



**UNIVERSIDAD  
DEL AZUAY**

**Facultad de Ciencia y Tecnología**

**Escuela de Ingeniería en Alimentos**

**Predicción *in silico* de los tiempos de retención de antibióticos  
identificados en muestras de leche mediante UHPLC/ESI Q-  
Orbitrap**

**Trabajo de graduación previo a la obtención del título de:**

**INGENIERO EN ALIMENTOS**

**Autor:**

**María Emilia Ayora Ortega**

**Director:**

**Dr. Cristian Rojas Villa**

**Cuenca – Ecuador**

**2023**

INDICE DE CONTENIDO	
INDICE DE TABLAS .....	4
AGRADECIMIENTOS .....	6
RESUMEN.....	7
ABSTRACT .....	8
Capítulo I .....	9
1 Marco Teórico .....	9
1.1 Leche .....	9
1.2 Antibióticos.....	9
1.3 Cromatografía líquida de ultra alto rendimiento .....	10
1.4 Modelado <i>in silico</i> .....	11
1.4.1 Generación de la base de datos.....	12
1.4.2 Modelos moleculares.....	12
1.4.3 Descripciones moleculares.....	14
1.4.4 Desarrollo del modelo .....	15
1.4.5 Interpretación del modelo.....	15
2 Materiales y métodos .....	17
2.1 Generación de la base de datos.....	17
2.2 Filtrado de la base de datos.....	17
2.3 Diseño y optimización de las geometrías moleculares.....	18
2.4 Cálculo de descriptores moleculares .....	19
2.5 Reducción no supervisada de descriptores moleculares .....	19
2.6 Desarrollo del modelo <i>in silico</i> .....	19
2.7 División de la base de datos.....	20
2.8 Validación del modelo .....	20
2.9 Grado de contribución de los descriptores moleculares .....	21
2.10 Mecanismo de acción.....	21
2.11 Análisis del dominio de aplicabilidad.....	21
Capítulo III .....	23
3 Resultados y discusiones .....	23
3.1 Generación de la base de datos y curado .....	23
3.2 Cálculo de descriptores moleculares.....	23
3.3 Desarrollo y validación del modelo <i>in silico</i> .....	23
3.4 Mecanismo de acción.....	26
3.5 Dominio de aplicabilidad .....	27
CONCLUSIONES.....	28

REFERENCIAS .....29

## INDICE DE FIGURAS

<b>FIGURA 1.1</b> Representación química del hidrógeno, deuterio y tritio. ....	14
<b>FIGURA 2.1</b> Antibióticos que tienen deuterio dentro de la estructura química. ....	18
<b>FIGURA 3.1</b> Modelo <i>in silico</i> para los antibióticos detectados mediante UHPLC. (a) tiempos de retención experimental versus tiempos de retención predichos. (b) dispersión de los residuos estandarizados vs. los tiempos de retención predichos. ....	25
<b>FIGURA 3.2</b> Antibióticos que presentan diferencia alta entre el tiempo de retención experimental y predicho por el modelo. ....	26
<b>FIGURA 3.3</b> Diagrama de William para definir el dominio de aplicabilidad del modelo <i>in silico</i> : valor de influencia vs. residuos estandarizados. ....	27

## INDICE DE TABLAS

<b>TABLA 1</b> Resultados del modelo <i>in silico</i> en calibración, validación cruzada y predicción. ....	24
---	----

## **DEDICATORIA**

Esta tesis va dedicada a mi abuelita Cecilia Toledo y mi tía Anita Ayora quienes desde el cielo siempre me protegieron y me alentaron. A mis padres José Ayora Toledo y Priscila Ortega Ormaza por su incondicional apoyo, por las enseñanzas que obtuve durante toda esta etapa de mi vida y por ser mi mayor fuente de inspiración; a mis hermanos José y Pablo quienes, con su amor y apoyo incondicional no me han dejado rendirme y estuvieron siempre alentándome. Todo esto es dedicado netamente a mi familia.

## **AGRADECIMIENTOS**

Quiero agradecer a Dios porque ha sido guía durante toda esta etapa de mi vida, a mis padres, porque sin su apoyo nada de esto hubiera sido posible, gracias por su amor incondicional, por enseñarme a nunca darme por vencida, por inculcarme valores que me han servido a lo largo de mi vida. A mis hermanos, por sus consejos, sus palabras de aliento y por estar siempre en cada momento de mi vida. A cada uno de los docentes que fueron parte de mi etapa universitaria y de manera especial a quien me guío durante este trabajo Dr. Cristian Rojas Villa, por brindarme sus conocimientos y su sabiduría. A mis compañeros con los que pude compartir experiencias y que día a día me enseñaban algo nuevo, gracias por los buenos y malos momentos. De manera especial a mi grupo de amigas “La Mafia” que me enseñaron el significado verdadero de la amistad, por su apoyo, por las risas y lágrimas compartidas, dejaron una huella en mi corazón. A toda mi familia que se ha hecho presente en cada logro demostrando su cariño sincero. Y a la Universidad del Azuay que me permitió conocer a grandes personas que quedaran guardadas en mi corazón por siempre y por dejarme los mejores recuerdos, experiencias de mi vida.

## RESUMEN

En este estudio, se utilizó una base de datos de 130 antibióticos identificados en muestras de leche para calibrar un modelo *in silico* basado en las relaciones cuantitativas de estructura-propiedad (QSPR). La propiedad experimental fue el tiempo de retención ( $t_R$ ) determinado mediante cromatografía líquida de ultra alta resolución (UHPLC). Para la optimización molecular, se utilizó el método semiempírico Tight Binding GFN2-xTB, de tal forma de calcular diversos descriptores moleculares dependientes de la conformación, los que se analizaron mediante la reducción no supervisada W-VSP. Posteriormente, se calibró una regresión lineal múltiple (MLR) utilizando la selección de variables mediante los algoritmos genéticos (GAs). Estos procedimientos se utilizaron para establecer un modelo de tres descriptores con buena calidad en calibración, validación cruzada y predicción. Además, para cumplir con los principios de QSPR con fines regulatorios, se definió el dominio de aplicabilidad y se proporcionó la interpretación del mecanismo de acción de los descriptores.

**Palabras clave:** *muestras de leche, antibióticos, tiempo de retención, UHPLC, in silico*



Dr. Cristian Rojas Villa  
**Director del Trabajo de Titulación**



Mgtr. María Fernanda Rosales Medina  
**Directora de Escuela**

Emilia Ayora.

María Emilia Ayora Ortega

**Autora**

## ABSTRACT

In this study, a database 130 antibiotics identified in milk samples was used to calibrate an *in silico* model based on quantitative structure-property relationships (QSPR). The experimental property was the retention time ( $t_R$ ) determined by means of ultrahigh performance liquid chromatography (UHPLC). For molecular optimization, the semi-empirical Tight Binding GFN2-xTB method was used in order to calculate several conformation-dependent molecular descriptors, which were analyzed by means of the W-VSP unsupervised approach. Afterwards, a multiple linear regression (MLR) was calibrated along with the Genetic Algorithms (GAs) variable subset selection. These procedures were used to establish a three-descriptor model with acceptable quality in fitting, cross-validation and prediction. In addition, to accomplish the QSAR's principles for regulatory purposes, the applicability domain was defined and the mechanistic interpretation of the descriptors was provided.

**Keywords:** milk samples, antibiotics, retention time, UHPLC, *in silico*



Dr. Cristian Rojas Villa  
**Thesis Supervisor**



Mgt. María Fernanda Rosales Medina  
**Faculty Director**

Emilia Ayora.

María Emilia Ayora Ortega

**Author**



## Capítulo I

### 1 Marco Teórico

#### 1.1 Leche

Por definición la leche es un líquido blanco que segregan de manera natural las glándulas mamarias de las vacas sanas o de cualquier otra especie, excluido el calostro y que es útil para alimento de sus crías (Badui, 2016). La leche se caracteriza por ser un alimento de consumo masivo a nivel mundial debido a que posee un valor nutricional alto, por lo que es considerado como un alimento completo y equilibrado dentro de la dieta humana. La leche proporciona un elevado contenido de nutrientes, tales como grasas, proteínas de alto valor biológico, minerales especialmente calcio y fósforo, hidratos de carbono (por lo general en forma de lactosa), vitaminas liposolubles y vitaminas del complejo B (Badui, 2016).

