



**UNIVERSIDAD  
DEL AZUAY**

**Facultad de Ciencia y Tecnología**

Escuela de Ingeniería en Alimentos

**Modelado *in silico* de los índices de retención de  
compuestos orgánicos volátiles identificados en  
muestras de quinua**

Trabajo de graduación previo a la obtención del título de:

**INGENIERO EN ALIMENTOS**

**Autor:**

Juan Fernando Cedillo Rodríguez

**Director:**

Dr. Cristian Rojas Villa

Cuenca – Ecuador

2024

## DEDICATORIA

*Lámpara y Luz: El Viaje de Psique*

*Con lámpara en mano, Psique se acerca,*

*¿Quién es este amor que la noche alberga?*

*"No busques ver, sino sentir", dice Eros,*

*"Pues la verdad yace más allá de los velos."*

Adaptado del mito de Eros y Psique

A mis amados padres, Fernando y Lorena, quienes, con su amor incondicional, sacrificios y apoyo constante, han sido el pilar fundamental para alcanzar esta meta. Gracias por sembrar en mí el valor de la perseverancia y la humildad. Este logro es por y para ustedes.

A mi hermano menor, Juan Francisco, mi mejor amigo y cómplice en las aventuras. Gracias por ser mi motivación para ser un ejemplo a seguir y por llenarme de risas en los momentos más difíciles. Tu admiración me di el impulso necesario para dar lo mejor de mí.

A mis queridos abuelos, Hugo y María, quienes, con sus sabias palabras y cálidos abrazos, me recordaron siempre mantener la fe y mantenerme firme ante las adversidades del camino. Su legado de trabajo duro y bondad quedará plasmado en mi corazón.

A mis tíos, Guido y Lorena, y primos, Diego, Lorena y Mateo, por su cariño y apoyo incondicional, me han demostrado el verdadero valor de la familia. Gracias por creer en mí y celebrar junto a mí cada uno de mis logros.

A mis amigos, compañeros y confidentes, quienes hicieron de esta aventura una experiencia inolvidable. Gracias por su comprensión, risas compartidas y por ser mi segunda familia. Juntos hemos creado recuerdos que los tendré presentes para siempre.

Este logro es el resultado del amor, sacrificio y apoyo de todos ustedes. Cada uno ha dejado una huella imborrable en mi corazón y en mi camino hacia el éxito.

Eternamente agradecido.

## AGRADECIMIENTOS

Este camino académico ha sido un viaje lleno de aprendizajes, desafíos y triunfos, ninguno de los cuales habría sido posible sin el apoyo de personas importantes en este camino. En primer lugar, extendiendo mi más sincera gratitud al Dr. Cristian Rojas Villa, mi director de tesis, cuya experiencia y dedicación han sido la brújula que guio este proyecto. Sus consejos precisos, su paciencia inagotable y su fe en mi trabajo fue la piedra angular sobre la que se forjó esta investigación.

La sabiduría y pasión transmitidas por mis profesores a lo largo de mi carrera universitaria han sido fundamentales en mi formación. Cada lección, cada enseñanza y cada retroalimentación han esculpido al profesional en que me he convertido. Merece un reconocimiento especial el Dr. Diego Suárez, cuyo entusiasmo contagioso y profundo conocimiento en la ingeniería en alimentos encendieron en mí una pasión que definió mi trayectoria académica.

Detrás de cada logro académico, se encuentra el amor inquebrantable y el respaldo continuo de mi familia. Mis padres, con su trabajo incansable y su fe en mi potencial, han sido el motor que me impulsó en cada etapa. Su ejemplo de trabajo duro y su apoyo emocional me han enseñado que no hay obstáculo imposible. La presencia constante de mi hermano, su empatía en mis momentos de estrés y sus palabras de ánimo, fueron el impulso que necesitaba para renovar mis energías y seguir adelante.

Los lazos de amistad forjados durante estos años universitarios han sido un tesoro invaluable. A Daniela, Sol, Jorge y Andrés, les debo momentos de alegría que aligeraron la carga académica, y su apoyo incondicional que me sostuvo en tiempos difíciles. Su compañía ha hecho de esta etapa una experiencia inolvidable, llena de crecimiento personal y recuerdos atesorados.

Finalmente, mi reconocimiento a la Universidad del Azuay, institución que ha sido mucho más que un centro de estudios. Sus aulas han sido testigos de mi evolución intelectual, sus laboratorios han visto nacer mis ideas más innovadoras y sus espacios han albergado conversaciones que han moldeado mi ética profesional. Gracias por proporcionar un ambiente que fomenta la excelencia, la curiosidad y el compromiso social.

Este proyecto de titulación no es solo el fruto de mi esfuerzo individual, sino el resultado de una sinergia colectiva. A todos ustedes, que han sido faro, ancla y viento a favor en mi viaje académico, mi gratitud es eterna. Sus contribuciones perduren en cada página de este trabajo y en cada logro futuro.

## RESUMEN

El objetivo de este trabajo fue desarrollar un modelo computacional basado en las relaciones cuantitativas estructura-propiedad (QSPR) para predecir los índices de retención ( $I$ ) de 61 compuestos orgánicos volátiles (VOCs) identificados en variedades de quinua blanca, roja y negra. El índice de retención se cuantificó mediante cromatografía de gases acoplado a espectrometría de masas, utilizando la columna capilar FS-SE-54-CB-1 (sílice fundida-5% fenilo 95% dimetilpolisiloxano). La geometría molecular de los compuestos volátiles se optimizó mediante la herramienta computacional Atomic Charger Calculator II. Posteriormente, el método de subconjuntos balanceados (BSM) permitió dividir la base de datos en conjuntos de calibración y predicción en una relación 80/20. Con los 48 compuestos del grupo de calibración, se aplicó la regresión de mínimos cuadrados ordinarios (OLS) con el método de reemplazo (RM) para la selección de descriptores. Para el diagnóstico del modelo se analizó el coeficiente de determinación ( $R^2$ ) y el error cuadrático medio ( $RMSE$ ). Se obtuvo un modelo óptimo con 4 descriptores, el cual presenta buen desempeño en calibración ( $R^2 = 0.956$  y  $RMSE = 32.59$ ) y buena capacidad de estimación de los  $I$  de los 13 compuestos del grupo de predicción ( $R^2 = 0.969$  y  $RMSE = 27.52$ ). Adicionalmente, se aplicaron diversas técnicas de validación interna para medir la estabilidad del modelo y descartar la presencia de sobreajuste. El modelo se desarrolló siguiendo los lineamientos de Organización para la Cooperación Económica y el Desarrollo para garantizar su fiabilidad, proporcionando una comprensión computacional de la retención cromatográfica de compuestos volátiles en muestras de quinua.

**Palabras clave:** *Quinua, índices de retención, relaciones cuantitativas estructura-propiedad, descriptores moleculares, compuestos orgánicos volátiles, in silico.*

## ABSTRACT

The aim of this work was to develop a computational model based on quantitative structure-property relationships (QSPR) to predict the retention index ( $I$ ) of 61 volatile organic compounds (VOCs) identified in white, red, and black quinoa varieties. The retention index was quantified by gas chromatography coupled with mass spectrometry, using the FS-SE-54-CB-1 capillary column (fused silica-5% phenyl 95% dimethylpolysiloxane) for separation. The molecular geometry of the volatile compounds was optimized using the Atomic Charger Calculator II webserver. Subsequently, the balanced subsets method (BSM) allowed the division of the database into training and prediction sets in a ratio 80/20. With the 48 compounds of the training set, ordinary least squares regression (OLS) was applied with the replacement method (RM) for descriptor selection. The coefficient of determination ( $R^2$ ) and the root mean squared error ( $RMSE$ ) were used to analyze the quality of the model. An optimal model with 4 descriptors was obtained, which exhibits good performance in calibration ( $R^2 = 0.956$  and  $RMSE = 32.59$ ) and good predictive ability when applied to the 13 compounds of the test set ( $R^2 = 0.969$  and  $RMSE = 27.52$ ). Additionally, internal validation approaches were applied to quantify the stability of the model and confirm the absence of change correlation. The model has been developed following the guidelines proposed by the Organisation for Economic Co-operation and Development to ensure its reliability, thus providing a computational understanding of the physicochemical phenomenon of chromatographic retention of volatile compounds present in quinoa samples.

**Keywords:** *Quinoa, volatile organic compounds, retention index, quantitative structure-property relationships, molecular descriptors, in silico.*

## ÍNDICE

DEDICATORIA .....	i
AGRADECIMIENTOS .....	ii
RESUMEN .....	iii
ABSTRACT .....	iv
INTRODUCCIÓN .....	1
CAPÍTULO I	
1. Marco teorico .....	2
1.1. Generalidades de la quinua .....	2
1.1.1. Composición química y beneficios nutricionales de la quinua.....	2
1.1.2. Usos y tipos de quinua .....	3
1.2. Caracterización de compuestos orgánicos volátiles de la quinua .....	5
1.2.1. Perfil aromático de diferentes variedades de quinua .....	5
1.2.2. Cambios en los compuestos orgánicos volátiles durante el procesamiento postcosecha.....	6
1.2.3. Actividad antioxidante de los compuestos volátiles .....	6
1.3. Identificación por espacio de cabeza en cromatografía de gases acoplado a espectrometría de masas.....	7
1.3.1. Principios y fundamentos de la técnica.....	7
1.3.2. Aplicaciones.....	8
1.3.3. Optimización y validación del modelo .....	9
1.3.4. Identificación y cuantificación.....	10
1.4. Relaciones cuantitativas estructura-propiedad.....	11
1.4.1. Fundamentos del modelo QSPR .....	11
1.4.2. Aplicaciones de los modelos QSPR para la predicción de propiedades fisicoquímicas .....	12
1.4.3. Uso de los modelos QSPR en el diseño molecular asistido por computadora.....	13
CAPÍTULO II	
2. Materiales y métodos .....	14
2.1. Generación de la base de datos .....	14

2.2.	Filtrado de la base de datos .....	15
2.3.	Representación molecular y optimización de geometrías .....	17
2.4.	Cálculo de los descriptores moleculares .....	18
2.5.	Aprendizaje no supervisado para la reducción de descriptores moleculares .....	19
2.6.	Desarrollo del modelo QSPR .....	20
2.7.	Validación del modelo QSPR .....	21
2.8.	Análisis de la contribución de los descriptores moleculares .....	22
2.9.	Mecanismo de acción de los descriptores moleculares .....	22
2.10.	Análisis del dominio de aplicabilidad .....	23
<b>CAPÍTULO III</b>		
3.	Resultados .....	24
3.1.	Generación de la base de datos .....	24
3.2.	Cálculo de los descriptores moleculares .....	24
3.3.	Desarrollo del modelo QSPR .....	24
3.4.	Validación del modelo QSPR .....	26
3.5.	Análisis del grado de contribución de los descriptores moleculares .....	29
3.6.	Análisis del dominio de aplicabilidad .....	31
<b>CONCLUSIONES .....</b>		<b>32</b>
<b>REFERENCIAS.....</b>		<b>33</b>

## INDICE DE FIGURAS

<b>Figura 1.</b> Moléculas que representan la presencia de dímeros .....	17
<b>Figura 2.</b> Índices de retención experimental vs. los índices de retención predichos de los VOCs en semillas de quinua medidos mediante HC-GC-MS .....	28
<b>Figura 3.</b> Grafica de dispersión de los residuos estandarizados vs. los índices de predicción predichos. ....	28
<b>Figura 4.</b> Grado de contribución de los descriptores moleculares para los VOCs presentes en muestras de quinua.....	29
<b>Figura 5.</b> Gráfico de residuos normalizados frente a valores de influencia para delimitar el espacio de aplicación válido del modelo QSPR. ....	31

## ÍNDICE DE TABLAS

<b>Tabla 1.</b> Cantidad de descriptores moleculares obtenidos por medio del método de reemplazo	26
<b>Tabla 2.</b> Resultados obtenidos para el modelo QSPR en los grupos de calibración, validación cruzada y predicción .....	27

## INTRODUCCIÓN

En la era actual impulsada por los avances científicos y tecnológicos, la química computacional ha nacido como un campo interdisciplinario de vanguardia que combina principios de la química, la física y la matemática con el poder del cómputo moderno. Esta disciplina ha revolucionado la forma en el que se aborda el estudio y la predicción de propiedades fisicoquímicas de compuestos orgánicos, impactando áreas tan diversas como el diseño de fármacos, la catálisis, la química de materiales y la industria alimentaria. Dentro de este contexto, las Relaciones Cuantitativas Estructura-Propiedad (QSPR) han demostrado ser herramientas poderosas e indispensables para el estudio y predicción de propiedades de interés de los compuestos químicos. Estos modelos establecen correlaciones matemáticas entre los descriptores moleculares, que codifican información estructural, electrónica y estérica de las moléculas, y propiedades fisicoquímicas de interés. De esta forma, los modelos QSPR permiten predecir las propiedades de nuevos compuestos sin la necesidad de llevar a cabo costosos y prolongados ensayos experimentales, acelerando así el proceso de descubrimiento y optimización molecular.

El presente trabajo de titulación se centra en el desarrollo de un modelo QSPR para la predicción de los índices de retención cromatográficos de compuestos orgánicos volátiles presentes en semillas de quinua. El índice de retención es la respuesta analítica en técnicas de laboratorio como la cromatografía de gases. Esta propiedad permite la identificación y cuantificación de compuestos en matrices complejas, lo cual es fundamental en diversas aplicaciones, desde el control de calidad en la industria alimentaria hasta la detección de contaminantes ambientales. Mediante un enfoque sistemático y riguroso basado en técnicas quimiinformáticas apropiadas, se llevará a cabo un exhaustivo proceso de selección de descriptores moleculares relevantes a partir de un conjunto inicial de miles de ellos. Los descriptores son variables numéricas que codifican información estructural, electrónica, estérica y topológica de las moléculas. Empleando métodos de aprendizaje automático supervisado y regresión lineal, se construirá un modelo QSPR robusto y altamente predictivo, capaz de estimar con una precisión, el índice de retención de los compuestos volátiles.

