. C. (2014). Organochlorine pesticides (OCPs) in South Asian region: A review. *Science of The Total Environment*, *476–477*, 705–717.

Alvascience. (2020). *alvaMolecule (software to view and prepare chemical datasets)* (1.0.4). https://www.alvascience.com/alvamolecule.

Alvascience. (2021). *alvaModel (Software to Model QSAR Data)* (2.0.2). https://www.alvascience.com/ alvamodel.

Alvascience. (2022). *alvaDesc (software for molecular descriptors calculation) version* (2.0.12). https://www.alvascience.com/alvadesc.

Ballabio, D., Consonni, V., Mauri, A., Claeys-Bruno, M., Sergent, M., & Todeschini, R. (2014). A novel variable reduction method adapted from space-filling designs. *Chemometrics and Intelligent Laboratory Systems*, *136*.

Bhushan, C., Bhardwaj, A., & Misra, S. S. (2013). State of Pesticide Regulations in India. *Centre for Science and Environment, JSTOR*, *2013*, 1–72.

Buchwald, P. (2000). Modeling liquid properties, solvation, and hydrophobicity: A molecular size-based perspective. *Perspectives in Drug Discovery and Design*, *19*(1), 19–45.

ChemAxon Ltd. (2022). *MarvinSketch* (22.11). https://www.chemaxon.com.

Duchowicz, P. R., Castro, E. A., & Fernández, F. M. (2006). Alternative algorithm for the search of an optimal set of descriptors in QSAR-QSPR studies. *Match*, *55*(1).

Estrada, E. (1996). Spectral moments of the edge adjacency matrix in molecular graphs. 1. Definition and applications to the prediction of physical properties of alkanes. *Journal of Chemical Information and Computer Sciences*, *36*(4).

FAO. (2020). *Crops and livestock products, production quantities of Spinach by country*.

Gangwal, R. P., Damre, M. V., & Sangamwar, A. T. (2016). Overview and Recent Advances in QSAR Studies. In *Chemometrics Applications and Research* (1st ed., pp. 29–60). Apple Academic Press.

Gere, D. R. (1983). Supercritical fluid chromatography. *Science*, *222*(4621), 253–259.

Gilden, R. C., Huffling, K., & Sattler, B. (2010). Pesticides and health risks. *Journal of Obstetric, Gynecologic, and Neonatal Nursing : JOGNN*, *39*(1), 103–110.

Grisoni, F., Consonni, V., & Todeschini, R. (2018). Impact of Molecular Descriptors on Computational Models. In *Methods in Molecular Biology* (Vol. 1825).

Hanai, T., & Hubert, J. (1984). Retention versus van der waals volume and π energy in liquid chromatography. *Journal of Chromatography A*, *290*(C).

Hong, H., Xie, Q., Ge, W., Qian, F., Fang, H., Shi, L., Su, Z., Perkins, R., & Tong, W. (2008). Mold2, molecular descriptors from 2D structures for chemoinformatics and toxicoinformatics. *Journal of Chemical Information and Modeling*, *48*(7), 1337–1344.

INEN. (2013). *Norma Técnica Ecuatoriana NTE INEN 1834:2013 Primera revisión*.

Ishibashi, M., Izumi, Y., Sakai, M., Ando, T., Fukusaki, E., & Bamba, T. (2015). High-throughput simultaneous analysis of pesticides by supercritical fluid chromatography coupled with high-resolution mass spectrometry. *Journal of Agricultural and Food Chemistry*, *63*(18), 4457–4463.

Jeschke, P. (2022). Manufacturing Approaches of New Halogenated Agrochemicals. In *European Journal of Organic Chemistry* (Vol. 2022, Issue 12).

Katritzky, A. R., Kuanar, M., Slavov, S., Hall, C. D., Karelson, M., Kahn, I., & Dobchev, D. A. (2010). Quantitative correlation of physical and chemical properties with chemical structure: Utility for prediction. *Chemical Reviews*, *110*(10), 5714–5789.

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., Zaslavsky, L., Zhang, J., & Bolton, E. E. (2021). PubChem in 2021: New data content and improved web interfaces. *Nucleic Acids Research*, *49*(D1).

Kubinyi, H. (2008). *QSAR : Hansch analysis and related approaches* (1st ed., Vol. 1). VCH.

Labute, P. (2000). A widely applicable set of descriptors. *Journal of Molecular Graphics and Modelling*, *18*(4–5).

Lipkowitz, K. B., & Boyd, D. B. (2007). Reviews in Computational Chemistry. In *Reviews in Computational Chemistry* (Vol. 9).

Matter, H., & Pötter, T. (1999). Comparing 3D Pharmacophore Triplets and 2D Fingerprints for Selecting Diverse Compound Subsets. *Journal of Chemical Information and Computer Sciences*, *39*(6).

McGregor, M. J., & Pallai, P. V. (1997). Clustering of large databases of compounds: Using the MDL “keys” as structural descriptors. *Journal of Chemical Information and Computer Sciences*, *37*(3).

Morita, M., Yamada, S., & Konno, T. (2022). Halogen atom effect of fluorinated tolanes on their luminescence characteristics. *New Journal of Chemistry*, *46*(10).