Al ser un alimento de consumo masivo, un problema que se presenta durante la producción de la leche es la presencia de antibióticos, los cuales son suministrados a las vacas para evitar diversos tipos de enfermedades, por ejemplo la mastitis (Máttar *et al.*, 2009; Salas *et al.*, 2013). La calidad de la leche y sus derivados se determinada por medio de las características fisicoquímicas y bacteriológicas, las cuales están reflejadas por la composición de los productos. Para que la leche se considere de buena calidad deber reunir las siguientes características (Badui, 2016; Damodaran & Parkin, 2018):

- Poseer una adecuada composición nutricional (contenido de proteínas, minerales, vitaminas, porcentaje de grasa, número de sólidos totales).
- Estar libre de sustancias extrañas y de residuos químicos e inhibidores (antibióticos, pesticidas y otros).
- Ausencia de patógenos extraños (por ejemplo, *Salmonella* o *E. coli*).
- Poseer adecuadas características organolépticas (sabor, olor, color).

Además, su calidad depende de las condiciones climáticas, de factores fisiológicos de los animales que la producen, así como la salud del animal; de las condiciones de salubridad en las que se ordeña y del manejo dado al producto hasta llegar al consumidor. También puede existir el almacenamiento de residuos de compuestos químicos, por ejemplo, en sus tejidos u órganos animales (Moncada Jiménez & Pelayo Consuegra, 2011). Es muy importante garantizar la buena calidad de la leche, ya sea para consumo directo o para los diversos procesos productivos de transformación en derivados lácteos (Moncada Jiménez & Pelayo Consuegra, 2011).

#### 1.2 Antibióticos

Los antibióticos se definen como sustancias químicas que tienen la capacidad de paralizar el desarrollo de ciertos microorganismos patógenos debido a su acción bacteriostática, o de

causarles la muerte mediante acción bactericida. Dichas sustancias pueden ser utilizadas con los siguientes fines: para el tratamiento o prevención de enfermedades infecciosas, estimulantes de crecimiento y otras (Fernández *et al.*, 2008).

La contaminación de la leche con residuos de antibióticos es considerada un problema de salud pública a nivel mundial. Es por ello que surge la importancia de controlar la presencia de residuos de antibióticos en los alimentos con el fin de evitar problemas de resistencia de la población a estos compuestos químicos y evitar la muerte de las bacterias ácido lácticas que se usan como cultivos fermentativos en la industria láctea. Hay que mencionar, además, que la presencia de estos residuos se debe a que no respetan la dosis permisibles y los tiempos de espera para la recuperación del ganado vacuno. Por lo general los residuos más comunes en la leche suelen ser sulfonamidas y nitrofuranos, que son los principales agentes usados para el control de la mastitis (inflamación de la ubre), y los plaguicidas organofosforados que se aplican como ectoparasiticidas para el control de moscas y garrapatas (Salas *et al.*, 2013).

La presencia de estos residuos, aparte de ser perjudicial para la salud del consumidor también afectan los procesos de industrialización de la leche, debido a que inhiben o retardan el desarrollo de las bacterias lácticas ocasionando pérdidas en las industrias lácteas. Por ejemplo, afectan a la formación de cuajado al momento de elaborar quesos, induciendo una maduración anormal, además ocurre la disminución de la producción de acidez durante el proceso de productos lácteos fermentados y la disminución del crecimiento de cultivos lácteos (Talero-Pérez *et al.*, 2014).

Por lo general, los antibióticos se clasifican según su mecanismo de acción en: inhibidores de la formación de la pared bacteriana, inhibidores de la síntesis proteica, inhibidores de la duplicación del ADN, inhibidores de las vías metabólicas e inhibidores de la membrana citoplasmática (Fernández, 2015).

### **1.3 Cromatografía líquida de ultra alto rendimiento**

La cromatografía líquida es una técnica analítica que se utiliza con el fin de separar componentes de una mezcla. Consta de una fase estacionaria (columna) y de una fase móvil (líquidos inmiscibles entre sí). La función de la fase móvil es arrastrar a la muestra a través de la fase estacionaria, la que puede ser sólida o compuesta por una película líquida soportada en un sólido inerte. Los componentes con mayor afinidad con la fase estacionaria se desplazarán con menor velocidad que aquellos que presentan menor afinidad. En años recientes, la cromatografía líquida de ultra alto rendimiento (UHPLC) se ha convertido en una de las técnicas de mayor uso para el análisis químico. Se caracteriza por ser un método con un tiempo de análisis más corto y un equilibrio de columna más rápido, por lo que resulta ideal para el desarrollo de nuevos métodos analíticos. La UHPLC permite el desarrollo de métodos más rápido y un análisis más preciso de muestras complejas debido a sus capacidades de alto rendimiento y alta resolución (Wang *et al.*, 2015).

Las características actuales del sistema (dispersión, volúmenes de permanencia y límites de presión) son bastante adecuadas para columnas empaquetadas con partículas de menos de 2 o 3  $\mu\text{m}$ , con un diámetro interno de 2 o 3 mm. Esta técnica de alta resolución ayuda al desarrollo rápido de métodos ya que aumenta su rendimiento entre 3 y 10 veces más que el método convencional. También se considera como una tecnología verde, debido a que utiliza menos solventes orgánicos y una pequeña cantidad de muestras, lo que se traduce en velocidades de flujo y un tiempo de análisis más bajo (las columnas son más pequeñas y cortas). Por otra parte, una de las más grandes limitaciones de este método es el costo del equipo, así como la compatibilidad con equipos existentes de HPLC. A pesar de su rápido crecimiento, en el mercado existe una selección limitada de fases estacionarias en comparación con la cromatografía gaseosa (GC). Esto se evidencia para la cromatografía de intercambio iónico y de exclusión por tamaño (Dong & Zhang, 2014).

#### **1.4 Modelado *in silico***

Actualmente las relaciones cuantitativas estructura-propiedad (QSPR), también conocidas como estudios *in silico*, se han convertido en una herramienta importante de amplia aplicación en la ciencia de los alimentos (dentro de lo que se conoce como química computacional de los alimentos). Esta metodología busca predecir las propiedades de las moléculas a partir de su estructura química representada por los descriptores moleculares. En la actualidad el desarrollo de modelos QSPR sigue en desarrollo y expansión, por lo han sido adoptados por diferentes organismos internacionales como una herramienta para explorar la información contenida en las estructuras químicas de las moléculas desconocidas. 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 (Rojas & Duchowicz, 2021).

El modelado QSPR busca desarrollar correlaciones matemáticas entre una respuesta, la cual está definida por la propiedad de los compuestos, y descriptores moleculares que se definen como atributos químicos codificados dentro de la estructura química de las moléculas analizadas. A lo largo del tiempo, ha existido cada vez un mayor interés entre los investigadores por utilizar este enfoque para modelar los fenómenos de retención en diversos tipos de cromatografía, ya que es útil para predecir el tiempo de retención ( $t_R$ ) o el índice de retención ( $I$ ) de compuestos no evaluados o no sintetizados. Asimismo, es útil para desarrollar modelos con fines regulatorios para la evaluación del peligro químicos de ciertas sustancias. También ayuda a la evaluación de la exposición y evaluación del peligro de los compuestos químicos, así como asistir en la identificación de sustancias persistentes, bioacumulables y tóxicas en el ambiente. Finalmente, esta metodología *in silico* es útil como alternativa a la experimentación con animales (*in vivo*) (Rojas & Duchowicz, 2021). Para la generación de modelos *in silico* basados en las relaciones cuantitativas estructura-propiedad se consideran las siguientes etapas:

### 1.4.1 Generación de la base de datos

La generación de la base de datos es el primer paso para el desarrollo de un modelo *in silico*. Para ello se recopila la información de los compuestos químicos junto con una respuesta de interés: antibióticos y tiempos de retención en este caso. Con la información inicial de la base de datos, se continúa con la inclusión de información adicional, por ejemplo, el número de registro CAS, la notación de la especificación de introducción lineal molecular simplificada (SMILES) y el número de identificación PubChem CID. Esta información está disponible en la quimioteca de acceso libre PubChem (Kim *et al.*, 2019) y sirve para filtrar la base de datos con el propósito de eliminar moléculas que presentan problemas para el desarrollo del modelo: Por ejemplo, estructuras que poseen deuterio, moléculas repetidas, así como compuestos con información faltante. Para el curado de los antibióticos se utilizará el programa alvaMolecule (Alvascience, 2022c; Mauri & Bertola, 2022), que servirá también identificar errores presentes en la estructura química de los antibióticos.

### 1.4.2 Modelos moleculares

Los compuestos químicos son representados mediante modelos moleculares, es decir, cualquier molécula puede ser modelada mediante un grafo molecular (representación gráfica). Los programas que se usan en química computacional poseen una interfaz gráfica que facilita el diseño de compuestos químicos (Rojas & Duchowicz, 2021). Un modelo químico debe ser aplicable para cualquier tipo de molécula, siendo el único limitante la capacidad del cálculo del sistema. Por lo tanto, el modelo molecular debe estar bien definido por una configuración dada de núcleos y electrones, permitiendo una solución aproximada por la ecuación de Schrödinger. Un modelo químico posee las siguientes características (Cuevas & Cortés, 2003):

- **Consistencia de tamaño:** debe generar el resultado que se obtendría de la solución de la ecuación de Schrödinger.
- **Variacional:** las energías no deben ser menores a las obtenidas de la solución de la ecuación de Schrödinger.
- **Eficiente:** debe ser de fácil implementación y de cálculo mediante herramientas computacionales.
- **Preciso:** los resultados experimentales deben ser reproducidos cuantitativamente.