Además, se implementarán rigurosamente diversas estrategias de validación interna y externa, incluyendo técnicas como validación cruzada y aleatorización-Y, con el objetivo de garantizar la solidez estadística del modelo y su aplicabilidad en un amplio dominio químico. Asimismo, se realizará un análisis de la contribución relativa de cada descriptor molecular al modelo, revelando información válida del mecanismo de acción que gobierna la compleja relación estructura-propiedad. Los resultados de este trabajo de titulación no sólo brindarán conocimientos fundamentales sobre los factores estructurales que influyen en el índice de retención de compuestos orgánicos volátiles, sino que también sentarán las bases para futuras aplicaciones en la industria alimentaria, como la optimización de perfiles aromáticos, el diseño de aditivos alimentarios y la determinación de compuestos responsables de sabores y olores característicos en diversos productos. Además, la metodología desarrollada podrá ser fácilmente transferible a otros sistemas químicos y propiedades de interés, convirtiéndose en una herramienta versátil y poderosa para la exploración racional del vasto espacio químico.

# CAPÍTULO I

## 1. MARCO TEORICO

### 1.1. Generalidades de la quinua

La quinua (*Chenopodium quinoa*) es, seguramente, uno de los cultivos más significativos del mundo, puesto que se trata de un pseudocereal con origen en los Andes Sudamericanos, específicamente en Perú y Bolivia, donde ha sido domesticado y consumido desde hace miles de años (Rubén, Burin, Pereyra, & Heras, 2015). Antiguamente, los incas llamaban a la quinua “*chisaya mama*”, que significa “*madre de todos los granos*”, pues creían que este alimento sagrado les era dado por los dioses como una fuente de nutrientes, energía y salud. En la época del imperio incaico el consumo de quinua estaba reservado mayormente para la realeza y la clase alta. Tras la conquista española, esta semilla comenzó a ser desplazado por los cereales europeos como el trigo, pero en las últimas décadas este pseudocereal ha experimentado un resurgimiento (Cusack, 1984; Mujica & Jacobsen, 2006). Uno de los aspectos más destacados de la quinua es que se trata de un alimento libre de gluten, por lo que se posiciona como una excelente alternativa para aquellas personas sensibles al gluten o que padecen la enfermedad celíaca. Asimismo, la quinua sobresale por ser una semilla altamente nutritiva, puesto que sus vitaminas, minerales, antioxidantes flavonoides, compuestos antiinflamatorios y todos los aminoácidos esenciales, son benéficos para la salud (Ranjan et al., 2023; Rubén et al., 2015).

#### *1.1.1. Composición química y beneficios nutricionales de la quinua*

En comparación con otros cereales, la quinua tiene mayor cantidad de proteína de origen vegetal: por cada 100 gramos de porción, la quinua aporta alrededor de 14 gramos de proteína de alta calidad; mientras que el arroz o la avena poseen solo entre 7 a 10 gramos de proteína, respectivamente (Pérez Ramos, Elías Peñafiel, & Delgado Soriano, 2017). La proteína de la quinua es considerada una de las proteínas más completas, tiene un alto porcentaje de digestibilidad, con una excelente biodisponibilidad. Además, la quinua posee todos los aminoácidos esenciales necesarios para el ser humano, convirtiéndola en una fuente ideal de nitrógeno especialmente para dietas vegetarianas y en poblaciones vulnerables (Campos-Rodriguez, Acosta-Coral, & Paucar-Menacho, 2022). Por otro lado, la quinua tiene un 60% más de fibra dietética que la mayoría de los cereales, lo cual beneficia el tránsito intestinal y favorece la saciedad. Complementariamente, este alimento es rico en magnesio, zinc, hierro y lisina (Villacrés et al., 2015), además, este pseudocereal contiene altos niveles de flavonoides, compuestos antioxidantes con acción antiinflamatoria, los cuales protegen a las células del daño oxidativo asociado con enfermedades degenerativas como cáncer, Alzheimer o párkinson (Bernuy, Villanueva, Suárez, & Vilchez, 2018; Shah, Khan, & Yan, 2022).

Los hidratos de carbono en la quinua son complejos, el almidón se digiere y absorbe lentamente, evitando así los picos altos de glucosa en la sangre después de las comidas (un alto valor glicémico después de comidas). Por esta razón, se considera un alimento apto para diabéticos y en dietas de control glicémico (Mohyuddin, 2019). Adicionalmente, diversos estudios coinciden en que la quinua reduce de manera favorable los niveles de colesterol total en sangre. Esta acción hipocolesterolémica se atribuye principalmente al contenido de compuestos insaponificables como los esteroides de su aceite, así como los antioxidantes polifenólicos (Bernuy et al., 2018; Shah et al., 2022).

### ***1.1.2. Usos y tipos de quinua***

La quinua posee un sabor agradable, delicado y ligeramente amargo, con una textura esponjosa al cocinarla. Resulta ser un ingrediente muy versátil para incluir en distintas preparaciones, tanto en platos salados como dulces. Los granos de quinua cocidos pueden consumirse de distintas maneras, por ejemplo, ensaladas, sopas, guisos, tortillas, postres e incluso en jugos y batidos (W. Rojas, Vargas Mena, & Pinto Porcel, 2016). Asimismo, se puede utilizar la harina de quinua para preparar panes, tortas, galletas, brownies, magdalenas y todo tipo de horneados libre de gluten. Se prepara también leche de quinua como alternativa vegetal saludable y sin alérgenos (Romo, Rosero, Forero, & Céron, 2006). Por otro lado, a lo largo de los Andes sudamericanos, en la zona originaria de este pseudocereal se han identificado cerca de 150 a 200 variedades diferentes de quinua, las cuales se clasifican principalmente en función del color y tamaño del grano, el tamaño de la planta, las características agronómicas, contenido de saponina y requisitos ambientales (Samame, Monteza, Diaz, & Cabrera, 2017; Voronov, Pleskachiov, Shitikova, Zargar, & Abdelkader, 2023).

#### ***Quinua por color del grano***

El color del perigonio o cubierta del grano es uno de los principales criterios para categorizar los distintos tipos de quinua.

- **Quinua blanca o quinua real**

Es la variedad más común y consumida, su grano es de color blanquecino o amarillo pálido. La planta alcanza un metro de altura con inflorescencias grandes. Destaca por su precocidad, dulzor, tamaño grande del grano, buen rendimiento y menor contenido de saponinas (Arzapalo Quinto, Huamán Cándor, Quispe Solano, & Espinoza Silva, 2015).

- **Quinua roja**

Su perigonio es de color rojizo, rosado o anaranjado. La semilla es pequeño, arrugado y de maduración intermedia. Posee un considerable contenido proteico (Arzapalo Quinto et al., 2015).

- **Quinua negra**

Presenta coloración púrpura oscuro o negro. El grano es pequeño, arrugado y de maduración tardía, rica en proteínas, antioxidantes y compuestos bioactivos como antocianinas (Arzapalo Quinto et al., 2015).

- **Quinoa amarilla**

Variedad de reciente obtención con perigonio amarillo dorado y grano grande de excelente calidad culinaria (Arzapalo Quinto et al., 2015).

### *Quinoa por contenido de saponinas*

La saponina es un glucósido triterpenoide que se encuentra de forma natural en la cubierta de los granos de quinoa. Le otorga el sabor amargo característico. Según su contenido en saponinas las quinuas pueden clasificarse en:

- **Quinoa dulce**

Variedades con bajo contenido de saponinas, menor al 0.11%. Su sabor es dulce y no requiere lavado previo ya que la saponina ha sido eliminada mediante selección genética (Delgado, Palacios, & Betancourt, 2009).

- **Quinoa amarga**

Variedades silvestres o no mejoradas con niveles de saponinas mayor al 0.11%. Tienen sabor amargo por lo que necesitan lavado exhaustivo antes de consumirlas para eliminar las saponinas solubles (Mamani, Bonifacio, Taboada, & Angulo, 2022).

### *Quinuas por ecotipo y origen*

Según la zona agro-ecológica y país de origen se distinguen (Barrientos, Carevic, & Delatorre, 2017; Vilca, Espinoza, & Vidal, 2015):

- **Quinoa de la puna o altiplano**

Proceden de la meseta del altiplano a más de 3000 msnm. Presentan ciclos cortos y granos pequeños.

- **Quinoa de valles interandinos**

Proceden de los cales y salares entre 2000 a 3000 msnm, son plantas altas, de ciclo largo y grano grande.

- **Quinoa de nivel del mar**

Este tipo procede de Ecuador, Chile y Colombia adaptadas para cultivarse por debajo de los 2000 msnm.

Otras clasificaciones consideran el tamaño de la planta (gigante, enana o baja), forma de grano (ovalado, esférico o lenticular), grosor de grano, días de madurez fisiológica e inclusive fragancia del grano. Como se puede ver, existe una amplia diversidad de tipos de quinua en los Andes, cada una con propiedades y usos particulares.

La mayor biodiversidad de quinua a nivel mundial se encuentra en bancos de germoplasma localizados en Bolivia y Perú (Mujica & Jacobsen, 2006). Asimismo, se considera que la quinua es un cultivo muy resiliente, pues crece en diversos pisos ecológicos y se adapta para tolerar condiciones extremas. Su diversificación le ha permitido sobrevivir diversos años y ser un recurso alimentario clave para las poblaciones andinas (Samame et al., 2017). En la actualidad, la quinua se sigue cultivando principalmente en Perú y Bolivia, pero su consumo se ha extendido ampliamente por todo el mundo, considerándola así un “súper alimento”. Se espera que su producción y consumo a nivel mundial siga en aumento debido a la creciente demanda de alimentos orgánicos, sostenibles y libres de alérgenos (Rubén et al., 2015).

## **1.2. Caracterización de compuestos orgánicos volátiles de la quinua**

### ***1.2.1. Perfil aromático de diferentes variedades de quinua***

La quinua al ser un cultivo con una variedad considerable de variedades y ecotipos que son cultivadas en distintas zonas de los Andes del Sur, puede diferir en sus características físicas, así como en sus cualidades organolépticas como el sabor y aroma. Varios estudios se han enfocado en caracterizar el perfil de compuestos orgánicos volátiles (VOCs: Volatile Organic Compounds)(Cserhádi, 2010) en diferentes tipos de quinua y entender su impacto en las propiedades sensoriales. Mediante el uso de técnicas como la cromatografía de gases acoplada a espectrometría de masas (GC-MS: Gas Chromatography-Mass Spectrometry), los investigadores han logrado identificar y cuantificar cientos de compuestos volátiles en variedades de quinua blanca, roja y negra. Algunas de las principales familias químicas halladas son: alcoholes, aldehídos, cetonas, piridinas, furanos, pirazinas, terpenos y compuestos azufrados (Almaguer, Kollmannsberger, Gastl, & Becker, 2023). Se ha detectado una diferenciación evidente en la composición y concentración relativas de los compuestos volátiles entre variedades de quinua. Por ejemplo, la quinua blanca muestra mayores niveles de (*E*)-2-nonenal, relacionado con aromas de avena y trigo, mientras que la quinua roja destaca por contenidos más altos de  $\beta$ -ionona responsable de notas florales a violeta (Balladares, Martínez, & Rojas, 2015). Asimismo, cada tipo de quinua presenta un distinto grupo de compuestos azufrados como disulfuros, tioles, tiofenos y sulfuros que influyen fuertemente en sus notas aromáticas. Algunos contribuyen con aromas agradables a cebolla, ajo, cítricos, mientras que otros causan olores desagradables (Belitz, Grosch, & Schieberle, 2008).

### ***1.2.2. Cambios en los compuestos orgánicos volátiles durante el procesamiento postcosecha***

Luego de la etapa de cultivo y cosecha, la quinua debe someterse a una serie de procesos postcosecha que tienen como objetivo limpiar y secar de manera correcta los granos para prolongar su vida útil y prepararlos para consumo humano. Algunas de estas etapas incluyen el secado inicial, descascarillado, separación de saponinas, lavado, secado final y almacenamiento (Song et al., 2021). Investigaciones recientes mediante GC-MS han analizado como estos procesos de postcosecha modifican el perfil original de compuestos orgánicos volátiles de las variedades de quinua cruda. Se ha observado que algunos grupos de compuestos como terpenos, norisoprenoides, bencenos y alcoholes disminuyen drásticamente durante el lavado acuoso. Esto se debe probablemente a la solubilización y eliminación de estos compuestos durante la inmersión para separar las saponinas amargas (Yang et al., 2022). Por otra parte, el tostado de la quinua puede aumentar los niveles de pirroles, piridinas y furanos, producto de reacciones térmicas como la degradación de azúcares durante el calentamiento. Asimismo, el almacenamiento prolongado incrementa la concentración de ciertos aldehídos como 2-decenal asociados a aromas rancios por posible oxidación de ácidos grasos insaturados (Chagua Lazo & Palomino Villazan, 2014). La investigación apunta a que la intensidad del tratamiento postcosecha influye de manera significativa tanto en la pérdida de compuestos volátiles beneficiosos para el aroma de la quinua, así como en la formación de nuevos compuestos indeseables que impactan negativamente las cualidades sensoriales. Mientras más drástico sea el proceso de lavado acuoso o tostado, mayores alteraciones provocará en el perfil original de los compuestos volátiles del grano fresco (Almaguer et al., 2023).

### ***1.2.3. Actividad antioxidante de los compuestos volátiles***

La quinua contiene varios metabolitos secundarios bioactivos como saponinas, fitoesteroles, flavonoides, y una variedad de compuestos orgánicos volátiles que contribuyen a su distintivo aroma (Luisetti, Lucero, & Ciappini, 2020). Muchos de los compuestos volátiles de la quinua exhiben interesantes propiedades antioxidantes relacionadas con su capacidad para inhibir la oxidación de biomoléculas como lipoproteínas, ácido desoxirribonucleico (ADN) y ácidos grasos poliinsaturados, mediante la neutralización de formas reactivas de oxígeno (Luisetti et al., 2020). Un dato importante a considerar es que la actividad antioxidante de extractos volátiles de quinua es similar a la de los compuestos puros, e incluso, superior en algunos casos a antioxidantes comerciales como el butilhidroxitolueno (BHT) y la vitamina C o ácido ascórbico (Laus, Gagliardi, Soccio, Flagella, & Pastore, 2012). El mecanismo antioxidante parece estar relacionado con la capacidad para donar átomos de hidrógeno o electrones para neutralizar radicales libres, además de posibles efectos quelantes de metales de transición (Laus et al., 2012). Se ha observado que el contenido de compuestos volátiles en la quinua depende del tipo y condiciones ambientales de

crecimiento; además de factores relacionados al procesamiento postcosecha como tiempo y temperatura de almacenamiento, tratamientos térmicos, presión, entre otros (Luisetti et al., 2020).