Punzi, J. S., Lamont, M., Haynes, D., & Epstein, R. L. (2005). USDA pesticide data program: Pesticide residues on fresh and processed fruit and vegetables, grains, meats, milk, and drinking water. *Outlooks on Pest Management*, *16*(3), 131–137.

Rencher, A., & Schaalje, G. (2001). Linear Models in Statistics. In *John Wiley & Sons, Inc* (2nd ed., Vol. 96, Issue 455). John Wiley & Sons, Inc.

Rojas, C., Alcívar León, C. D., Contreras Aguilar, E., Mazón Ayala, P. V., & Muñoz, D. (2022). Quantitative Structure–Property Relationship for the Retention Index of Volatile and Semi-Volatile Compounds of Coffee. *Chemistry Proceedings*, 48.

Rojas, C., Aranda, J. F., Pacheco Jaramillo, E., Losilla, I., Tripaldi, P., Duchowicz, P. R., & Castro, E. A. (2021). Foodinformatic prediction of the retention time of pesticide residues detected in fruits and vegetables using UHPLC/ESI Q-Orbitrap. *Food Chemistry*, *342*(May 2020), 128354.

Rojas, C., & Duchowicz, P. R. (2021). *Química computacional de los alimentos : relaciones cuantitativas estructura-actividad/propiedad (QSAR/QSPR)* (1ra ed.). Editorial Acribia, S.A.

Rojas, C., Duchowicz, P. R., & Castro, E. A. (2019). Foodinformatics: Quantitative Structure-Property Relationship Modeling of Volatile Organic Compounds in Peppers. *Journal of Food Science*,

84(4), 770–781.

Rojas, C., Duchowicz, P. R., Tripaldi, P., & Diez, R. P. (2015a). QSPR analysis for the retention index of flavors and fragrances on a OV-101 column. *Chemometrics and Intelligent Laboratory Systems*, *140*.

Rojas, C., Duchowicz, P. R., Tripaldi, P., & Diez, R. P. (2015b). Quantitative structure-property relationship analysis for the retention index of fragrance-like compounds on a polar stationary phase. *Journal of Chromatography A*, *1422*.

Rojas, C., Tripaldi, P., Pérez-González, A., Duchowicz, P. R., & Pis Diez, R. (2018). A retention index-based QSPR model for the quality control of rice. *Journal of Cereal Science*, *79*, 303–310.

Roy, K., & Das, R. (2014). A Review on Principles, Theory and Practices of 2D-QSAR. *Current Drug Metabolism*, *15*(4), 346–379.

Roy, K., Kar, S., & Das, R. N. (2015a). *A Primer on QSAR/QSPR Modeling: fundamental concepts* (1st ed.). Springer International Publishing.

Roy, K., Kar, S., & Das, R. N. (2015b). Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment. In *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment* (1st ed.). Academic Press.

Rücker, C., Rücker, G., & Meringer, M. (2007). Y-randomization and its variants in QSPR/QSAR. *Journal of Chemical Information and Modeling*, *47*(6).

Speck-Planche, A., Guilarte-Montero, L., Yera-Bueno, R., Rojas-Vargas, J. A., García-López, A., Uriarte, E., & Molina-Pérez, E. (2011). Rational design of new agrochemical fungicides using substructural descriptors. *Pest Management Science*, *67*(4).

Taylor, L. T. (2009). Supercritical fluid chromatography for the 21st century. *The Journal of Supercritical Fluids*, *47*(3), 566–573.

Thayer, K. A., Heindel, J. J., Bucher, J. R., & Gallo, M. A. (2012). Role of environmental chemicals in diabetes and obesity: A national toxicology program workshop review. *Environmental Health Perspectives*, *120*(6), 779–789.

Todeschini, R., & Consonni, V. (2008). Descriptors from Molecular Geometry. In *Handbook of Chemoinformatics* (Vol. 3, pp. 1004–1033). John Wiley & Sons, Ltd.

Todeschini, R., & Consonni, V. (2009). *Molecular Descriptors for Chemoinformatics* (2nd ed., Vol. 1). John Wiley & Sons .

Todeschini, R., Lasagni, M., & Marengo, E. (1994). New molecular descriptors for 2D and 3D structures. Theory. *Journal of Chemometrics*, *8*(4), 263–272.

Vu-Duc, N., Nguyen-Quang, T., Le-Minh, T., Nguyen-Thi, X., Tran, T. M., Vu, H. A., Nguyen, L. A., Doan-Duy, T., Van Hoi, B., Vu, C. T., Le-Van, D., Phung-Thi, L. A., Vu-Thi, H. A., Chu, D. B., & Plaza-Bolaños, P. (2019). Multiresidue Pesticides Analysis of Vegetables in Vietnam by Ultrahigh-Performance Liquid Chromatography in Combination with High-Resolution Mass Spectrometry (UPLC-Orbitrap MS). *Journal of Analytical Methods in Chemistry*, *2019*.

WHO/UNEP. (1990). *Consecuencias sanitarias del empleo de plaguicidas en la agricultura* (p. 128).

Organización Mundial de la Salud.

Xue, L., & Bajorath, J. (2012). Molecular Descriptors in Chemoinformatics, Computational Combinatorial Chemistry, and Virtual Screening. *Combinatorial Chemistry & High Throughput Screening*, *3*(5).