#### 1.4.2.1 Modelos moleculares bidimensionales

Para una representación bidimensional de los compuestos, basta un método de optimización basado en la mecánica molecular MM+. Este método se fundamenta en las leyes de la mecánica clásica, es decir, las moléculas se tratan como un conjunto de átomos que se representan mediante esferas en el espacio unidos entre sí por enlaces representados por resortes (Jensen,

2017). A partir de las moléculas optimizadas con la mecánica molecular, se pueden obtener diversos grafos moleculares (Janezic *et al.*, 2015). Un grafo consta de un grupo de vértices (átomos) y un grupo de aristas entre los vértices (enlaces químicos). Gracias a la representación de los grafos moleculares se puede calcular diversos tipos de matrices, denominadas matrices grafo-teóricas. Estas matrices se caracterizan por permitir el cálculo de distintos tipos de descriptores moleculares. Las matrices grafo-teóricas más destacadas son (Janezic *et al.*, 2015; Todeschini & Consonni, 2009):

- **Matrices de vértices:** son matrices cuadradas  $|V| \times |V|$ , es decir tanto las filas como columnas representan a los átomos o vértices del grafo molecular.
- **Matrices de aristas:** son matrices cuadradas  $|E| \times |E|$ , cuyos elementos codifican la información concerniente a los enlaces. La matriz más importante es la de adyacencia A.
- **Matrices de incidencia:** son matrices no cuadradas, ya que no brinda información referente a las relaciones entre dos diferentes conjuntos de átomos, enlaces, ciclos, subestructuras o trayectos moleculares.

#### 1.4.2.2 Modelos moleculares tridimensionales

La representación tridimensional (geométrica) se caracteriza por brindar mayor información a la conectividad atómica. Permite distinguir estructuras moleculares que presentan isomería. El primer grupo de isómeros son los geométricos *cis-* o *trans-*, que se distinguen entre sí por los sustituyentes se encuentran del mismo lado o de lados opuestos de un plano de referencia común para ambos estereoisómeros, respectivamente. Otro grupo de isómeros tridimensionales son los enantiómeros: si la rotación es hacia la izquierda la molécula es levógira (*L-*); mientras que si la rotación es hacia el lado opuesto es dextrógira (*D-*). También existen los isómeros (*R-*) y (*S-*), estos isómeros se distinguen por la dirección en la que se encuentra la molécula, si se encuentra en sentido de las manecillas del reloj se utiliza *R* (rectus). Por otra parte, si la dirección es en contra de las manecillas del reloj se utiliza *S* (siniestro). En el caso particular del benceno existen tres tipos de isómeros, el primero es *orto-* (sustituyentes en las posiciones 1 y 2), *meta-* (sustituyentes de las posiciones 1 y 3) y *para-* (sustituyentes en las posiciones 1 y 4) (Morrison & Boyd, 1998).

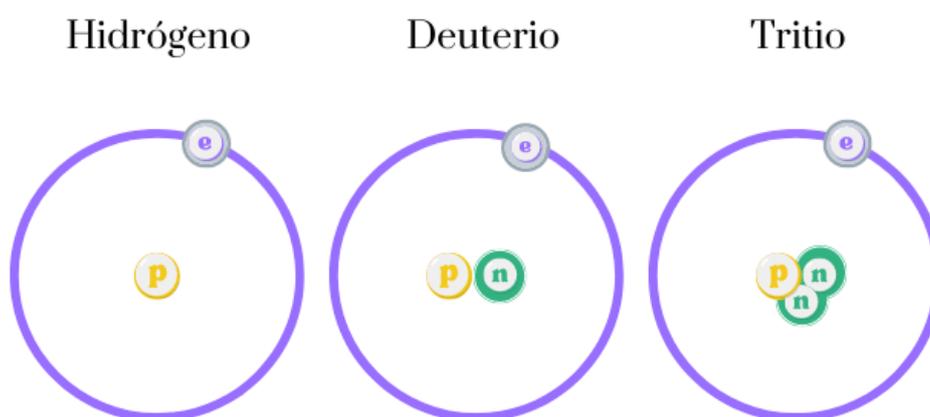
#### 1.4.2.3 Método semiempírico Tight Binding GFN2-Xtb

Los métodos semiempíricos permiten la optimización de geometrías moleculares (entre otras propiedades) con bajos costos computacionales. Entre los diversos métodos semiempíricos disponibles, se tiene al método GFN2-xTB (Bannwarth *et al.*, 2021; Bannwarth *et al.*, 2019), perteneciente a la familia de métodos Tight Binding o de Ligadura Fuerte. Este método permite la optimización molecular de geometrías, calcular frecuencias vibracionales y energías de interacción no covalente, además de otras propiedades. Adicionalmente, el método demuestra

un gran desempeño en dinámica molecular, simulaciones de espectros de masas, además de ser útil como entrada para generar campos de fuerza intermoleculares. También adquiere importancia para el cálculo automatizado de espectros de resonancia magnética nuclear (RMN), donde el muestreo estructural GFN-xTB se utiliza para generar el conjunto conformero-rotámero que a su vez define la forma detallada de un espectro de RMN. Las principales características de este método son:

- El conjunto base de GFN2-xTB consta de una base mínima de funciones de valencia centradas en los átomos. Se utilizan funciones de polarización con la finalidad de mejorar la descripción de estados hipervalentes.
- El hamiltoniano GFN2-xTB representa el primer método Tight Binding que incluye interacciones electrostáticas y efectos de correlación de intercambio hasta segundo orden en la expansión multipolar. Adicionalmente, el método incluye una descripción de interacciones de dispersión a nivel conocido como D4.
- A diferencia de otros métodos semiempíricos, el método GFN2-xTB sigue una estrategia de parametrización global y por elemento, en lugar de por pares de elementos como es usual.

En los métodos de optimización de geometrías semiempíricos, la función de onda trabaja sobre los electrones. Por este motivo, los compuestos que tienen hidrógeno pesado (deuterio o tritio) son invariantes a estos métodos. En la Figura 1.1 se muestra que la única diferencia que existe entre el hidrógeno, deuterio y tritio es el número de protones presentes en el núcleo: cero, uno y dos, respectivamente.



**FIGURA 1.1** Representación química del hidrógeno, deuterio y tritio.

### 1.4.3 Descripciones moleculares

Los descriptores moleculares son representaciones matemáticas de una molécula que se obtienen a través de la aplicación de un algoritmo específico a una representación molecular definida, o a un procedimiento experimental específico (Consonni & Todeschini, 2011; Mauri *et al.*, 2017). El campo de los descriptores moleculares es multidisciplinario y engloba diferentes

teorías en distintos niveles. 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. Por ende, se pueden distinguir dos 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). Cada descriptor molecular contiene una pequeña parte de la información química contenida en su estructura molecular y se clasifican en diferentes categorías: 0D o constitucionales, como el número de átomos de carbono o de nitrógeno en una molécula; 1D también conocidos como descriptores de conteo de grupos funciones o fragmentos centrados en el átomo. Los descriptores 2D se basan en una representación topológica, por ejemplo, índices topológicos, índices de conectividad, índices de información, entre otros. Finalmente, los descriptores moleculares tridimensionales (3D) consideran la distribución espacial de los átomos obtenidos mediante una optimización semiempírica o cuántica, aquí se pueden citar los descriptores 3D-MorSE, WHIM, CATS 3D y otros más.

#### **1.4.4 Desarrollo del modelo**

Para el desarrollo del modelo basado en las relaciones cuantitativas estructura-propiedad (QSPR), la base de datos de antibióticos, representados por los descriptores moleculares y el  $t_R$ , se divide en grupos de calibración y predicción. Con el grupo de calibración se aplica el aprendizaje supervisado para calibrar el modelo, mientras que el grupo de predicción se utiliza para medir la capacidad predictiva del modelo. El método más usado para modelar una variable aleatoria continua es la regresión de mínimos cuadrados ordinarios (OLS), también conocida como regresión lineal múltiple (MLR) (Rojas & Duchowicz, 2021). Este enfoque permite establecer relaciones lineales entre un grupo de variables explicativas (independientes) y una variable dependiente (respuesta cuantitativa) (Rencher & Schaalje, 2008). El método de los OLS brinda una estimación de los coeficientes de 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). Esta estimación coincide con el obtenido mediante el método de máxima verosimilitud. Una vez que se ha obtenido el modelo, se somete a validación interna para analizar la estabilidad del modelo a la exclusión de un cierto número de moléculas. Finalmente, el modelo se aplica para calcular los tiempos de retención de los antibióticos del grupo de predicción (capacidad predictiva del modelo). Se espera que los parámetros en calibración, validación interna y predicción sean los mejores y lo más cercano entre ellos.