### **1.3. Identificación por espacio de cabeza en cromatografía de gases acoplado a espectrometría de masas**

#### ***1.3.1. Principios y fundamentos de la técnica***

La identificación por espacio de cabeza acoplada a cromatografía de gases y espectrometría de masas (HS-GC-MS: Head Space-Gas Chromatography-Mass Spectrometry) (Feng, Wang, Ji, Min, & Yan, 2021) es una técnica analítica habitualmente empleada para la identificación y cuantificación de compuestos en una gran variedad de matrices, especialmente para analitos volátiles y semivolátiles (Gutiérrez Bouzán & Droguet, 2002). Esta técnica se basa en la inyección directa de muestras en la fase gaseosa sobre la columna cromatográfica, sin división de flujo hacia el detector de ionización de llama. De esta manera, los analitos no son separados al inyectar la muestra y la columna no se sobrecarga, permitiendo la detección de trazas de analitos incluso en matrices complejas (Han et al., 2023). La técnica involucra tres componentes acoplados: el espacio de cabeza (HS), la cromatografía de gases (GC) y la espectrometría de masas (MS). El espacio de cabeza se refiere a la fase gaseosa sobre una muestra sólida o líquida, la cual contendrá los analitos volátiles y vaporizables (Gutiérrez Bouzán & Droguet, 2002). En la GC, los componentes de una mezcla compleja se separan según la velocidad por la que pasan por la columna cromatográfica, debido a las diversas interacciones con la fase estacionaria. Al salir de la columna, estos componentes entran al espectrómetro de masas secuencialmente (Feng et al., 2021), donde son ionizados (generalmente por impacto electrónico), acelerados y luego separados en un analizador basado en relaciones masa/carga. El detector mide la abundancia iónica y genera los espectros de masas, los cuales son patrones representativos de la estructura de cada analito.

Entre las principales ventajas del HS-GC-MS se encuentran: alta sensibilidad y bajos límites de detección, identificación inequívoca de los compuestos mediante sus espectros de masas, capacidad de análisis tanto cualitativo como cuantitativo con amplio rango lineal, y versatilidad para acomodar distintos tipos de muestras con una preparación mínima (Feng et al., 2021). La sensibilidad analítica está en el orden de los picogramos, muy por debajo de las exigencias de numerosas aplicaciones regulatorias. Además, los mejores instrumentos recientes incluyen sistemas de inyección más avanzados, interfaces GC-MS más eficientes, potentes sistemas de vacío y nuevos tipos de columnas cromatográficas (Hwang et al., 2019). Dentro de las principales etapas de un análisis por HS-GC-MS se encuentran (Cozzolino, 2023):

1. Preparación de la muestra: a pesar que se requiere poca preparación de muestra, usualmente involucra etapas como pesaje, filtraciones, diluciones o ajustes en la matriz para mejorar la efectividad del análisis.

2. Generación del espacio de cabeza: se coloca la muestra en un vial sellado y sometido a una temperatura constante, lo que provoca la vaporización de los analitos volátiles.
3. Inyección en el cromatógrafo de gases: se toma una alícuota definida del espacio de cabeza formado sobre la muestra y se inyecta directamente a la columna cromatográfica.
4. Separación: la separación ocurre en la columna cromatográfica capilar basándose en la afinidad diferencial de cada analito por la fase estacionaria seleccionada.
5. Detección por espectrometría de masas: los analitos entran al espectrómetro de masas de manera secuencial donde son ionizados, acelerados y separados para generar los espectros característicos.
6. Identificación y cuantificación: los espectros son comparados con librerías para identificar cada pico. Además, mediante estándares se puede generar curvas de calibración para la cuantificación de los analitos.

### *1.3.2. Aplicaciones*

La técnica de identificación por espacio de cabeza en cromatografía de gases acoplada a espectrometría de masas (HS-GC-MS) ha encontrado numerosas aplicaciones en la industria alimentaria debido a su versatilidad y capacidad para analizar compuestos volátiles y semivolátiles presentes en diversas matrices alimentarias. En este sentido, una de las principales aplicaciones de HS-GC-MS es el análisis de aromas y compuestos volátiles responsables del sabor y el aroma característicos de diversos productos. Por ejemplo, este método permite realizar un análisis cuantitativo y cualitativo de los compuestos volátiles que se encuentran en frutas, verduras, carnes, productos lácteos, bebidas alcohólicas y no alcohólicas, entre otros (Plutowska & Wardencki, 2008). El análisis de aromas se considera un punto muy importante en la evaluación de la calidad de las matrices alimentarias, así como para monitorear los cambios que ocurren durante el procesamiento, el almacenamiento y la maduración (Cajka, Hajslova, Pudil, & Ridellova, 2009). Además, la información obtenida mediante HS-GC-MS puede utilizarse para el desarrollo de nuevos productos, la optimización de procesos y la autenticación de alimentos y bebidas (Vélez, Tabares, & Durán, 2022). El control de calidad y la seguridad alimentaria es un ámbito en el cual se ha usado dicha técnica, permitiendo la detección y cuantificación de compuestos indeseables, contaminantes y residuos en alimentos y bebidas. Algunos ejemplos incluyen el análisis de residuos de disolventes, compuestos derivados del proceso de fermentación, compuestos tóxicos generados durante el procesamiento térmico y contaminantes ambientales como hidrocarburos aromáticos policíclicos (PAHs) (Antolín, Acitores, Jiménez, & Vega).

La técnica HS-GC-MS se emplea también en la identificación de compuestos que desencadenan olores y sabores indeseables, lo cual es fundamental con el fin de proporcionar un producto alimenticio de excelente calidad sensorial (Vélez et al., 2022). Adicional a esto, HS-GC-MS también se ha utilizado en la autenticación y trazabilidad de alimentos, lo que es especialmente importante para proteger a los consumidores de fraudes y asegurar la seguridad y calidad de los

productos. Mediante el análisis de los perfiles de compuestos volátiles y semivolátiles es posible diferenciar alimentos de distintos orígenes geográficos, variedades o métodos de producción (Cajka et al., 2009). Por ejemplo, HS-GC-MS se ha aplicado en la autenticación de aceites de oliva, vinos, cafés, mieles y otros productos alimenticios, utilizando los compuestos volátiles como “huellas digitales” únicas (Antolín et al.). Así, la HS-GC-MS se la encuentra como una excelente y valiosa herramienta en el campo de la investigación y desarrollo de nuevos productos alimenticios. Asimismo, es crucial mejorar los procesos y la calidad de los productos alimenticios, optimizando los cambios de los compuestos volátiles durante el procesamiento, almacenamiento y vida útil de los alimentos; lo que claramente influye en su aceptabilidad (Plutowska & Wardencki, 2008). Es importante mencionar que HS-GC-MS es también utilizada para evaluar el impacto de diferentes ingredientes, aditivos y tecnologías de procesamiento en el perfil de compuestos volátiles de los alimentos, lo que facilita la creación de productos innovadoras con un perfil sensorial deseado (Vélez et al., 2022).

### ***1.3.3. Optimización y validación del modelo***

La optimización y validación del método de cromatografía de gases acoplada a espectrometría de masas (GC-MS) es una etapa importante en el análisis de alimentos, ya que permite el análisis cuantitativo y cualitativo precisa de compuestos volátiles y semivolátiles presentes en matrices complejas. El proceso de optimización se enfoca en ajustar diversos parámetros instrumentales y condiciones analíticas para obtener una separación eficiente de los analitos, una buena sensibilidad y una alta resolución. Algunos de los aspectos clave a considerar son la selección de la columna cromatográfica adecuada, el programa de temperatura del horno, la velocidad de flujo del gas portador, la temperatura de inyección y la fuente de ionización (Chinnici et al., 2009). Además, la fase de preparación de la muestra es crucial, ya que tiene un impacto significativo en la calidad de los resultados obtenidos. Se utilizan distintas técnicas de extracción, como la extracción líquido-líquido, la extracción en fase sólida (SPE: Solid Phase Extraction) o la microextracción en fase sólida (SPME: Solid Phase MicroExtraction), para separar los analitos de interés de la matriz alimentaria y concentrarlos antes de su inyección en el sistema GC-MS (Risticvic, Carasek, & Pawliszyn, 2008).

Una vez optimizado el método, es necesario realizar una validación rigurosa para asegurar su idoneidad y confiabilidad. La validación implica la evaluación de parámetros como la linealidad, precisión, exactitud, límites de detección y cuantificación, selectividad, robustez y estabilidad de las muestras (Chinnici et al., 2009). Estos parámetros se determinan mediante experimentos específicos y el análisis de muestras fortificadas con concentraciones conocidas de los analitos. La linealidad se evalúa mediante la construcción de curvas de calibración y el cálculo del coeficiente de correlación. La precisión se determina a través de la repetibilidad y la precisión intermedia, utilizando muestras replicadas y analizadas en diferentes días. La exactitud se evalúa mediante ensayos de recuperación, donde se añaden cantidades conocidas de analitos a las muestras y se calcula el porcentaje de recuperación (Risticvic et al., 2008). Los límites de detección y

cuantificación se calculan a partir de la relación señal/ruido y la desviación estándar de las mediciones de muestras con bajas concentraciones de analitos. La selectividad se evalúa analizando muestras blanco y muestras fortificadas para verificar la ausencia de interferencias. La robustez se determina mediante pruebas de estabilidad frente a pequeñas variaciones en los parámetros del método (Chinnici et al., 2009; Risticevic et al., 2008).

#### ***1.3.4. Identificación y cuantificación***

La identificación de los compuestos se realiza mediante la comparación de los espectros de masas obtenidos con los de una biblioteca de espectros de referencia. Estas bibliotecas contienen los espectros de masas de numerosos compuestos, junto con información sobre sus tiempos de retención y otros datos relevantes (Standards & Technology, 2017). Las principales bibliotecas de espectros de masas utilizadas en HS-GC-MS incluyen (Standards & Technology, 2017):

- NIST (National Institute of Standards and Technology)
- Wiley Registry
- Biblioteca de espectros de masas del fabricante del equipo

La identificación de compuestos no se basa únicamente en la comparación de espectros de masas, sino que se complementa con el análisis de los tiempos de retención cromatográficos y el uso de patrones analíticos de referencia, cuando estos se encuentran disponibles (Plutowska & Wardencki, 2008). Adicionalmente, para cuantificar los compuestos identificados, se emplean métodos como la calibración externa o el método del estándar interno. En ambos casos, se construyen curvas de calibración utilizando soluciones estándar de concentración conocida (Koek, Jellema, van der Greef, Tas, & Hankemeier, 2011). En el método de calibración externa, se preparan soluciones estándar de los analitos de interés y se inyectan en el sistema HS-GC-MS. Para determinar las concentraciones de los compuestos analitos en las muestras, se genera inicialmente una curva de calibración. Esta curva se construye relacionando concentraciones conocidas de estándares con las áreas o alturas de los picos cromatográficos que generan. Posteriormente, las concentraciones de los analitos presentes en las muestras problema se calculan interpolando sus respectivas áreas o alturas de pico en la curva de calibración previamente establecida (Plutowska & Wardencki, 2008).

Por el contrario, en el método del estándar interno se añade una cantidad conocida de un compuesto de referencia (estándar interno) a las soluciones estándar y a las muestras. La cuantificación se realiza mediante la relación entre las áreas o alturas de los picos del analito y del estándar interno. Este método compensa las variaciones en la inyección y el proceso de extracción, mejorando la precisión y la exactitud de la cuantificación. Además de los métodos de calibración, existen otros enfoques para la cuantificación, como el método de adición de estándar y el método de dilución isotópica. Estos métodos se utilizan cuando hay efectos de matriz significativos o se requiere una mayor exactitud en la cuantificación. Para garantizar la confiabilidad de los resultados obtenidos mediante HS-GC-MS, es necesario evaluar y validar los parámetros de calidad analítica, como la

linealidad, la precisión, la exactitud, los límites de detección y cuantificación, la selectividad y la robustez del método (Koek et al., 2011).

## **1.4. Relaciones cuantitativas estructura-propiedad**

### ***1.4.1. Fundamentos del modelo QSPR***

Las relaciones cuantitativas estructura-propiedad (QSPR: Quantitative Structure-Property Relationships) son herramientas de modelado que relacionan propiedades fisicoquímicas de compuestos químicos con descriptores numéricos que codifican información acerca de su estructura química (C. Rojas & Duchowicz, 2021). Estos modelos se basan en el principio que menciona “las propiedades moleculares están determinadas por la estructura química del compuesto”. De esta forma, compuestos similares tenderán a poseer propiedades experimentales comparables (C. Rojas & Duchowicz, 2021). Los modelos QSPR consisten en ecuaciones matemáticas obtenidas a partir de un conjunto de datos que contienen los valores de una propiedad dada medidos experimentalmente y los valores de un grupo de descriptores moleculares, para un grupo de compuestos con cierta similitud estructural (C. Rojas & Duchowicz, 2021; Varmuza, Dehmer, & Bonchev, 2012).

La gran mayoría de modelos QSPR reportados en la literatura se basan en análisis multivariantes, en especial la regresión lineal múltiple (MLR: Multiple Linear Regression) (C. Rojas & Duchowicz, 2021). Su finalidad es modelar la propiedad objetivo como una función lineal de los descriptores seleccionados. Sin embargo, además de los métodos mencionados anteriormente, también se utilizan técnicas quimiométricas avanzadas como el análisis de componentes principales (PCA), la regresión de mínimos cuadrados parciales (PLS), el análisis discriminante (DA), las máquinas de soporte vectorial (SVM), las redes neuronales artificiales (ANN) algoritmos genéticos (GAs), entre otros; para el tratamiento de datos en la identificación y cuantificación de compuestos. (C. Rojas & Duchowicz, 2021). Los descriptores moleculares son las variables independientes del modelo que codifican matemáticamente los aspectos estructurales relevantes del grupo de moléculas que determinan la propiedad modelada. Entre los descriptores más utilizados están los constitucionales (peso molecular, número de átomos y enlaces, etc.), topológicos (índices de conectividad), geométricos (volumen, área superficial), cuánticos (cargas, orbitales HOMO/LUMO) y otros que consideran características estructurales específicas (Varmuza et al., 2012). El éxito de un modelo QSPR depende en gran medida de la calidad y cantidad de datos experimentales empleados, la selección apropiada de descriptores y el método matemático-estadístico usado para derivar la ecuación. También es crítico validar adecuadamente los modelos antes de emplearlos para hacer predicciones sobre nuevas moléculas (Varmuza et al., 2012).