#### **1.4.5 Interpretación del modelo**

La interpretación de un modelo QSPR se refiere en analizar la capacidad de ajuste de los datos, la estabilidad del modelo frente a perturbaciones internas (validación cruzada) y

fundamentalmente a la capacidad predictiva del mismo. Adicionalmente, se analizan datos que presentan valor de residuo alto (datos atípicos), es decir, la diferencia entre el tiempo de retención experimental y predicho por el modelo; así como analizar la presencia de patrones en la gráfica de residuos y tiempos de retención predichos. Otro aspecto importante es el análisis del mecanismo de acción de los descriptores presentes en el modelo, lo que contribuye a su aplicación para fines regulatorios. El mecanismo de acción se obtiene mediante la definición de cada descriptor molecular y la manera en la que se relaciona con la propiedad experimental ( $t_R$ ). Sin embargo, no siempre es posible obtener dicha interpretación debido a la complejidad matemática involucrada en la definición de algunos descriptores moleculares (Rojas & Duchowicz, 2021). Adicionalmente, en los modelos QSPR las predicciones dependen de la naturaleza química y la propiedad experimental de los compuestos utilizados para calibrar el modelo. El dominio de aplicabilidad (AD) (Jaworska *et al.*, 2005) se refiere a la construcción de un espacio químico teórico dentro del cual las predicciones de nuevos compuestos son confiables. En los modelos de regresión, el enfoque comúnmente utilizado es el valor de influencia, el cual se mide la distancia de cada objeto del conjunto de predicción con respecto al centro del modelo definido por el grupo de calibración (valor de influencia crítico) (Rojas & Duchowicz, 2021).

## Capítulo II

### 2 Materiales y métodos

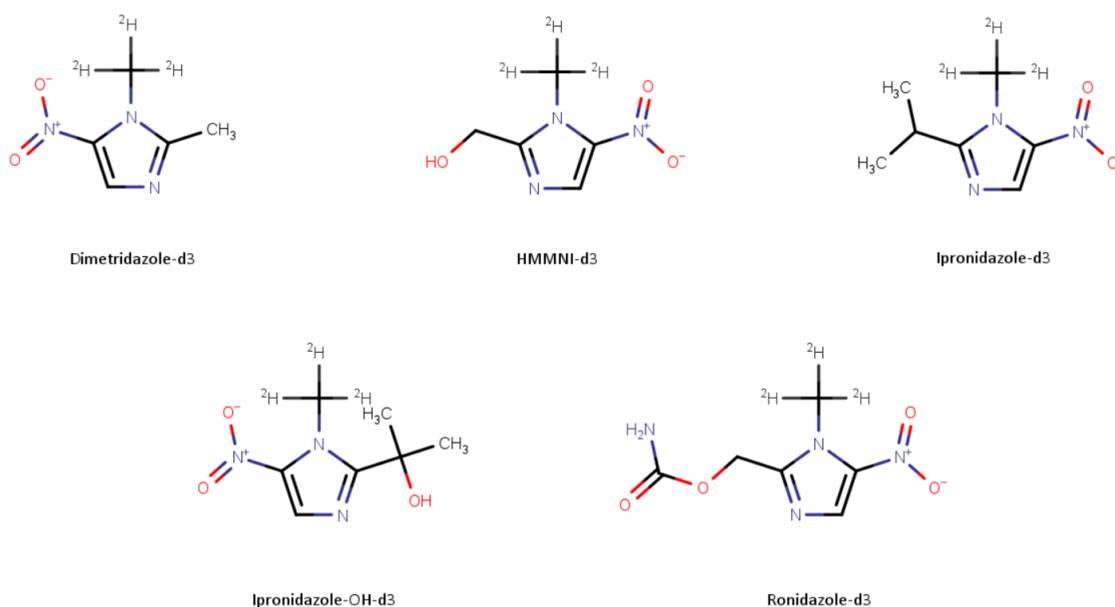
#### 2.1 Generación de la base de datos

Para generar la base de datos se recopiló la información experimental del artículo científico “*Development and Validation of a Multiclass Method for Analysis of Veterinary Drug Residues in Milk Using Ultrahigh Performance Liquid Chromatography Electrospray Ionization Quadrupole Orbitrap Mass Spectrometry*” (Wang *et al.*, 2015). En esta investigación se cuantificaron los  $t_R$  de 130 antibióticos identificados en muestras de leche mediante la cromatografía líquida de ultra alto rendimiento/espectrometría de masa de ionización por electrospray cuadrupolo Orbitrap (UHPLC/ESI Q-Orbitrap). Los antibióticos se aislaron mediante un método modificado de extracción con líquido asistido por salinización (SOSLE), que implica la precipitación de las proteínas de la leche con un tampón de extracción (ácido oxálico y ácido etilendiaminotetraacético EDTA a pH = 3) y acetonitrilo. A continuación, se separan las fases de acetonitrilo/agua por salificación usando sulfato de amonio, y se procede a la extracción en fase sólida (SPE) utilizando cartuchos adsorbentes poliméricos de fase inversa. Para el análisis químico las muestras fueron inyectadas de manera individual en la columna Hypersil Gold (100 mm x 2.1 mm, 1.9  $\mu$ m) y el Q-Orbitrap se operó en Full MS/dd-MS2 modo positivo para determinar el tiempo de retención y la identidad de cada antibiótico. Las fases móviles fueron metanol y agua, cada una compuesta con formiato de amonio y ácido fórmico al 0.10%. Para la identificación de los analitos se usaron soluciones estándar de diversas drogas veterinarias. Posteriormente, para cada antibiótico se obtuvo el número de registro CAS, la notación de la especificación de introducción lineal molecular simplificada (SMILES) y el número de identificación PubChem CID a partir de la quimioteca de acceso libre PubChem (Kim *et al.*, 2019).

#### 2.2 Filtrado de la base de datos

En el filtrado de la base de datos se excluyeron los antibióticos que tienen deuterio dentro de su estructura química, debido a que el método de optimización de geometría semiempírico Tight Binding GFN2-Xtb trabaja sobre los electrones. Los compuestos excluidos de la base de datos son (Figura 2.1):

- *Dimetridazol-d3* (PubChem CID 16212238 y CAS 64678-69-9).
- *HMMNI-d3* (*Hidroxí Dimetridazol-d3*) (PubChem CID 16212240 y CAS 1015855-78-3).
- *Ipronidazol-d3* (PubChem CID 16212242 y CAS 1015855-83-0).
- *Ipronidazol-OH-d3* (PubChem CID 71749080 y CAS 1156508-86-9).
- *Ronidazol-d3* (PubChem CID 16212243 y CAS 1015855-87-4).



**FIGURA 2.1** Antibióticos que tienen deuterio dentro de la estructura química.

### 2.3 Diseño y optimización de las geometrías moleculares

Las estructuras moleculares de los antibióticos fueron diseñadas en el programa HyperChem (Hypercube Inc.), donde se optimizaron las geometrías de acuerdo al campo de fuerzas de la mecánica molecular MM+. Una vez diseñadas y optimizadas las geometrías, se utilizó el programa alvaMolecule (Alvascience, 2020)(Mauri & Bertola, 2022) para el curado de las estructuras moleculares. Durante el curado se identifican errores presentes en los compuestos, por ejemplo: valencia inusual, átomos no estándar y aromaticidad. El programa permite utilizar diversas herramientas de estandarización para corregir diseños erróneos de las estructuras moleculares, para eliminar características específicas de las moléculas o para estandarizar aspectos estructurales definidos: 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, eliminar centros quirales y orientación de los enlaces.

Posteriormente, se afinaron las geometrías mediante una optimización basada en el método semiempírico Tight Binding GFN2-xTB implementado en el programa xTB (Bannwarth *et al.*, 2021; Bannwarth *et al.*, 2019). Se utilizó la metadinámica para muestrear el espacio conformacional molecular y optimizar la geometría del conformero de energía más baja encontrado para cada antibiótico (Grimme, 2019). Los efectos del solvente (agua) se tuvieron en cuenta implícitamente a través del método analítico linealizado de Poisson-Boltzmann, también implementado en el programa xTB (Ehlert *et al.*, 2021). Además, las energías de ionización vertical se calcularon utilizando una opción ad-hoc incluida en el mismo programa.

## 2.4 Cálculo de descriptores moleculares

Las estructuras químicas optimizadas de los antibióticos se usaron para calcular 5633 descriptores moleculares en el programa alvaDesc (Alvascience, 2022a; Mauri & Bertola, 2022). Este programa calcula descriptores independientes (0D a 2D) y dependientes (3D) de la conformación molecular de acuerdo a la optimización GFN2-xTB. Los descriptores calculados en alvaDesc se encuentran agrupados en 32 bloques lógicos. A continuación, se realizó una reducción de la dimensionalidad de los descriptores mediante la eliminación de aquellos con valores constantes, casi constantes y con valores faltantes. Complementariamente, en el programa xTB se calcularon seis descriptores cuánticos: momento dipolar (MD), diferencia de energía HOMO-LUMO (H-L gap en Debye), energía libre de Gibbs de solvatación (Gsolv en kcal/mol), energía de ionización (IP en eV), afinidad electrónica (EA en eV) y electronegatividad (EN en eV).

## 2.5 Reducción no supervisada de descriptores moleculares

Los métodos de reducción no supervisados consideran únicamente la matriz de descriptores para reducir la dimensionalidad de los datos de tal manera que preserve la información contenida en la matriz de datos inicial, es decir, se emplean para excluir aquellos descriptores que presentan ruido, redundancia y multicolinealidad. Entre los diversos enfoques disponibles para la reducción no supervisada de descriptores, en este trabajo se ha usado el método V-WSP (algoritmo propuesto por Wootton, Sergent y Phan-Tan-Luu), el cual selecciona un subconjunto representativo de variables que tengan mínima correlación en el espacio multidimensional dentro de un umbral determinado (por ejemplo 0.95) (Ballabio *et al.*, 2014). Esta reducción consiste en comparar el coeficiente absoluto de correlación de Pearson entre pares de descriptores ( $R_{ij}$ ) y el valor de umbral preestablecido ( $thr = 0.95$ ); de tal forma que si se cumple que  $R_{ij} \geq thr$ , se eliminará el descriptor que presenta la mayor correlación promedio con los demás. De esta manera, se seleccionan las variables más representativas para desarrollar el modelo computacional.

## 2.6 Desarrollo del modelo *in silico*

Con la finalidad de establecer la relación cuantitativa estructura-propiedad se utilizó la regresión de mínimos cuadrados ordinarios (OLS) acoplado a los algoritmos genéticos (GAs) (Kubat, 2017; Leardi, 2020) para la selección supervisada de descriptores. Los GAs se basan en la teoría de la evolución de Darwin, es decir, se parte de una población de “*cromosomas*”, a partir de la cual se genera un proceso evolutivo mediante reproducción y mutación con la finalidad de optimizar una función objetivo (por ejemplo, el error cuadrático medio RMSE). Aquí, cada cromosoma es un vector binario de  $p$  bits (número de variables), en donde bits con valor 1 indican que los descriptores correspondientes están presentes en el modelo y viceversa para los bits nulo, durante la corrida, el tamaño de la población es constante, manteniendo a los mejores

cromosomas mientras que los peores son eliminados (Rojas & Duchowicz, 2021). El desarrollo del modelo mediante mínimos cuadrados acoplados a los algoritmos genéticos se desarrollará en el programa alvaModel (Alvascience, 2022b; Mauri & Bertola, 2022).

## 2.7 División de la base de datos

Con el fin de asegurar la confiabilidad del modelo QSPR para los  $t_R$  de los antibióticos, la base de datos se divide en conjuntos de calibración y predicción de acuerdo al método de subconjuntos balanceados (BSM) (Rojas & Duchowicz, 2021). Este método se fundamenta en el análisis de conglomerados  $k$ -medias ( $k$ -MCA) 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. El grupo de calibración se usa para calibrar el modelo, mientras que el conjunto de predicción se usa para medir la capacidad predictiva del mismo. La división de una base de datos debe ser realizada de tal manera que se alcance relaciones estructura-propiedad similares. Por esta razón, la partición BSM considera la propiedad experimental y los descriptores independientes de la conformación, con el propósito de evitar sesgo debido a los descriptores conformacionales.

## 2.8 Validación del modelo

El modelo se valida mediante protocolos de validación interna de dejar-uno-fuera (LOO) y dejar-varios-fuera (LMO). 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 y se usan las restantes para recalibrar el modelo; seguidamente se predice la propiedad de las moléculas excluidas. Los métodos LMO utilizados en este trabajo son: Monte Carlo, Bootstrap y Ventanas venecianas. En el enfoque Monte Carlo se define subconjuntos de entrenamiento (80% de las moléculas) y de evaluación (20% de los compuestos), este proceso se repite tantas veces como sea necesario (iteraciones). La calidad de la validación se efectúa por comparación de las predicciones acumuladas versus las respuestas del subconjunto de evaluación (Krakowska *et al.*, 2016; Xu *et al.*, 2004). Por otra parte, en el método Bootstrap (Efron, 1979, 1982, 1987) se utiliza el muestro aleatorio con reemplazo, en donde se conserva constante el tamaño original de la base de datos ( $n$ ), es decir, los objetos del subconjunto de entrenamiento estarán conformados por objetos repetidos con el propósito de equilibrar con número de elementos excluidos. El modelo se calcula con los compuestos del conjunto de calibración para predecir las respuestas de los elementos del conjunto de validación (Rojas & Duchowicz, 2021). Por otra parte, las ventanas venecianas realizan una división del grupo de calibración de acuerdo a  $k$ -grupos de validación cruzada, es decir, cada objeto del grupo de validación se selecciona a partir del primer objeto del grupo de

calibración y los subsiguientes cada  $k$ -ésimo objeto; los cuales se excluyen uno a la vez, se recalibra el modelo y posteriormente se realiza la predicción de las moléculas excluidas.

Finalmente, se aplica la aleatorización-Y (Lindgren *et al.*, 1996) para descartar la presencia de correlación casual en el modelo. Es decir, los tiempos de retención del vector respuesta se cambian de forma aleatoria de tal forma que no correspondan a sus correspondientes antibióticos. Este proceso se repite cientos de veces a fin de obtener una estimación promedio del modelo aleatorizado (Rojas & Duchowicz, 2021). Se espera que los resultados de la aleatorización-Y sean pobres con respecto a los de calibración. La validación del modelo QSPR también se realizó en el programa alvaModel (Alvascience, 2022c; Mauri & Bertola, 2022).

## **2.9 Grado de contribución de los descriptores moleculares**

Para medir el grado de contribución de los descriptores presentes en el modelo se estandarizan los coeficientes del modelo de regresión lineal. De esta manera, mientras más alto sea el valor absoluto del coeficiente de un descriptor, mayor será la importancia del mismo para predecir el tiempo de retención y viceversa.

## **2.10 Mecanismo de acción**

El mecanismo de acción se obtiene mediante la definición de cada descriptor molecular (si es posible) y la manera en que se relaciona con la propiedad en estudio ( $t_R$ ). Sin embargo, no siempre es posible obtener dicha interpretación debido a la complejidad matemática involucrada en la definición de algunos descriptores moleculares. El mecanismo de acción ayuda a brindar mayor confianza al modelo desarrollado, principalmente para fines regulatorios (Rojas & Duchowicz, 2021).

## **2.11 Análisis del dominio de aplicabilidad**

Los modelos QSPR son modelos reduccionistas o locales, es decir, las predicciones están acotadas por la naturaleza química y la propiedad de los compuestos utilizados para calibrar el modelo (principio de congenericidad). Por lo tanto, el dominio de aplicabilidad (AD) (Jaworska *et al.*, 2005) se refiere a un espacio químico teórico dentro del cual las predicciones del tiempo de retención de nuevos compuestos son confiables. Si una molécula cae fuera de este dominio, su predicción se considera una extrapolación del modelo, y por lo tanto poco confiable. Para definir el dominio de aplicabilidad se usa el enfoque basado en el valor de influencia (valor umbral de influencia también conocido como umbral crítico). Si el valor de influencia es grande, la observación determina su respuesta sin considerar el resto de las observaciones. Por el contrario, si los valores son pequeños esto indica que las observaciones contribuyen equitativamente en la construcción de las respuestas (Rojas & Duchowicz, 2021). Por lo tanto,

moléculas que poseen valores altos de influencia se encuentran distantes del centro del modelo y se las considera predicciones poco confiables o extrapolaciones del modelo.

## Capítulo III

### 3 Resultados y discusiones

#### 3.1 Generación de la base de datos y curado

Con los 130 tiempo de retención de antibióticos identificados en muestras de leche cruda datos reportados por (Wang *et al.*, 2015). Se realizó una verificación del nombre químico y se obtuvo el número de registro CAS y SMILES canónico en la quimioteca de acceso libre PubChem (Kim *et al.*, 2021). Posteriormente, se excluyeron cinco antibióticos que poseen hidrógeno pesado dentro de su estructura química. Seguidamente, con el propósito de obtener una base de datos estandarizada se procedió a analizar cada molécula con el software alvaMolecule (Alvascience, 2022c; Mauri & Bertola, 2022), donde se implementaron diversos de curado.