#### *1.4.2. Aplicaciones de los modelos QSPR para la predicción de propiedades fisicoquímicas*

Las Relaciones Cuantitativas Estructura-Propiedad (QSPR) han adquirido una importancia significativa en el campo de la química de los alimentos, especialmente en la predicción de propiedades fisicoquímicas de compuestos presentes en matrices alimentarias. Estas propiedades desempeñan un rol clave en la calidad, seguridad y aceptabilidad de los alimentos por parte de los consumidores. Una de las aplicaciones principales de los modelos QSPR en el ámbito alimentario es la predicción de propiedades organolépticas, tales como el aroma, el sabor y la textura, las cuales influyen directamente en la percepción sensorial de los productos. Los compuestos responsables de estas propiedades, como los compuestos volátiles y los precursores de aroma y sabor, pueden ser modelados utilizando descriptores moleculares adecuados (X. Xu, Luan, Liu, & Gao, 2011). Por ejemplo, los modelos QSPR se han utilizado para predecir los umbrales de percepción de aroma y los descriptores sensoriales de compuestos volátiles en alimentos, lo que permite optimizar la formulación y el procesamiento de productos alimenticios (Guichard, 2015).

Los modelos QSPR también se han aplicado en la predicción de propiedades físicas y químicas relevantes en el procesamiento y almacenamiento de alimentos, como puntos de ebullición, solubilidades, constantes de disociación, coeficientes de partición, entre otras (Katritzky et al., 2004). Estas propiedades son importantes para comprender y controlar procesos como la extracción, la fermentación, la cristalización y la estabilidad de los alimentos. Por ejemplo, los modelos QSPR han sido utilizados para predecir la solubilidad de compuestos bioactivos en matrices alimentarias, lo que facilita el desarrollo de alimentos funcionales (Dearden & Worth, 2007). Otra aplicación relevante de los modelos QSPR en el ámbito de los alimentos es la predicción de propiedades ambientales y toxicológicas de aditivos, contaminantes y residuos presentes en los productos alimenticios, estas propiedades son importantes para evaluar el impacto ambiental y los riesgos para la salud asociados al consumo de alimentos (Gramatica, 2007). Los modelos QSPR se han utilizado para predecir la biodegradabilidad, la bioacumulación y la toxicidad de compuestos presentes en alimentos, lo que contribuye a la evaluación de riesgos y a la selección de aditivos y procesos más seguros y sostenibles (Gramatica, 2007).

Además de la predicción de propiedades, los modelos QSPR también se han utilizado en el diseño y la optimización de ingredientes y aditivos alimentarios. Mediante la combinación de estos modelos con técnicas de optimización y diseño molecular, es posible explorar y generar nuevos compuestos con propiedades deseadas, como mayor estabilidad, mejores características sensoriales o mayor bioactividad (Pripp, Isaksson, Stepaniak, Sørhaug, & Ardö, 2005). Esta aplicación es particularmente útil en el desarrollo de nuevos aditivos, antioxidantes, conservantes, colorantes y compuestos bioactivos para alimentos funcionales (Pripp et al., 2005).

### *1.4.3. Uso de los modelos QSPR en el diseño molecular asistido por computadora*

El diseño molecular asistido por computadora (CAMD: Computer-Aided Molecular Design) es un enfoque que combina técnicas de modelado molecular, optimización y bases de datos de propiedades para diseñar nuevas moléculas con propiedades específicas. Este enfoque se basa en la premisa de que las propiedades de una molécula están determinadas por su estructura, y utiliza modelos QSPR como herramientas predictivas clave. El proceso de CAMD comienza con la definición de las propiedades objetivo para molécula a diseñar, como propiedades físicas, químicas, biológicas o sensoriales (Gani & Ng, 2015). A continuación, se aplican modelos QSPR para predecir estas propiedades en función de la estructura molecular. Los modelos QSPR se utilizan en el proceso de CAMD de varias maneras (Gani & Ng, 2015):

1. Generación de estructuras candidatas: los modelos QSPR se emplean para generar estructuras moleculares candidatas que potencialmente cumplan con las propiedades objetivo definidas.
2. Predicción de propiedades: una vez generadas las estructuras candidatas, los modelos QSPR se utilizan para predecir sus propiedades de interés, como solubilidad, reactividad, toxicidad, sabor, aroma, entre otras.
3. Optimización de estructuras: los resultados de las predicciones se utilizan en algoritmos de optimización para refinar y ajustar las estructuras moleculares, con el objetivo de maximizar o minimizar las propiedades deseadas.
4. Validación y selección: las estructuras optimizadas se validan utilizando criterios adicionales, como la viabilidad sintética, la estabilidad y la disponibilidad de recursos, antes de seleccionar las moléculas más prometedoras para su posterior síntesis y evaluación experimental.

A pesar de sus ventajas, el diseño molecular asistido por computadora y el uso de modelos QSPR también presentan desafíos, como la necesidad de datos experimentales de alta calidad para el desarrollo y validación de los modelos, la incertidumbre inherente a las predicciones y la complejidad de algunas propiedades difíciles de modelar (Gani & Ng, 2015).

## CAPÍTULO II

### 2. MATERIALES Y MÉTODOS

#### 2.1. Generación de la base de datos

Para la generación de la base de datos se recopiló la información experimental del artículo científico titulado “*Characterization of volatile compounds in differently coloured Chenopodium quinoa seeds before and after cooking by headspace-gas chromatography-ion mobility spectrometry*” (Yang et al., 2021). De este artículo se extrajeron los índices de retención ( $I$ ) de 78 compuestos orgánicos volátiles (VOCs: volatile organic compounds) identificados en quinua de color blanco, rojo y negro; antes y después de la cocción. Las 11 muestras de quinua fueron obtenidas de Bolivia, Perú y China, y se almacenaron a  $-20^{\circ}\text{C}$ . Se analizaron sus propiedades colorimétricas  $L^*$ ,  $a^*$  y  $b^*$  por medio de espectrofotometría NS800 y se realizó análisis proximales de proteínas, grasa y carbohidratos de acuerdo (AOAC: Association of Official Analytical Chemists). Los VOCs fueron analizados por espectrometría de movilidad iónica (IMS: Ion Mobility Spectrometry) (FlavourSpec®, Gesellschaft für analytische Sensorsysteme mbH, Dortmund, Germany) con un cromatógrafo de gases Agilent 990 instalado internamente (Agilent Technologies, Palo Alto, CA, EE.UU.). El procesamiento quimiométrico de los datos del IMS se realizó empleando el software Laboratory Analytical Viewer (LAV), GC-IMS Library Search y los complementos Report y Gallery plot. El LAV se usó para visualizar el espectro de análisis y realizar un análisis cuantitativo. Los complementos Report Gallery permitieron comparar directamente las muestras utilizando las gráficas de vista superior y con las huellas dactilares moleculares.

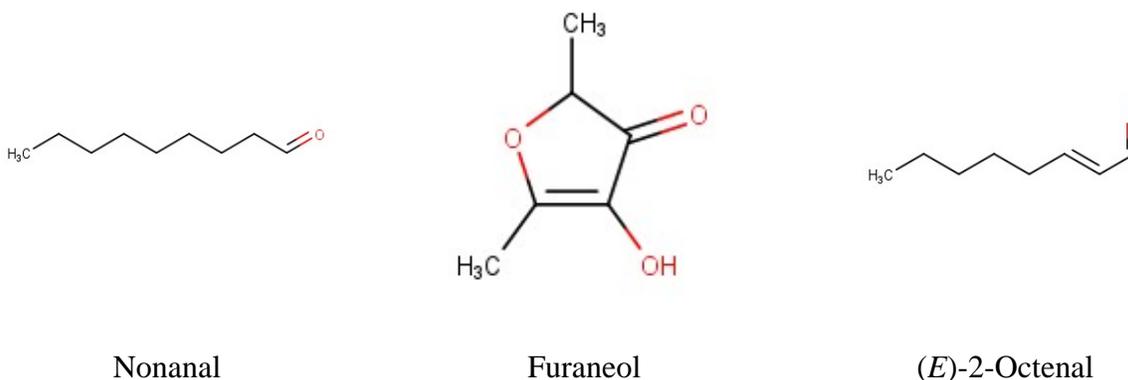
Para el análisis químico se pesaron 2 g de semilla de quinua, se lavaron con 10 mL de agua destilada 3 veces y luego se transfirieron a un vial de muestreo de vidrio de espacio de cabeza de 20 mL (Zhejiang Hamag Technology Co. Ltd., Ningbo, China). Las muestras se incubaron a  $4^{\circ}\text{C}$  con 2.6 mL de agua destilada durante 20 minutos en una unidad de muestreo (CTC Analytics AG, Zwingen, Suiza). Las muestras también se incubaron a  $100^{\circ}\text{C}$  durante 20 minutos en viales para simular el proceso de cocción (Zhang et al., 2018). Después de la incubación se inyectaron 500  $\mu\text{L}$  de gas del espacio de cabeza en el inyector del equipo programado a  $85^{\circ}\text{C}$  en modo sin división. Las muestras inyectadas circularon a través de la columna capilar FS-SE-54CB-1 (fase de sílice fundida-5% fenilo 95% dimetilpolisiloxilano-espesor de película 1  $\mu\text{m}$ , 15m  $\times$  0.23 mm, Agilent Technologies) en condiciones isotérmicas de  $60^{\circ}\text{C}$  utilizando nitrógeno (99.999% de pureza, Beijing Chen-Xin-Shun-Xing Gas Raw Material Sales Co.Ltd., Beijing, China) como gas portador en un flujo de 2mL/min durante 2 min, 10 mL/min durante 8min, 100 mL/min durante 10min y 150 mL/min durante 10min. La importancia de la columna capilar FS-SE-54CB-1 se debe a que gracias a su material de sílice fundida con una fase estacionaria de 5% de fenilo y 95% de dimetilpolisiloxano, otorga una ligera polaridad que permite una buena resolución para una amplia gama de compuestos volátiles. Asimismo, sus dimensiones proporcionan una alta eficiencia y

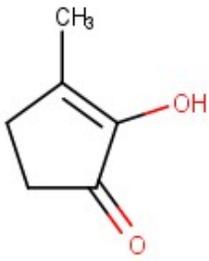
resolución en la separación de compuestos volátiles complejos, lo cual la convierte en una columna ideal para el análisis rápido de compuestos volátiles, debido a que permite tiempos de análisis relativamente cortos sin comprometer la resolución. Por último, reduce el ruido de fondo y mejora la sensibilidad del análisis al minimizar el sangrado de la columna gracias a su delgado espesor de película. (Agilent Technologies, 2009; Yang et al., 2021).

Después de la separación por GC, los analitos se ionizaron en la cámara de ionización de IMS en modo de iones positivos por una fuente de ionización por radiación- $\beta$ - $^3\text{H}$  con una actividad de 300 MBq (megabecquerel). Los iones se condujeron al tubo de deriva a 45°C en un voltaje lineal constante de 400V/cm y un flujo de nitrógeno de 150 mL/min. Los iones derivados se capturaron en 30 ms. El índice de retención (*I*) de los compuestos orgánicos volátiles se calculó usando n-cetonas C<sub>4</sub>-C<sub>9</sub> (Sinopharm Chemical Reagent Co., Ltd Beijing, China) como referencia externa. Los VOCs se identificaron comparando el *I* y el tiempo de deriva con la biblioteca GC-IMS. Todos los experimentos de análisis HS-GC-IMS se realizaron por duplicado (Yang et al., 2021). Posteriormente, para cada compuesto volátil se obtuvo el número de registro CAS (Voigt, 2003), la notación de la especificación de introducción lineal molecular simplificada (SMILES: Simplified Molecular Input Line Entry System) (Polanski & Gasteiger, 2016) y el número de identificación PubChem CID a partir de la quimioteca de acceso libre PubChem (Kim et al., 2023).

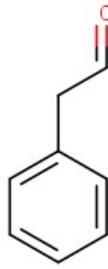
## 2.2. Filtrado de la base de datos

Durante el filtrado de la base de datos, se observó que existen 17 VOCs declarados como dímeros, es decir, compuestos formados por la unión de 2 moléculas más pequeñas idénticas mediante interacciones intermoleculares débiles (puentes de hidrógeno, fuerzas de Van der Waals o interacciones iónicas) (McMurry, Mondragón, & Pozo, 2008). Las moléculas dimerizadas se presentan en la Figura 1.

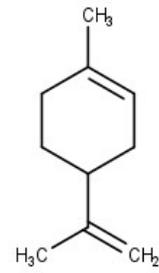




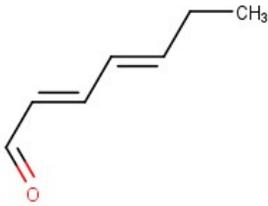
Cicloocteno



Fenilacetaldehído



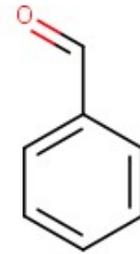
Limoneno



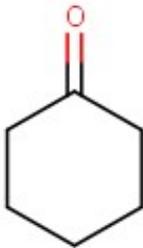
(*E,E*)-2,4-Heptadienal



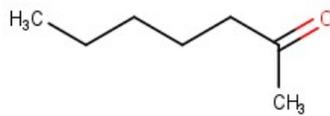
Octanal



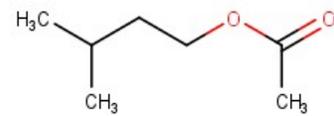
Benzaldehído



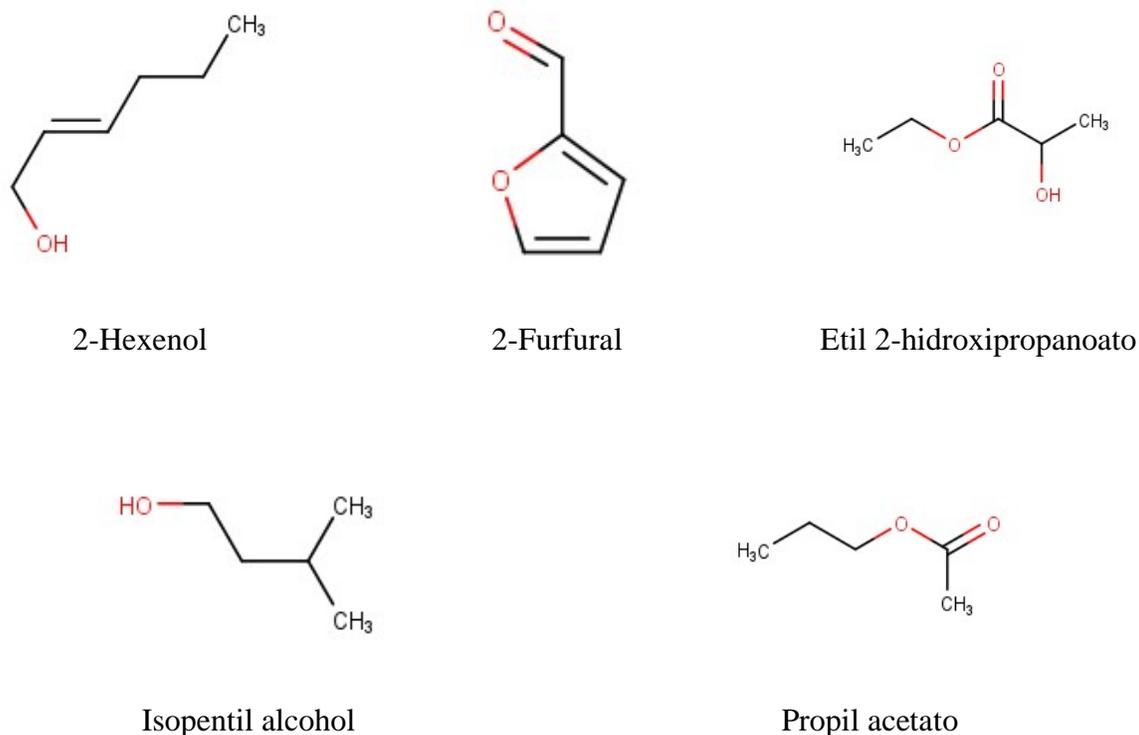
Ciclohexanona



2-Heptanona



Isoamil acetato



**Figura 1.** Moléculas que representan la presencia de dímeros

Los dímeros (Figura 1) se consideraron como moléculas no dimerizadas y se utilizó el *I* promedio para el modelado. De esta manera se obtuvieron 61 compuestos orgánicos volátiles para el desarrollo de relación cuantitativa estructura-propiedad.