#### 3.2 Cálculo de descriptores moleculares

Se calcularon 5633 descriptores moleculares en el programa alvaDesc (Alvascience, 2022a; Mauri & Bertola, 2022). En una primera etapa de reducción no supervisada se excluyeron 1792 descriptores con valores constantes, 99 con valores casi constantes y 31 descriptores con al menos un valor faltante. De esta manera se obtuvieron 3711 descriptores moleculares que fueron sometidos a la reducción no supervisada V-WSP con un umbral de correlación  $thr = 0.95$ . Este procedimiento permitió excluir 1544 descriptores moleculares. A esta base de datos reducida de descriptores se fusionaron los seis descriptores calculados en el programa xTB.

#### 3.3 Desarrollo y validación del modelo *in silico*

Para el desarrollo del modelo, la base de datos fue dividida en grupos de calibración y predicción de acuerdo al enfoque del método de subconjuntos balanceados. Se usaron 88 moléculas para calibrar el modelo y las 37 restantes para estimar la capacidad predictiva del modelo. Para la selección supervisada de los 2173 descriptores mediante el los algoritmos genéticos acoplados al método de mínimos cuadrados ordinarios se utilizó el grupo de calibración para ajustar los modelos en el programa alvaModel (Alvascience, 2022b; Mauri & Bertola, 2022). Se analizó la validación cruzada de ventanas venecianas con cinco grupos de exclusión para controlar la presencia de sobreajuste en el modelo. Este procedimiento permitió obtener un modelo óptimo de tres descriptores moleculares:

$$t_R = 0.481 + 0.100 MLOGP2 - 1.201 ESOL + 1.221 CATS2D\_07\_NL \quad (\text{Ec. 1})$$

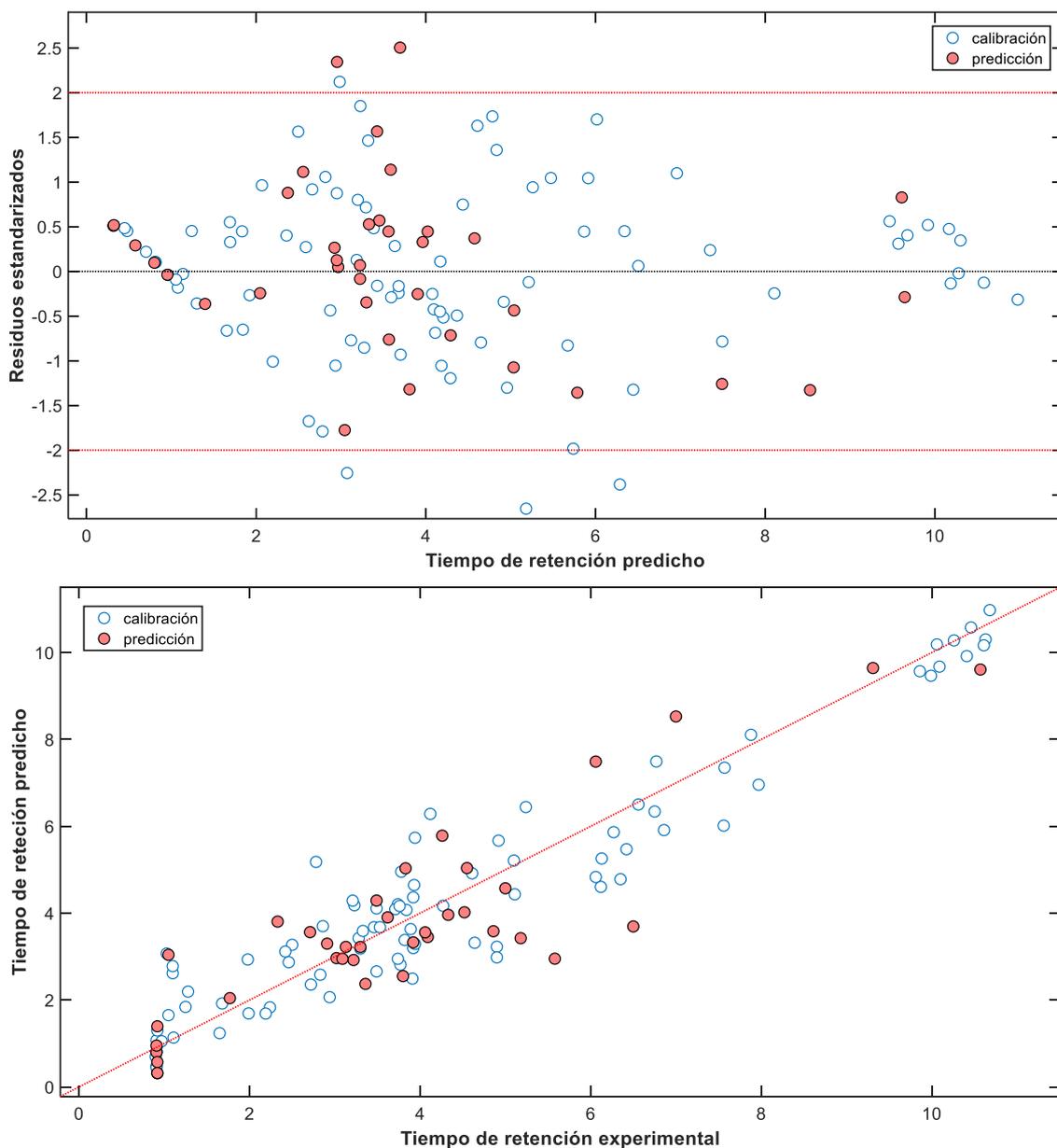
Los tres descriptores que forman parte de la relación cuantitativa estructura-propiedad para la predicción de los tiempos de retención de antibióticos identificados en muestras de leche pertenecen a las propiedades moleculares (*MLOGP2* y *ESOL*) y descriptores farmacóforos (*CATS2D\_07\_NL*). *MLOGP2* es el coeficiente de reparto octanol-agua de Moriguchi cuadrático, *ESOL* es la solubilidad estimada para la solubilidad acuosa utilizando un modelo de consenso para el LogP (coeficiente de reparto octanol-agua) y *CATS2D\_07\_NL* indica la presencia de pares de átomos negativo (N) y lipofílico (L) separados por una distancia topológica de 7 enlaces.

El modelo QSPR de la Ec. 1 tiene una capacidad descriptiva de los tiempos de retención del 90.6 %. Adicionalmente, el modelo fue sometido a varios enfoques de validación cruzada: dejar-uno-fuera, ventanas venecianas con 5 grupos, Monte Carlo (20% de exclusión con 1000 iteraciones) y Bootstrap con 1000 iteraciones. La similitud entre el coeficiente de determinación y el error cuadrático de validación y calibración indica la ausencia de sobreajuste (Tabla 3.1). Adicionalmente, se ha analizado la correlación casual en el modelo mediante la aleatorización-Y del tiempo de retención con 1000 iteraciones, obteniendo valores promedio de  $R^2 = 0.034$ , RMSE = 2.766, lo que claramente indica que al cambiar de forma aleatoria la respuesta, se obtienen modelos con poca calidad descriptiva (verifica la ausencia de correlación casual). Finalmente, al aplicar el modelo para la estimación de los tiempos de retención de las moléculas del grupo de predicción, se observa que el mismo tiene una capacidad de predicción del 85.7 % (Tabla 1), y en consecuencia, es válido para la predicción de los  $t_R$  de antibióticos detectados mediante UHPLC.