### 2.3. Representación molecular y optimización de geometrías

Para el diseño de los VOCs se empleó el programa Marvin (ChemAxon Ltd., 2023), utilizando la representación de la molécula mediante el SMILES isomérico, para de esta manera obtener una representación molecular dependiente de la geometría estructural. Esta representación tridimensional permite una descripción de los detalles conformacionales de los VOCs. Posteriormente, se procedió a la optimización de las geometrías moleculares empleando el enfoque de mecánica molecular MM+ (Molecular Mechanic Plus), el cual es un campo de fuerzas empírico empleado en química computacional para calcular las propiedades estructurales y energéticas de moléculas. Este método se basa en la mecánica clásica y utiliza ecuaciones Newtonianas para modelar interacciones atómicas y es apropiado para el análisis de conformacional y optimización geométrica (Allinger, 1977). Posteriormente, se afinó la optimización de las estructuras de los compuestos utilizando el Atomic Charger Calculator II (Raček et al., 2020). Esta herramienta computacional permite realizar el cálculo rápido de cargas atómicas parciales utilizando más de 20 conjuntos de datos empíricos junto a parámetros provenientes de la literatura científica.

Además, esta aplicación permite el análisis y visualización de la distribución de cargas, lo cual resulta de gran utilidad para comprender propiedades electrónicas e interacciones intermoleculares de los sistemas estudiados. Para este efecto, se introdujeron las moléculas en la aplicación y se configuró su cálculo mediante el método de ecualización de electronegatividad EEM (Electronegativity Equalization Method), que consiste en igualar las electronegatividades atómicas en el estado de equilibrio molecular utilizando valores predefinidos de electronegatividad para cada tipo de átomo y resolviendo un sistema de ecuaciones para obtener las cargas parciales que satisfacen dicha condición (Mortier, Ghosh, & Shankar, 1986). El cálculo utilizó los parámetros Geidl 2015 (Cheminf\_b3lyp\_npa), el cual constituye un conjunto de cargas parciales atómicas derivadas de la química cuántica publicados. Específicamente, es un conjunto de parámetros desarrollado por aplicaciones quimioinformáticas, obtenido mediante el funcional de densidad B3LYP y el análisis de población natural npa (Natural Population Analysis) para la derivación de cargas parciales atómicas a partir de la densidad electrónica calculada (Geidl et al., 2015).

#### **2.4. Cálculo de los descriptores moleculares**

Los descriptores moleculares son variables numéricas que describen diversas propiedades físicas, químicas y topológicas de una molécula. Estas medidas se utilizan en estudios de las relaciones cuantitativas estructura-propiedad para predecir el índice de retención cromatográfico. Los descriptores moleculares permiten comprender la estructura y el comportamiento de las moléculas en la predicción del índice de retención cromatográfico (Roberto Todeschini & Consonni, 2009). Con las estructuras químicas de los compuestos orgánicos volátiles optimizadas, se calcularon diversos descriptores moleculares, así como, las llaves de los sistemas de acceso molecular (MACCS: Molecular ACCess System) (Duan, Dixon, Lowrie, & Sherman, 2010) en el programa alvaDesc (Alvascience, 2023a). Los descriptores moleculares se encuentran agrupados en 33 bloques lógicos que se presentan a continuación:

- Índices constitucionales (Constitutional indices)
- Descriptores de anillo (Ring descriptors)
- Índices topológicos (Topological indices)
- Recuento de caminos y trayectos (Walk and path counts)
- Índices de conectividad (Connectivity indices)
- Índices de información (Information indices)
- Descriptores basados en matrices 2D (2D matrix-based descriptors)
- Autocorrelaciones 2D (2D autocorrelations)
- Autovalores de Burden (Burden eigenvalues)
- Descriptores tipo P\_VSA (P\_VSA-like descriptors)
- Índices ETA (ETA: Extended Topochemical Atom indices)
- Índices de adyacencia de aristas (Edge adjacency indices)
- Descriptores geométricos (Geometrical descriptors)

- Descriptores basados en matrices 3D (3D matrix-based descriptors)
- Autocorrelaciones 3D (3D autocorrelations)
- Descriptores RDF (RDF, Radial Distribution Function descriptors)
- Descriptores 3D-MoRSE (3D-MoRSE descriptors)
- Descriptores WHIM (WHIM: Weighted Holistic Invariant Molecular descriptors)
- Descriptores GETAWAY (GETAWAY: Geometry, Topology and Atom Weights Assambly descriptors)
- Perfiles moleculares de Randic (Randic molecular profiles)
- Recuento de grupos funcionales (Functional group counts)
- Fragmentos centrados en el átomo (Atom-centred fragments)
- Índices de estado E centrados en el átomo (Atom-type E-state indices)
- Descriptores farmacóforos (Pharmacophore descriptors)
- Pares de átomos 2D (2D Atom Pairs)
- Pares de átomos 3D (3D Atom Pairs)
- Descriptores de carga (Charge descriptors)
- Propiedades moleculares (Molecular properties)
- Índices tipo fármaco (Drug-like indices)
- Descriptores CATS 3D (CATS: Chemically Advanced Template Search 3D descriptors)
- Descriptores WHALES (WHALES: Weighted Holistic Atom Localization and Entity Shape descriptors)
- Descriptores MDE (MDE: Molecular Descriptor Evaluator descriptors)
- Descriptores de quiralidad (Chirality descriptors)
- Sistema de acceso molecular (MACCS: Molecular Access System)

En el contexto del presente trabajo de titulación, no se consideraron los índices tipo fármaco debido a que la naturaleza y el propósito de dicho bloque no se alinean con los objetivos específicos planteados en la investigación. Estos índices están diseñados para medir qué tan similares son las moléculas bajo estudio en comparación con compuestos conocidos y establecidos como fármacos o agentes terapéuticos en términos de sus propiedades estructurales y fisicoquímicas (Mauri, 2020).

## **2.5. Aprendizaje no supervisado para la reducción de descriptores moleculares**

Para abordar la reducción de descriptores moleculares, se implementó un enfoque no supervisado con el objetivo de seleccionar un subconjunto representativo de descriptores que preserve la mayor cantidad de información y elimine la redundancia presente en el conjunto inicial de descriptores. Los conjuntos de descriptores moleculares pueden presentar ruido, redundancia y multicolinealidad, lo cual perjudica el desarrollo de modelos predictivos. El algoritmo V-WSP (Variable reduction based on Wootton, Sergent, Phan-Tan-Luu's algorithm), representa una alternativa ampliamente utilizada para este propósito. Este método selecciona un subconjunto

representativo de variables de tal forma que se encuentre una mínima correlación entre las mismas (umbral de 0.95) dentro de un espacio multidimensional definido, preservando así la mayor cantidad de información. El algoritmo V-WSP funciona de la siguiente manera (Ballabio et al., 2014):

1. Se calcula la matriz de correlación entre todos los descriptores del bloque correspondiente.
2. Los descriptores se ordenan de manera decreciente según la suma de los valores absolutos de sus coeficientes de correlación con los demás descriptores del mismo bloque.
3. Se selecciona el primer descriptor de la lista ordenada y se incluye en un subconjunto inicial de descriptores seleccionados.
4. Para cada descriptor restante en la lista ordenada, se calcula su correlación con los descriptores ya seleccionados. Si la correlación máxima es menor que un umbral predefinido, se agrega este descriptor al subconjunto seleccionado.
5. Se repite el paso anterior hasta que todos los descriptores hayan sido considerados.

De esta manera, se logra una reducción efectiva de la dimensionalidad del conjunto de descriptores preservando la información esencial y eliminando la redundancia inherente.

## **2.6. Desarrollo del modelo QSPR**

Una vez obtenido el conjunto reducido de descriptores moleculares más representativos e informativos, se procede al desarrollo del modelo QSPR (Quantitative Structure-Property Relationship) mediante técnicas del aprendizaje automático supervisado. En esta etapa, se empleó el conjunto de datos  $\mathbf{X}$ , la cual contiene los valores de los descriptores seleccionados para el conjunto de compuestos y el vector de datos de salida  $\mathbf{Y}$ , que corresponde a los índices de retención experimentales que se desean modelar. El enfoque consiste en aplicar la regresión de mínimos cuadrados ordinarios (OLS: Ordinary Least Squares), un método ampliamente aceptado y utilizado en la elaboración de modelos de regresión lineal. Esta técnica buscó encontrar los coeficientes del modelo lineal que minimicen la suma de los cuadrados de los residuos entre los valores predichos y los valores experimentales (Rencher & Schaalje, 2008). Sin embargo, debido a la alta dimensionalidad se acopló el método de reemplazo (RM: Replacement Method) para la selección supervisada de descriptores (Duchowicz, Castro, & Fernández, 2006). El RM parte de un modelo inicial (semilla), sobre el cual se calcula el error relativo de los coeficientes de cada descriptor. Seguidamente, el descriptor con el mayor error relativo se reemplaza por otro descriptor del conjunto que inicialmente no haya sido incluido. Este proceso iterativo continúa de tal forma que se optimiza la desviación estándar residual ( $s$ ) del modelo en cada paso. El algoritmo del RM explora diferentes combinaciones hasta converger en un modelo parsimonioso y predictivo (modelo óptimo). La combinación de la regresión OLS y RM para la selección de descriptores permitió encontrar un modelo QSPR óptimo con las siguientes características:

- Buena capacidad de ajuste a los datos de entrenamiento
- Alta capacidad predictiva para nuevos compuestos

Este enfoque del modelo QSPR mediante aprendizaje automático supervisado y selección de variables iterativa aseguró la obtención de un modelo robusto, interpretable y confiable. La calibración de los modelos y la selección supervisada de descriptores se aplicó en el programa MATLAB (The MathWorks Inc., 2022).

## 2.7. Validación del modelo QSPR

Para garantizar la confiabilidad y solidez del modelo, se implementaron diversas estrategias de validación interna y externa. En primera instancia, para la validación externa del modelo se aplicó el Método de Subconjuntos Balanceados (BSM: Balanced Subsets Method) (C. Rojas, Duchowicz, Tripaldi, & Diez, 2015a, 2015b) para dividir de manera racional la base de datos en subconjuntos de calibración y predicción. Este enfoque se basa en un análisis de conglomerados de  $k$ -medias, lo que permite una partición equilibrada de los datos en función de su similitud estructural y de la propiedad fisicoquímica. El modelo se construye utilizando el conjunto de calibración, mientras que el grupo de predicción se utiliza para evaluar la capacidad predictiva del modelo sobre datos, es decir, VOCs que no han sido utilizados para el desarrollo del modelo.

Por otro lado, también se aplicó la validación cruzada mediante protocolos que permitan evaluar la estabilidad del modelo frente a perturbaciones en el grupo de calibración. La técnica de validación cruzada de dejar-uno-afuera (LOO: Leave-One-Out) construye el modelo utilizando todos los compuestos excepto uno, y se evalúa la capacidad predictiva del modelo en el compuesto excluido. Este proceso se repite hasta que todos los compuestos de calibración hayan sido excluidos del conjunto de calibración (Arlot & Celisse, 2010). Asimismo, la técnica de dejar-varios-fuera (LMO: Leave-Many-Out) excluye un porcentaje de moléculas del grupo de calibración, el modelo se recalibra y se usa para predecir la propiedad del grupo de moléculas excluidas. En la técnica de dejar-varios-fuera se ha utilizado el método Monte Carlo, el cual implica, en cada iteración, seleccionar aleatoriamente un subconjunto de compuestos para ser excluido del conjunto de entrenamiento y utilizarlos para validar el modelo. La calidad del modelo se obtiene mediante comparación de las predicciones acumuladas de las diversas iteraciones (por ejemplo 1000) (Krakowska, Custers, Deconinck, & Daszykowski, 2016; Q. S. Xu, Liang, & Du, 2004). Otro enfoque LMO es el Bootstrap, en el que se construyen múltiples conjuntos de entrenamiento mediante muestreo aleatorio con reemplazo a partir del conjunto de calibración. Los compuestos no seleccionados constituyen el conjunto de validación. También el método Bootstrap se repite tantas veces como sea posible, de tal manera que se estime el valor promedio de las predicciones (Efron, 1979, 1982, 1987). El último método de validación LMO es el de Ventanas Venecianas, en el que los datos se dividen en varios subconjuntos consecutivos denominados ventanas venecianas. A continuación, se selecciona el primer elemento de cada grupo para formar el set de validación, el modelo se construye utilizando el conjunto de entrenamiento y se evalúa su capacidad predictiva sobre el conjunto de validación. Este proceso se repite hasta que se haya incluido en el set de validación el último compuesto de cada grupo (Ballabio & Consonni, 2013).

Por otra parte, se aplicó la Aleatorización-Y para descartar la presencia de correlación casual en el modelo. Esta técnica consiste en permutaciones aleatorias de la propiedad a modelar de tal forma que no corresponda a sus respectivos descriptores moleculares. Se construyen cientos de modelos con estos datos permutados y se calculan sus métricas de desempeño, las cuales deben ser significativamente inferiores a las del modelo original, ya que no debe existir correlaciones verdaderas en los datos aleatorios. Si el modelo original supera significativamente a los modelos aleatorios, se concluye que ha capturado correlaciones verdaderas y no casuales, evaluando así su significancia estadística (Eriksson, Johansson, & Wold, 1997; Lindgren, Hansen, Karcher, Sjöström, & Eriksson, 1996).

La validación del modelo QSPR se realizó en el programa alvaModel (Alvascience, 2023b), una herramienta especializada en el desarrollo y validación de modelos QSPR.

## **2.8. Análisis de la contribución de los descriptores moleculares**

Después de validar el modelo se evaluó la importancia relativa de cada descriptor en la predicción de la propiedad modelada. Para ello, se analizaron los coeficientes estandarizados del modelo de regresión lineal de mínimos cuadrados ordinarios. Estos coeficientes estandarizados proporcionaron una medida de la influencia relativa de cada descriptor molecular sobre la variable respuesta, lo que permitió comparar directamente la contribución de los diferentes descriptores en el modelo (independientemente de sus escalas originales). Un coeficiente estandarizado con un valor absoluto alto para un descriptor particular indicaba una mayor importancia de ese descriptor en la predicción de la propiedad modelada, mientras que un valor absoluto bajo sugiere una contribución menor. Además, el signo del coeficiente estandarizado aporta información valiosa sobre la naturaleza de la relación entre el descriptor y la propiedad experimental. Un coeficiente positivo implica un efecto sinérgico, es decir, un aumento en el valor del descriptor conlleva a un incremento en la propiedad predicha; mientras que un coeficiente negativo indica un efecto antagónico (el aumento del valor del descriptor resulta en una disminución de la propiedad predicha).

## **2.9. Mecanismo de acción de los descriptores moleculares**

Después de evaluar la contribución relativa de cada descriptor molecular incluido en el modelo QSPR, se procedió a analizar el mecanismo de acción, es decir, analizar cómo influyen los mismos en la predicción del índice de retención ( $I$ ). Comprender el mecanismo de acción brinda mayor confianza al modelo, especialmente para fines regulatorios, y proporciona una interpretación más profunda de las relaciones estructura-propiedad identificadas. El mecanismo de acción se obtiene mediante la definición de cada descriptor molecular y la explicación de cómo se relaciona con la propiedad en estudio, ya sea de forma sinérgica (incrementando el  $I$ ) o antagónica (disminuyendo el  $I$ ). Sin embargo, no siempre es posible obtener una interpretación completa debido a la complejidad matemática involucrada en la definición de algunos descriptores moleculares.

## 2.10. Análisis del dominio de aplicabilidad

El dominio de aplicabilidad (AD: Applicability Domain) es un aspecto fundamental en los modelos QSPR, debido a que define una región química teórica dentro de la cual las nuevas predicciones son confiables. Dada la naturaleza específica del modelo, se analiza el AD mediante el valor de influencia ( $h$ ) o leverage. Este enfoque cuantifica la distancia de un compuesto del grupo de predicción con respecto al centroide del conjunto de calibración en el espacio multidimensional de los descriptores. En el conjunto de calibración, estos valores oscilan entre 0 y 1, en cuanto al grupo de predicción solo se tiene valores por debajo de 0. Para el AD se analiza el valor umbral crítico ( $h^*$ ). Si el valor de influencia  $h$  de un compuesto es alto, indica que se encuentra lejos del centro del modelo, lo que sugiere que su predicción es poco confiable o representa una extrapolación del modelo; mientras que valores de influencia bajos indican que los índices de retención de esas moléculas son confiables (C. Rojas & Duchowicz, 2021).

## CAPÍTULO III

### 3. RESULTADOS

#### 3.1. Generación de la base de datos

El estudio comenzó con una base de datos de 78 compuestos identificados en muestras de semillas de quinua, junto con sus índices de retención ( $I$ ). Durante el proceso de filtración y curación de los datos, se identificaron 17 dímeros, los cuales se fusionaron y se utilizó el índice de retención promedio como valor experimental. Después de este proceso, se obtuvo una base de datos curada con 61 compuestos orgánicos volátiles (VOCs), para los cuales se obtuvo el número de registro CAS, notaciones SMILES y números PubChem CID. A continuación, se optimizaron las geometrías de los compuestos utilizando los campos de fuerza de la MM+ y se calcularon las cargas atómicas parciales mediante el método EEM y el conjunto de parámetros Geidl 2015 en la aplicación Atomic Charger Calculator II. Los archivos de salida de esta aplicación se utilizaron posteriormente para el cálculo de descriptores moleculares con el fin de desarrollar una relación cuantitativa estructura-propiedad.

#### 3.2. Cálculo de los descriptores moleculares

Después de optimizar las estructuras de los 61 compuestos orgánicos volátiles (VOCs), se calcularon 5633 descriptores y 166 claves estructurales MACCS utilizando el programa alvaDesc (Alvascience, 2023a). Con el objetivo de reducir la dimensionalidad del espacio de descriptores moleculares y eliminar información redundante o poco informativa, se implementó un enfoque de reducción no supervisada en dos etapas. En la primera etapa, se excluyeron 2578 descriptores con valores constantes, 64 descriptores con valores casi constantes y 43 descriptores que presentaban al menos un valor faltante. En la segunda etapa, se aplicó el método de reducción no supervisada V-WSP con un umbral de correlación  $thr = 0.95$ . Este método permitió identificar y eliminar descriptores altamente correlacionados, evitando así la redundancia de información y mejorando la calidad de los datos.

#### 3.3. Desarrollo del modelo QSPR

Para el desarrollo del modelo, la base de datos se dividió en dos conjuntos: calibración y predicción, utilizando el método de subconjuntos balanceados (BSM). Este método asignó 49 compuestos al conjunto de calibración y 12 moléculas al conjunto de predicción. El conjunto de calibración se empleó para calibrar y optimizar el modelo, mientras que el conjunto de predicción se utilizó para evaluar el poder predictivo del modelo desarrollado. Durante el desarrollo del modelo, se aplicó el método de reemplazo RM (Replacement Method) en conjunto con la regresión de mínimos cuadrados ordinarios para seleccionar un subconjunto óptimo de descriptores

moleculares, es decir, los descriptores que exhibieran la mayor correlación con el índice de retención (I) de los compuestos orgánicos volátiles. La implementación del método RM se realizó en dos etapas. Primero, se aplicó la selección supervisada en cada bloque de descriptores moleculares. El número de descriptores seleccionados se presentan en la Tabla 1.

<b>Bloque</b>	<b>Familia de Descriptor Molecular</b>	<b>Seleccionado</b>
B1	Índices constitucionales	19
B2	Descriptores de anillo	13
B3	Índices topológicos	26
B4	Recuentos de caminatas y rutas	22
B5	Índices de conectividad	18
B6	Índices de información	20
B7	Descriptores basados en matrices 2D	27
B8	Autocorrelaciones 2D	21
B9	Autovalores de Burden	25
B10	Descriptores tipo P_VSA	23
B11	Índices ETA	21
B12	Índices de adyacencia de aristas	34
B13	Descriptores geométricos	20
B14	Descriptores basados en matrices 3D	25
B15	Autocorrelaciones 3D	22
B16	Descriptores RDF	23
B17	Descriptores 3D-MoRSE	19
B18	Descriptores WHIM	26
B19	Descriptores GETAWAY	29
B20	Perfiles moleculares Randic	23
B21	Recuento de grupos funcionales	12
B22	Fragmentos centrados en átomos	14
B23	Índices de estado E centrados en átomos	18
B24	Descriptores en farmacóforos	14
B25	Pares de átomos 2D	14
B26	Pares de átomos 3D	1
B27	Descriptores de carga	13

B28	Propiedades moleculares	15
B29	Descriptores CATS 3D	13
B30	Descriptores WHALES	14
B31	Descriptores MDE	11
B32	Descriptores de quiralidad	26
MACCS	Sistema de acceso molecular	14
	<b>TOTAL</b>	<b>635</b>

**Tabla 1.** Cantidad de descriptores moleculares obtenidos por medio del método de reemplazo

En una segunda etapa, los mejores descriptores de cada bloque se combinaron (conjunto de 635) y se aplicó nuevamente el método RM para encontrar el modelo óptimo. La selección de los modelos con los mejores descriptores moleculares en ambas etapas se realizó de manera que se minimice la desviación estándar residual ( $s$ ), se maximice el coeficiente de determinación ( $R^2$ ) y minimice correlación máxima entre descriptores ( $R^2_{ijmax}$ ). Como resultado, el modelo QSPR final está constituido por cuatro descriptores moleculares, como se muestra en la Ecuación 1.

$$I = 251.0692 + 100.7720 \text{ SpMAD}_B(s) + 417.7484 \text{ MATS1p} \\ + 81.4098 \text{ Chi0}_{EA}(ri) + 496.9204 \text{ H2m}$$

#### Ecuación 1

Los descriptores seleccionados para el modelo QSPR son:

- **SpMAD<sub>B</sub>(s)**: Desviación absoluta media espectral de la matriz de carga ponderada por el estado intrínseco I.
- **MATS1p**: Autocorrelación de Moran a desplazamiento 1 ponderado por la polarización.
- **Chi0<sub>EA</sub>(ri)**: Índice tipo de conectividad de orden 0 de la matriz de adyacencia ponderado por la integral de resonancia.
- **H2m**: Autocorrelación H a desplazamiento 2 ponderado por la masa.

### 3.4. Validación del modelo QSPR

Posteriormente, se procedió a validar exhaustivamente el modelo mediante el coeficiente de determinación ( $R^2$ ) y el error cuadrático medio ( $RMSE$ ). Durante la validación cruzada, se aplicaron las técnicas de dejar-uno-fuera, Ventanas Venecianas con 5 grupos, Monte Carlo con 20% de exclusión y 1000 iteraciones y Bootstrap con 1000 iteraciones de remuestreo. Por otra parte, la capacidad predictiva se obtuvo al predecir los I de los VOCs del grupo de predicción. Todos estos enfoques de validación cruzada revelaron resultados consistentes y satisfactorios,

corroborando la solidez y confiabilidad del modelo QSPR obtenido en la predicción de los índices de retención (Tabla 2).

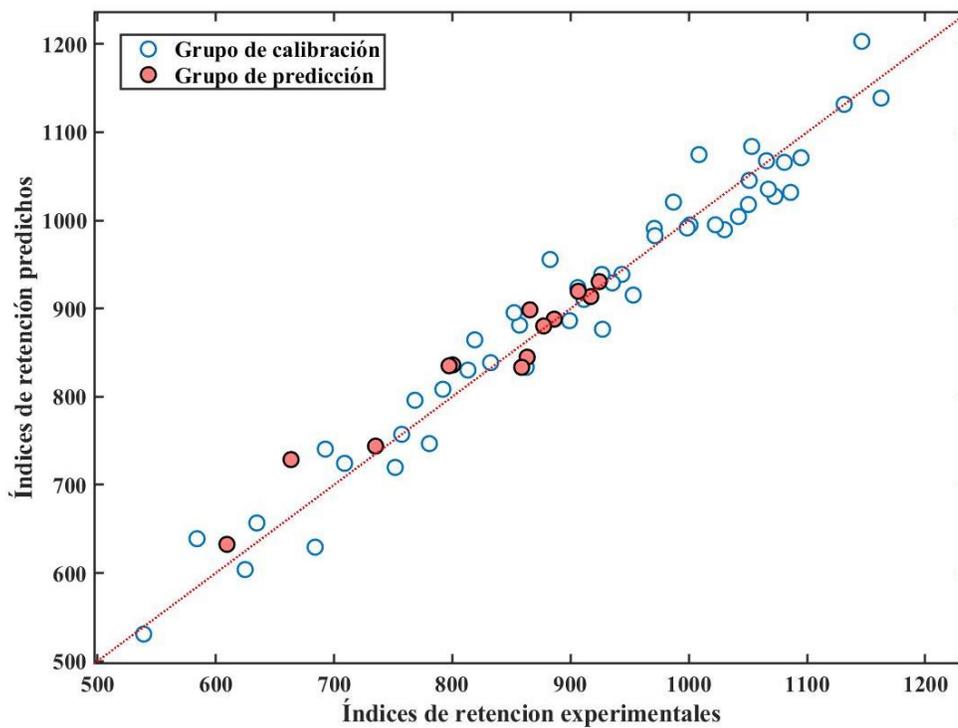
	$R^2$	$RMSE$
Calibración	0.956	32.590
Validación cruzada		
<i>Dejar-uno-fuera</i>	0.941	37.883
<i>Ventanas venecianas</i>	0.942	37.556
<i>Monte Carlo</i>	0.940	38.082
<i>Bootstrap</i>	0.934	39.914
Predicción	0.969	27.521

**Tabla 2.** Resultados obtenidos para el modelo QSPR en los grupos de calibración, validación cruzada y predicción

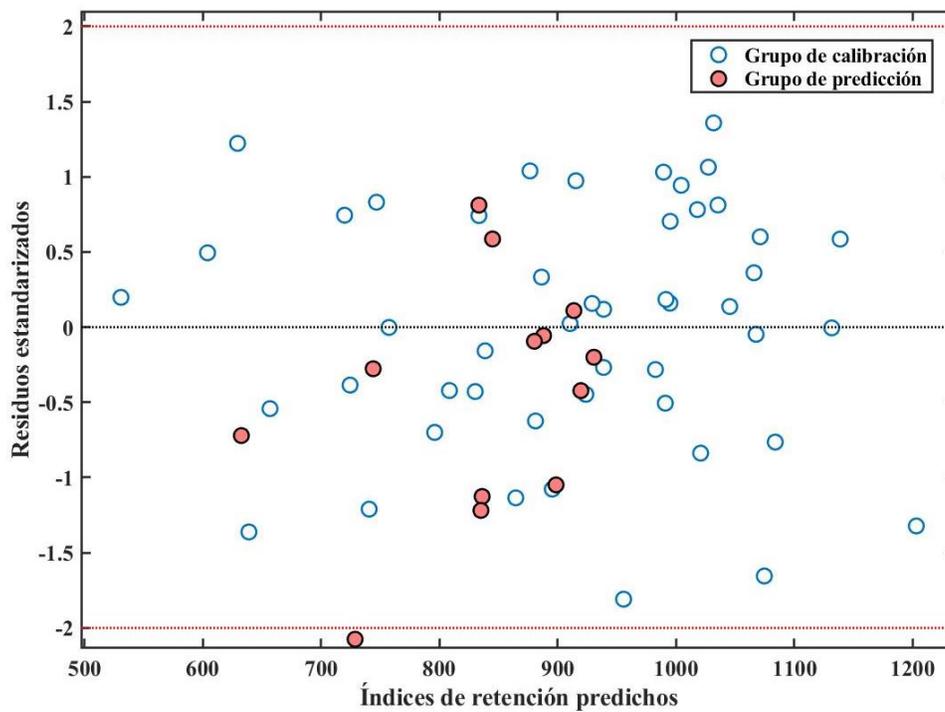
Los resultados indican que el modelo *in silico* desarrollado tiene una capacidad de ajuste del 95.6%. Por otro lado, los estadísticos de validación cruzada indican buena estabilidad del modelo frente a perturbaciones internas del mismo; mientras que la aleatorización-Y ( $R^2 = 0.0821$  y  $RMSE = 157.1759$ ) indican que el modelo no presenta correlación casual. Finalmente, al aplicar el modelo para la predicción del *I* para los 12 VOCs del grupo de predicción se evidencia una capacidad predictiva del 96.9%. Todos estos resultados garantizan la validez y robustez del modelo para la predicción precisa de los índices de retención de los compuestos orgánicos volátiles detectados mediante HS-GC-MS.