**TABLA 1** Resultados del modelo *in silico* en calibración, validación cruzada y predicción.

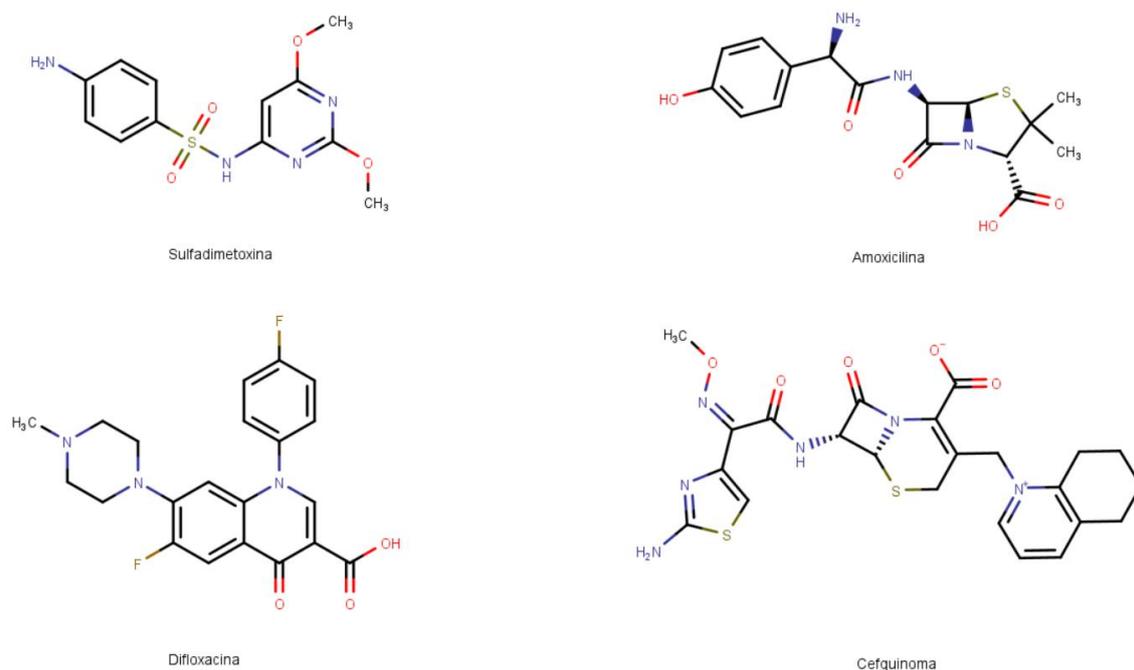
	<b>R<sup>2</sup></b>	<b>RMSE</b>
<b>Calibración</b>	0.906	0.864
<b>Validación cruzada</b>		
Dejar-uno-fuera	0.899	0.893
Ventanas venecianas	0.897	0.906
Bootstrap	0.899	0.891
Monte Carlo	0.900	0.891
<b>Predicción</b>	0.857	1.066

En la Figura 3.1a, se observa la relación entre el  $t_R$  experimental y predicho por el modelo (Ec. 1). Se observa que existe un buen ajuste de los puntos experimentales del grupo de calibración y predicción alrededor de la línea de ajuste perfecto. Lo que confirma los resultados presentados en la Tabla 1 para la calibración y predicción del modelo.



**FIGURA 3.1** Modelo in silico para los antibióticos detectados mediante UHPLC. (a) tiempos de retención experimental versus tiempos de retención predichos. (b) dispersión de los residuos estandarizados vs. los tiempos de retención predichos.

Por otra parte, en la Figura 3.1b 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  $t_R$  de los antibióticos determinados mediante UHPLC. Al definir un intervalo de dos veces el error cuadrático medio de calibración ( $\pm 2$  RMSC) se identifican 4 antibióticos que presentan residuo alto. Estos compuestos son *Sulfadimetoxina* (PubChem CID 5323 y CAS 122-11-2), *Amoxicilina* (PubChem CID 33613 y CAS 26787-78-0), *Difloxacina* (PubChem CID 56206 y CAS 98106-17-3), *Cefquinoma* (PubChem CID 5464355 y CAS 84957-30-2). La estructura química de estos antibióticos se muestra en la Figura 3.2.



**FIGURA 3.2** Antibióticos que presentan diferencia alta entre el tiempo de retención experimental y predicho por el modelo.

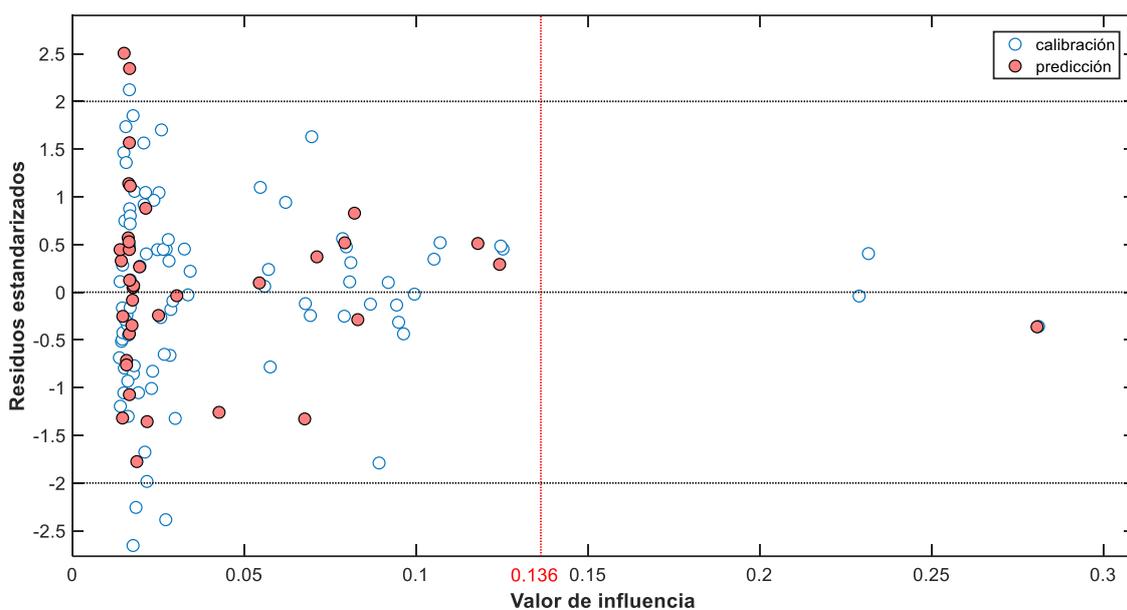
### 3.4 Mecanismo de acción

Mediante el cálculo de los coeficientes de regresión estandarizados se determinó el grado de contribución de cada descriptor para predecir el  $t_R$ :  $ESOL$  (-0.949) >  $MLOGP2$  (0.273) >  $CATS2D\_07\_NL$  (0.180). El descriptor  $ESOL$  es la solubilidad estimada (LogS) para la solubilidad acuosa utilizando un modelo de consenso para el LogP (Sorkun *et al.*, 2019). Este descriptor se fundamenta en una estimación que depende de la masa molecular (MW), número de enlaces rotables (RBN), la proporción aromática (AP) y el modelo de consenso para el LogP implementado en el programa alvaDesc (Alvascience, 2022a).  $ESOL$  tiene un efecto antagónico (negativo) para la predicción del  $t_R$ , es decir, que el tiempo de retención disminuye a medida que la solubilidad estimada para la solubilidad acuosa utilizando el modelo de consenso para el LogP aumenta. Por otra parte, el descriptor  $MLOGP2$  se basa en el coeficiente de reparto octanol-agua de Moriguchi lineal (MLOGP), el cual se construye mediante un modelo matemático que considera 13 parámetros estructurales. Para la calibración se utilizan 3576 moléculas orgánicas que incluyen compuestos alifáticos, aromáticos y heterocíclicos que contienen diversos tipos de átomos (C, H, N, O, S, P, F, Cl, Br, I) (Todeschini & Consonni, 2009). Este descriptor presenta un comportamiento sinérgico (positivo) para la predicción de los tiempos de retención. Finalmente,  $CATS2D\_07\_NL$  es un descriptor que cuantifica la presencia de pares de átomos negativos (N) y lipofílicos (L) en el esqueleto de los antibióticos separados por una distancia topológica de 7 enlaces. De esta forma, el  $t_R$  de los antibióticos aumenta cuando existen cada 7 enlaces, átomos de carbono, azufre o fósforo de grupos COOH, SOOH o POOH,

respectivamente (N), y átomos de cloro, bromo, yodo o un átomo de azufre adyacente únicamente a dos átomos de carbono (L) (Fechner *et al.*, 2003).

### 3.5 Dominio de aplicabilidad

El dominio de aplicabilidad define el espacio químico teórico dentro del cual el conjunto de calibración del modelo ha sido desarrollado y es aplicable con el fin de poder realizar la predicción de nuevos compuestos. El valor de influencia estableció un valor umbral de 0.136 (Figura 3.3), lo que significa que cualquier antibiótico por debajo del mismo se encuentra dentro del dominio de aplicabilidad. De esta manera las predicciones son el resultado de la interpolación del modelo. En este modelo se evidencia que únicamente una molécula supera el valor de influencia crítico y por consiguiente se encuentran fuera del dominio (predicción poco confiable).



**FIGURA 3.3** Diagrama de William para definir el dominio de aplicabilidad del modelo in silico: valor de influencia vs. residuos estandarizados.