La Figura 2 muestra una excelente correlación entre los índices de retención experimentales y los valores predichos por el modelo QSPR representado por la Ecuación 1. Esta correlación se observa tanto para el conjunto de calibración como para el conjunto de predicción. Los datos se distribuyen cerca de la línea de ajuste perfecto, mostrando una buena tendencia lineal. Esta Figura respalda los buenos parámetros estadísticos presentados en la Tabla 2.

Por otro lado, la Figura 3 presenta muestra el comportamiento de los residuos estandarizados frente a los índices de retención predichos por el modelo QSPR. Se puede observar que los residuos estandarizados se distribuyen de manera aleatoria alrededor de la línea cero, sin exhibir patrones sistemáticos o tendencias evidentes. Esta distribución aleatoria de los residuos es una característica deseable, ya que indica que el modelo de OLS utilizado es apropiado para modelar los índices de retención. Además, es indispensable destacar que en este gráfico no se identificaron moléculas cuyos  $RMSE$  se encuentren fuera del límite de  $\pm 2$  veces el error cuadrático medio del modelo.



**Figura 2.** Índices de retención experimental vs. los índices de retención predichos de los VOCs en semillas de quinua medidos mediante HC-GC-MS



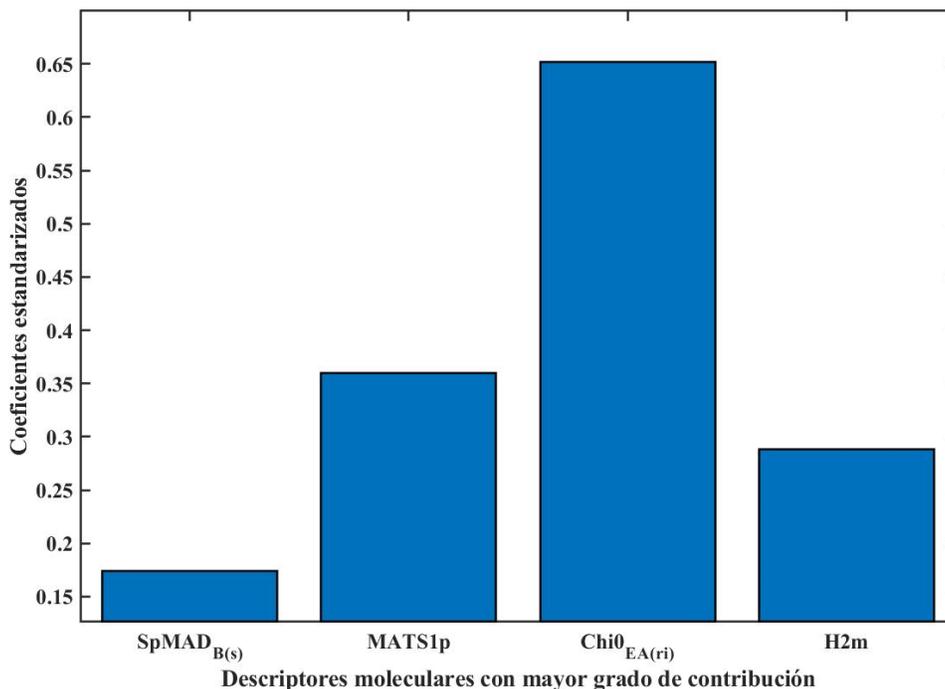
**Figura 3.** Grafica de dispersión de los residuos estandarizados vs. los índices de predicción predichos.

### 3.5. Análisis del grado de contribución de los descriptores moleculares

Para analizar la contribución de los descriptores, se calcularon los coeficientes de regresión estandarizados de los mismos. Estos coeficientes estandarizados permiten evaluar y comparar la importancia relativa de los descriptores en el modelo, donde un valor absoluto más alto indica una mayor contribución. Los valores obtenidos para los coeficientes estandarizados de los cuatro descriptores son:

$$Chi0_{EA}(ri) (0.6518) > MATS1p (0.3597) > H2m (0.2880) > SpMAD_{B(s)}(0.1739)$$

Cabe destacar que todos los descriptores presentan valores positivos de sus coeficientes en el modelo (Ecuación 1), lo que indica un efecto sinérgico en el modelo, es decir, que contribuyen de manera positiva a la predicción de los índices de retención (Figura 4). El descriptor con el coeficiente estandarizado de mayor magnitud es  $Chi0_{EA}(ri)$ , lo que sugiere su predominante influencia en el modelo desarrollado. Por otro lado, el descriptor con el coeficiente estandarizado de menor magnitud es  $SpMAD_{B(s)}$ . Si bien  $SpMAD_{B(s)}$  tiene una contribución menos significativa en el modelo obtenido, es importante destacar que todos los descriptores seleccionados aportan información relevante y complementaria para la predicción del  $I$  de los compuestos orgánicos volátiles.



**Figura 4.** Grado de contribución de los descriptores moleculares para los VOCs presentes en muestras de quinua

El descriptor *Chi0\_EA(ri)*, que representa el índice tipo de conectividad de orden 0 de la matriz de adyacencia ponderado por la integral de resonancia. Pertenece a la familia de los índices de adyacencia de arista, que son descriptores moleculares topológicos que cuantifican información sobre las relaciones de adyacencia entre los átomos o vértices en una molécula, considerando no solo las conexiones directas, sino también las conexiones a través de otros átomos intermedios. Estos índices se basan en la idea de que las propiedades fisicoquímicas de las moléculas están estrechamente relacionadas con su estructura molecular y las interacciones entre sus átomos constituyentes. Al capturar información sobre las adyacencias de arista, que van más allá de las simples conexiones de enlace, estos descriptores logran codificar información estructural más detallada y compleja. Los índices de adyacencia de arista son descriptores moleculares que se obtienen a partir de las matrices de adyacencia de arista de los grafos moleculares. Estas matrices representan las conexiones entre todos los átomos (vértices) mediante caminos de longitud específica dentro de la molécula. Estos índices cuantifican la distribución de dichos caminos en la estructura molecular, lo cual permite distinguir entre isómeros estructurales y capturar características estructurales significativas de las moléculas (R Todeschini & Consonni, 2010).

Por otra parte, la autocorrelación de Moran a desplazamiento 1 ponderado por la polarización. (*MATSp*) es un descriptor que pertenece al grupo de las autocorrelaciones 2D. Son una clase de descriptores moleculares que describen la distribución espacial de ciertas propiedades atómicas en una molécula, considerando únicamente la estructura molecular 2D (grafo molecular). Se basan en la función de autocorrelación, que mide la correlación entre valores de una propiedad en función de la distancia topológica entre los átomos. Específicamente, el descriptor de autocorrelación de Moreau-Broto (ATS) es el más conocido y se calcula sumando los productos de una propiedad atómica (como masa, polarizabilidad, etc.) para todos los pares de átomos a una determinada distancia topológica *k* en el grafo molecular. Esto se repite para todas las distancias topológicas presentes, generando un conjunto de valores ATS\_k que describen cómo esa propiedad se distribuye a lo largo de la estructura topológica de la molécula. Estos descriptores son invariantes a las traslaciones y rotaciones, y capturan información sobre la distribución atómica de las propiedades fisicoquímicas en la estructura 2D de la molécula (R Todeschini & Consonni, 2010).

La autocorrelación H a desplazamiento 2 ponderado por la masa (*H2m*), es un descriptor perteneciente al grupo de los descriptores GETAWAY (GEometry, Topology and Atom-Weights Assembly), los cuales son una clase de descriptores moleculares que codifican información sobre la geometría 3D, la topología y las propiedades atómicas de una molécula. Se basan en la matriz de influencia molecular (**MIM: Molecular Influence Matrix**), que es una representación matricial de la molécula obtenida a partir de las coordenadas cartesianas de sus átomos, la cual se denota por la siguiente ecuación:

$$H = \mathbf{M} \times (\mathbf{M}^T \times \mathbf{M})^{-1} \times \mathbf{M}^T$$

donde **M** representa la matriz molecular en las coordenadas cartesianas *x*, *y*, y *z*, de todos los átomos de las moléculas en una conformación específica. Las coordenadas se obtienen con respecto al centro geométrico para evitar invarianzas traslacionales. La diagonal de la matriz

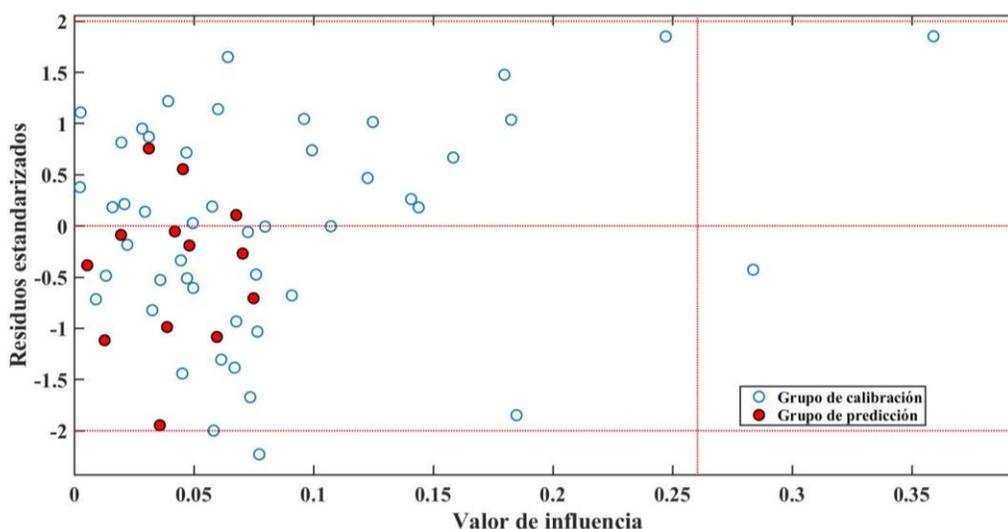
molecular,  $h_{ij}$  contiene los valores de influencia (leverages) que capturan la influencia de cada átomo en la forma general de la molécula. Los elementos fuera de la diagonal representan el grado de accesibilidad e interacción entre pares de átomos. A partir de la **MIM** y la matriz de distancias geométricas, se deriva la matriz de influencia/distancia (R Todeschini & Consonni, 2010).

Por último, **SpMAD<sub>B</sub>(s)**, desviación absoluta media espectral de la matriz de carga ponderada por el estado intrínseco **I**, pertenece a los descriptores basados en las matrices 2D, que son una clase importante de descriptores moleculares que se derivan directamente de las representaciones matriciales de la estructura molecular a partir de un grafo bidimensional. Estas matrices codifican información puramente topológica sobre la conectividad y distancias entre los átomos en la molécula. Los descriptores basados en matrices 2D codifican información sobre la conectividad atómica, distribución de fragmentos, simetría molecular, efectos estéricos, propiedades másicas y otras propiedades estructurales relevantes de la topología 2D de la molécula (R Todeschini & Consonni, 2010).

### 3.6. Análisis del 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). En este estudio se estableció un valor umbral de influencia crítico  $h^* = 0.260$  (Figura 5). Se observa que no existen moléculas del grupo de predicción que sobrepasen el umbral  $h^* = 0.260$ , lo que indica que todas las predicciones del grupo de predicción son interpolaciones del modelo, y por lo tanto confiables.

**Figura 5.** Gráfico de residuos normalizados frente a valores de influencia para delimitar el espacio de aplicación válido del modelo QSPR.



## CONCLUSIONES

El presente trabajo de titulación ha demostrado el uso de los enfoques computacionales, particularmente los modelos QSPR, en la predicción de los índices de retención cromatográfico de compuestos orgánicos volátiles identificados en semillas de quinua. A través de un enfoque sistemático y riguroso, se ha logrado desarrollar una relación cuantitativa estructura-propiedad capaz de capturar las complejas relaciones entre la estructura química de los compuestos y los índices de retención. El proceso de selección de descriptores moleculares basado en técnicas quimiométricas y algoritmos del aprendizaje automático, ha permitido identificar cuatro descriptores óptimos entre miles de ellos para modelar y predecir de forma apropiada los índices de retención de los VOCs. En efecto el modelo presenta una calidad descriptiva de los datos de 95.6% y una capacidad predictiva de 96.9%. Además, la implementación de diversas estrategias de validación interna ha garantizado la solidez estadística del modelo y su aplicabilidad en un amplio dominio químico. Este trabajo representa un avance fundamental en el campo de la química computacional de los alimentos, debido a que la metodología puede ser fácilmente transferible a otros compuestos orgánicos volátiles y otras matrices alimentarias con el fin de explorar de forma racional el vasto espacio químico. Finalmente, este modelo ha sido desarrollado considerando los cinco principios de modelado QSAR/QSPR definidos por la Organización para la Cooperación Económica y el Desarrollo (OECD).