## CONCLUSIONES

En este estudio se desarrolló un modelo *in silico* basado en una relación cuantitativa estructura-propiedad para los tiempos de retención de antibióticos detectados mediante cromatografía líquida de ultra alto rendimiento/espectrometría de masa de ionización por electrospray cuadrupolo Orbitrap (UHPLC/ESI Q-Orbitrap). La optimización de los compuestos de acuerdo a un método semiempírico moderno permitió calcular diversos tipos de descriptores moleculares que se sometieron a una reducción no supervisada para aligerar el cálculo computacional involucrado en la selección supervisada. La aplicación de métodos del aprendizaje automático permitió seleccionar tres descriptores moleculares que brindan una capacidad de ajuste del 90.6 % y una predictividad del 85.7 %. Este modelo *in silico* podría ser utilizado como un material de apoyo para los investigadores en el área de la química de los alimentos, de manera más específica para quienes trabajan en la identificación de contaminantes en diversas matrices alimentarias.

## REFERENCIAS

- Alvascience. (2022a). alvaDesc (Software for Molecular Descriptors Calculation) version 2.0.14, <https://www.alvascience.com>.
- Alvascience. (2022b). alvaModel (software to model QSAR data) version 2.0.6, <https://www.alvascience.com>.
- Alvascience. (2022c). alvaMolecule (Software to View and Prepare Chemical Datasets) version 2.0.2, <https://www.alvascience.com>.
- Badui, S. (2016). *Química de los Alimentos* (5a ed.). México, D.F. (México): Pearson Educación.
- 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, 147-154.
- Bannwarth, C., Ehlert, S., & Grimme, S. (2019). GFN2-xTB—An accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *Journal of Chemical Theory and Computation*, 15(3), 1652-1671.
- Bannwarth, C., Caldeweyher, E., Ehlert, S., Hansen, A., Pracht, P., Seibert, J., Spicher, S., & Grimme, S. (2021). Extended tight-binding quantum chemistry methods. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 11(2), e1493.
- Consonni, V., & Todeschini, R. (2011). Structure–Activity Relationships by Autocorrelation Descriptors and Genetic Algorithms. In *Chemoinformatics and Advanced Machine Learning Perspectives: Complex Computational Methods and Collaborative Techniques*, (pp. 60-94). Hershey (USA): IGI Global.
- Cuevas, G., & Cortés, F. (2003). *Introducción a la Química Computacional*: Fondo de Cultura Económica.
- Damodaran, S., & Parkin, K. L. (2018). *Fennema Química de los Alimentos*. Zaragoza (España): Editorial Acribia S.A.
- Dong, M. W., & Zhang, K. (2014). Ultra-high-pressure Liquid Chromatography (UHPLC) in Method Development. *TrAC Trends in Analytical Chemistry*, 63, 21-30.
- Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7, 1-26.
- Efron, B. (1982). *The jackknife, the Bootstrap and Other Resampling Plans*: SIAM.
- Efron, B. (1987). Better Bootstrap Confidence Intervals. *Journal of the American statistical Association*, 82(397), 171-185.
- Ehlert, S., Stahn, M., Spicher, S., & Grimme, S. (2021). Robust and Efficient Implicit Solvation Model for Fast Semiempirical Methods. *Journal of Chemical Theory and Computation*, 17(7), 4250-4261.
- Fechner, U., Franke, L., Renner, S., Schneider, P., & Schneider, G. (2003). Comparison of Correlation Vector Methods for Ligand-based Similarity Searching. *Journal of Computer-Aided Molecular Design*, 17(10), 687-698.

- Fernández, P., Moreno, A., Leza, J., Lizasoain, I., Moro, M., & Portolés, A. V. (2008). *Farmacología básica y clínica*. Madrid: Médica Panamericana.
- Fernández, P. L. (2015). *Velázquez Farmacología Básica y Clínica*: Editorial Médica Panamericana.
- Grimme, S. (2019). Exploration of Chemical Compound, Conformer, and Reaction Space with Meta-dynamics Simulations based on Tight-binding Quantum Chemical Calculations. *Journal of Chemical Theory and Computation*, 15(5), 2847-2862. <https://doi.org/10.1021/acs.jctc.9b00143>
- Hypercube Inc. HyperChem™ Professional version 8.0. <http://www.hyper.com>.
- Janezic, D., Milicevic, A., Nikolic, S., & Trinajstic, N. (2015). *Graph-theoretical Matrices in Chemistry*. CRC Press.
- Jaworska, J., Nikolova-Jeliazkova, N., & Aldenberg, T. (2005). QSAR Applicability Domain Estimation by Projection of the Training Set Descriptor Space: A Review. *ATLA*, 33(5), 445-459.
- Jensen, F. (2017). *Introduction to Computational Chemistry* (Third ed.). Chichester (UK): Wiley.
- Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., & Yu, B. (2019). PubChem 2019 Update: Improved Access to Chemical Data. *Nucleic acids research*, 47(D1), D1102-D1109.
- Krakowska, B., Custers, D., Deconinck, E., & Daszykowski, M. (2016). The Monte Carlo validation framework for the discriminant partial least squares model extended with variable selection methods applied to authenticity studies of Viagra® based on chromatographic impurity profiles. *Analyst*, 141(3), 1060-1070.
- Kubat, M. (2017). *An Introduction to Machine Learning* (Vol. 2). Cham (Switzerland): Springer.
- Leardi, R. (2020). Genetic Algorithms in Chemistry. In S. Brown, R. Tauler & B. Walczak (Eds.), *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis* Second ed., vol. 1 (pp. 617–634). Amsterdam (Netherlands): Elsevier.
- Lindgren, F., Hansen, B., Karcher, W., Sjöström, M., & Eriksson, L. (1996). Model Validation by Permutation Ttests: Applications to Variable Selection. *Journal of chemometrics*, 10(5-6), 521-532.
- Máttar, S., Calderón, A., Sotelo, D., Sierra, M., & Tordecilla, G. (2009). Detección de Antibióticos en Leches: Un Problema de Salud Pública. *Revista de Salud Pública*, 11, 579-590.
- Mauri, A., Consonni, V., & Todeschini, R. (2017). Molecular Descriptors. In *Handbook of computational chemistry*, (pp. 2065-2093). Gewerbestrasse (Switzerland): Springer.
- Mauri, A., & Bertola, M. (2022). Alvascience: A New Software Suite for the QSAR Workflow Applied to the Blood–Brain Barrier Permeability. *International Journal of Molecular Sciences*, 23(21), 12882.
- Moncada Jiménez, A., & Pelayo Consuegra, B. H. (2011). Análisis químico, biológico y fisicoquímico de la leche: calidad y contenido nutricional. In CANILEC (Ed.), *El Libro Blanco de la Leche y los Productos Lácteos*, (pp. 66-79). México D.F. (México).

- Morrison, R. T., & Boyd, R. N. (1998). *Química Orgánica* (Quinta ed.). México D.F. (México): Pearson.
- Rencher, A. C., & Schaalje, G. B. (2008). *Linear Models in Statistics*. Hoboken (USA): John Wiley & Sons, Inc.
- Rojas, C., & Duchowicz, P. R. (2021). *Química Computacional de los Alimentos: Relaciones Cuantitativas Estructura-Actividad/Propiedad (QSAR/QSPR)*. Zaragoza (España): Editorial Acribia, S.A.
- Salas, P., Calle, S., Falcón, N., Pinto, C., & Espinoza, J. (2013). Determinación de Residuos de Antibióticos Betalactámicos Mediante un Ensayo Inmunoenzimático en Leche de Vacas Tratadas Contra Mastitis. *Revista de investigaciones Veterinarias del Perú*, 24(2), 252-254.
- Sorkun, M. C., Khetan, A., & Er, S. (2019). AqSolDB, A Curated Reference Set of Aqueous Solubility and 2D Descriptors for a Diverse Set of Compounds. *Scientific data*, 6, 143.
- Talero-Pérez, Y. V., Medina, O. J., & Rozo-Núñez, W. (2014). Técnicas Analíticas Contemporáneas para la Identificación de Residuos de Sulfonamidas, Quinolonas y Cloranfenicol. *Universitas Scientiarum*, 19(1), 11-28.
- Todeschini, R., & Consonni, V. (2009). *Molecular Descriptors for Chemoinformatics* (Second ed.). Weinheim (Germany): Wiley-VCH Verlag GmbH & Co.
- Wang, J., Leung, D., Chow, W., Chang, J., & Wong, J. W. (2015). Development and Validation of A Multiclass Method for Analysis of Veterinary Drug Residues in Milk Using Ultrahigh Performance Liquid Chromatography Electrospray Ionization Quadrupole Orbitrap Mass Spectrometry. *Journal of Agricultural and Food Chemistry*, 63(41), 9175-9187.
- Xu, Q. S., Liang, Y. Z., & Du, Y. P. (2004). Monte Carlo Cross-validation for Selecting a Model and Estimating the Prediction error in Multivariate Calibration. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 18(2), 112-120.