## REFERENCIAS

- Agilent Technologies, I. (2009). Agilent J&W Ultra Inert GC Column Selection Guide. In I. Agilent Technologies (Ed.). Santa Clara, CA.
- Allinger, N. L. (1977). Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms. *Journal of the American Chemical Society*, 99(25), 8127-8134.
- Almaguer, C., Kollmannsberger, H., Gastl, M., & Becker, T. (2023). Characterization of the aroma profile of quinoa (*Chenopodium quinoa* Willd.) and assessment of the impact of malting on the odor-active volatile composition. *Journal of the Science of Food and Agriculture*, 103(5), 2283-2294.
- Alvascience. (2023a). alvaDesc (Software for Molecular Descriptors Calculation) version 2.0.12, <https://www.alvascience.com>.
- Alvascience. (2023b). alvaMolecule (Software to View and Prepare Chemical Datasets) version 1.0.4, <https://www.alvascience.com>.
- Antolín, B., Acitores, M., Jiménez, J. J., & Vega, M. OPTIMIZATION AND VALIDATION OF GC/MS METHODS FOR TRACE ANALYSIS OF VOCs AND PESTICIDES IN WATER.
- Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection.
- Arzapalo Quinto, D., Huamán Córdor, K., Quispe Solano, M., & Espinoza Silva, C. (2015). Extracción y caracterización del almidón de tres variedades de quinua (*Chenopodium quinoa* Willd) negra collana, pasankalla roja y blanca junín. *Revista de la Sociedad Química del Perú*, 81(1), 44-54.
- Ballabio, D., & Consonni, V. (2013). Classification tools in chemistry. Part 1: linear models. PLS-DA. *Analytical Methods*, 5(16), 3790-3798.
- 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.
- Balladares, M. C., Martínez, F. M., & Rojas, N. D. (2015). *Determinación de tocoferoles totales en semillas de quinua sometidas a diferentes tratamientos de cocción*.
- Barrientos, E., Carevic, F., & Delatorre, J. (2017). La sustentabilidad del altiplano sur de Bolivia y su relación con la ampliación de superficies de cultivo de quinua. *Idesia (Arica)*, 35(2), 7-15.
- Belitz, H.-D., Grosch, W., & Schieberle, P. (2008). *Food chemistry*: Springer Science & Business Media.
- Bernuy, N., Villanueva, M. E., Suárez, S., & Vilchez, C. (2018). Influencia del consumo de quinua (*Chenopodium quinoa* Willd.) sobre la acumulación del tejido adiposo y actividad antioxidante en tejidos de ratas obesas. *Archivos Latinoamericanos de Nutrición (ALAN)*, 68(2), 122-131.
- Cajka, T., Hajslova, J., Pudil, F., & Ridellova, K. (2009). Traceability of honey origin based on volatiles pattern processing by artificial neural networks. *Journal of Chromatography A*, 1216(9), 1458-1462.
- Campos-Rodriguez, J., Acosta-Coral, K., & Paucar-Menacho, L. M. (2022). Quinoa (*Chenopodium quinoa*): Composición nutricional y Componentes bioactivos del grano y la hoja, e impacto del tratamiento térmico y de la germinación. *Scientia Agropecuaria*, 13(3), 209-220.

- Chagua Lazo, G. S., & Palomino Villazan, L. L. (2014). Estudio comparativo de la capacidad antioxidante y compuestos fenólicos en quinua (*Chenopodium quinoa*) expandida de tres variedades provenientes del departamento de Junín.
- ChemAxon Ltd. (2023). MarvinSketch version 23.7, <http://www.chemaxon.com>.
- Chinnici, F., Duran Guerrero, E., Sonni, F., Natali, N., Natera Marin, R., & Riponi, C. (2009). Gas Chromatography– Mass Spectrometry (GC– MS) Characterization of Volatile Compounds in Quality Vinegars with Protected European Geographical Indication. *Journal of Agricultural and Food Chemistry*, 57(11), 4784-4792.
- Cozzolino, D. (2023). Cereals, Pseudocereals, Flour, and Bakery Products. In *Emerging Food Authentication Methodologies Using GC/MS* (pp. 47-63): Springer.
- Cserhádi, T. (2010). *Chromatography of aroma compounds and fragrances*: Springer Science & Business Media.
- Cusack, D. (1984). Quinoa-grain of the Incas.
- Dearden, J., & Worth, A. (2007). In silico prediction of physicochemical properties. *JRC Scientific and Technical Reports, EUR, 23051*, 1-68.
- Delgado, A. I., Palacios, J. H., & Betancourt, C. (2009). Evaluación de 16 genotipos de quinua dulce (*Chenopodium quinoa* Willd.) en el municipio de Iles, Nariño (Colombia). *Agronomía Colombiana*, 27(2), 159-167.
- Duan, J., Dixon, S. L., Lowrie, J. F., & Sherman, W. (2010). Analysis and comparison of 2D fingerprints: insights into database screening performance using eight fingerprint methods. *Journal of Molecular Graphics and Modelling*, 29(2), 157-170.
- 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 Commun. Math. Comput. Chem*, 55, 179-192.
- Efron, B. (1979). Bootstrap methods: another look at the jackknife annals of statistics 7: 1–26. *View Article PubMed/NCBI Google Scholar*, 24.
- 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.
- Eriksson, L., Johansson, E., & Wold, S. (1997). *QSAR model validation*. Paper presented at the Quantitative Structure-Activity Relationships in Environmental Sciences VII. Proceedings of the 7th International Workshop on QSAR in Environmental Sciences 24-28 June 1997.
- Feng, D., Wang, J., Ji, X.-J., Min, W.-X., & Yan, W.-J. (2021). Analysis of volatile organic compounds by HS-GC-IMS in powdered yak milk processed under different sterilization conditions. *Journal of Food Quality*, 2021, 1-10.
- Gani, R., & Ng, K. M. (2015). Product design–molecules, devices, functional products, and formulated products. *Computers & Chemical Engineering*, 81, 70-79.
- Geidl, S., Bouchal, T., Raček, T., Svobodová Vařeková, R., Hejret, V., Křenek, A., . . . Koča, J. (2015). High-quality and universal empirical atomic charges for cheminformatics applications. *Journal of cheminformatics*, 7, 1-10.
- Gramatica, P. (2007). Principles of QSAR models validation: internal and external. *QSAR & Combinatorial Science*, 26(5), 694-701.
- Guichard, E. (2015). Interaction of aroma compounds with food matrices. In *Flavour development, analysis and perception in food and beverages* (pp. 273-295): Elsevier.
- Gutiérrez Bouzán, M. C., & Droguet, M. (2002). La cromatografía de gases y la espectrometría de masas: identificación de compuestos causantes de mal olor. *Gutiérrez, MC; Droguet, M.*

- La cromatografía de gases y la espectrometría de masas: identificación de compuestos causantes de mal olor". Boletín Intexter, juliol 2002, núm. 122, p. 35-41.*
- Han, J., Fu, C., Wan, Y., Guo, J., Wei, C., & Liu, W. (2023). Characterizing the Volatile Profile and Flavor of Four Kinds of Hotpot Seasonings in China Using HS-GC-MS, HS-GC-IMS, and Sensory Evaluation. *European Journal of Lipid Science and Technology*, 125(8), 2200197.
- Hwang, J. B., Lee, S., Yeum, J., Kim, M., Choi, J. C., Park, S.-J., & Kim, J. (2019). HS-GC/MS method development and exposure assessment of volatile organic compounds from food packaging into food simulants. *Food Additives & Contaminants: Part A*, 36(10), 1574-1583.
- Katritzky, A. R., Fara, D. C., Yang, H., Tämm, K., Tamm, T., & Karelson, M. (2004). Quantitative measures of solvent polarity. *Chemical reviews*, 104(1), 175-198.
- Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., . . . Yu, B. (2023). PubChem 2023 update. *Nucleic acids research*, 51(D1), D1373-D1380.
- Koek, M. M., Jellema, R. H., van der Greef, J., Tas, A. C., & Hankemeier, T. (2011). Quantitative metabolomics based on gas chromatography mass spectrometry: status and perspectives. *Metabolomics*, 7, 307-328.
- 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.
- Laus, M. N., Gagliardi, A., Soccio, M., Flagella, Z., & Pastore, D. (2012). Antioxidant activity of free and bound compounds in quinoa (*Chenopodium quinoa* Willd.) seeds in comparison with durum wheat and emmer. *Journal of Food Science*, 77(11), C1150-C1155.
- Lindgren, F., Hansen, B., Karcher, W., Sjöström, M., & Eriksson, L. (1996). Model validation by permutation tests: applications to variable selection. *Journal of chemometrics*, 10(5-6), 521-532.
- Luisetti, J., Lucero, H., & Ciappini, M. C. (2020). Optimization of the extraction of antioxidant compounds from quinoa (*Chenopodium quinoa* Willd.). *Agronomía Colombiana*, 38(3), 436-441.
- Mamani, E. R., Bonifacio, A., Taboada, G. C., & Angulo, F. R. (2022). Determinación de la pureza y calidad de grano de quinua (*Chenopodium quinoa* Willd.) dulce y amarga de variedades comerciales, en Viacha-La Paz. *CIBUM SCIENTIA*, 1(1), 7-15.
- Mauri, A. (2020). alvaDesc: A tool to calculate and analyze molecular descriptors and fingerprints. *Ecotoxicological QSARs*, 801-820.
- McMurry, J., Mondragón, C. H., & Pozo, V. G. (2008). *Química orgánica*: Cengage learning México.
- Mohyuddin, S. (2019). Quinoa is beneficial to the comprehensive nutritional value of potential health. *Pakistan Journal of Science*, 71(2).
- Mortier, W. J., Ghosh, S. K., & Shankar, S. (1986). Electronegativity-equalization method for the calculation of atomic charges in molecules. *Journal of the American Chemical Society*, 108(15), 4315-4320.
- Mujica, A., & Jacobsen, S.-E. (2006). La quinua (*Chenopodium quinoa* Willd.) y sus parientes silvestres. *Botanica económica de los Andes Centrales*, 32, 449-457.

- Pérez Ramos, K., Elías Peñafiel, C., & Delgado Soriano, V. (2017). Bocado con alto contenido proteico: un extruido a partir de quinoa (*Chenopodium quinoa* Willd.), tarwi (*Lupinus mutabilis* Sweet) y camote (*Ipomoea batatas* L.). *Scientia Agropecuaria*, 8(4), 377-388.
- Plutowska, B., & Wardencki, W. (2008). Application of gas chromatography–olfactometry (GC–O) in analysis and quality assessment of alcoholic beverages—A review. *Food chemistry*, 107(1), 449-463.
- Polanski, J., & Gasteiger, J. (2016). Computer representation of chemical compounds. *Handbook of Computational Chemistry*; Leszczynski, J., Puzyn, T., Eds, 1-43.
- Pripp, A. H., Isaksson, T., Stepaniak, L., Sørhaug, T., & Ardö, Y. (2005). Quantitative structure activity relationship modelling of peptides and proteins as a tool in food science. *Trends in Food Science & Technology*, 16(11), 484-494.
- Raček, T., Schindler, O., Toušek, D., Horský, V., Berka, K., Koča, J., & Svobodová, R. (2020). Atomic Charge Calculator II: web-based tool for the calculation of partial atomic charges. *Nucleic acids research*, 48(W1), W591-W596.
- Ranjan, S., Sow, S., Ghosh, M., Padhan, S. R., Kumar, S., Gitari, H., . . . Nath, D. (2023). Nutraceutical properties and secondary metabolites of quinoa (*Chenopodium quinoa* Willd.): A review. *International Journal of Food Properties*, 26(2), 3477-3491.
- Rencher, A. C., & Schaalje, G. B. (2008). *Linear Models in Statistics*. USA: John Wiley & Sons, Inc.
- Risticvic, S., Carasek, E., & Pawliszyn, J. (2008). Headspace solid-phase microextraction–gas chromatographic–time-of-flight mass spectrometric methodology for geographical origin verification of coffee. *Analytica Chimica Acta*, 617(1-2), 72-84.
- Rojas, C., & Duchowicz, P. R. (2021). *Química Computacional de los Alimentos: Relaciones Cuantitativas Estructura-Actividad/Propiedad (QSAR/QSPR)*: Editorial Acribia, S.A.
- 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, 126-132.
- 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, 277-288.
- Rojas, W., Vargas Mena, A., & Pinto Porcel, M. (2016). La diversidad genética de la quinoa: potenciales usos en el mejoramiento y agroindustria. *Revista de Investigación e Innovación Agropecuaria y de Recursos Naturales*, 3(2), 114-124.
- Romo, S., Rosero, A., Forero, C., & Céron, E. (2006). Potencial nutricional de harinas de Quinoa (*Chenopodium Quinoa* W) variedad piartal en los Andes colombianos primera parte. *Bioteología en el sector Agropecuario y Agroindustrial*, 4(1), 112-125.
- Rubén, D., Burin, D., Pereyra, E., & Heras, A. I. (2015). *Quinoa, regalo ancestral: historia, contexto, tecnología, políticas*: Fundación Nueva Gestión.
- Samame, J. C. M., Monteza, V. A. T., Diaz, J. C. A., & Cabrera, H. I. M. (2017). Clasificación automática de tipos de semilla de quinoa a través de descriptores de color. *INGENIERÍA: Ciencia, Tecnología e Innovación*, 4(2).
- Shah, S., Khan, Y., & Yan, H. (2022). The phytochemical, pharmacological and medicina I evaluation of quinoa ( *Chenopodium quinoa* Willd.). *Pakistan Journal of Weed Science Research*, 28(2), 138.
- Song, J., Shao, Y., Yan, Y., Li, X., Peng, J., & Guo, L. (2021). Characterization of volatile profiles of three colored quinoas based on GC-IMS and PCA. *Lwt*, 146, 111292.

- Standards, N. I. o., & Technology. (2017). Libro del Web de Química del NIST, SRD 69. In: US Secretary of Commerce USA.
- The MathWorks Inc. (2022). MATLAB R2022b, <http://www.mathworks.com>.
- Todeschini, R., & Consonni, V. (2009). *Molecular descriptors for chemoinformatics*: John Wiley & Sons.
- Todeschini, R., & Consonni, V. (2010). Molecular descriptors for chemoinformatics: volume i: alphabetical listing/volume ii: appendices, references. *Methods and Principles in Medicinal Chemistry Series*.
- Varmuza, K., Dehmer, M., & Bonchev, D. (2012). *Statistical modelling of molecular descriptors in QSAR/QSPR*: Wiley Online Library.
- Vélez, J. P. A., Tabares, S. U., & Durán, N. P. (2022). Estudio de los compuestos orgánicos volátiles realizado por medio de microextracción en fase sólida por espacio de cabeza HS-SPME y cromatografía de gases acoplada a espectrometría de masas GC-MS de la cebolla de rama. *Ciencia Transdisciplinaria en la Nueva Era*, 60.
- Vilca, S. M., Espinoza, P., & Vidal, A. P. (2015). Multiplicación de semilla de variedades y ecotipos de quinua en valle de majes-Arequipa. *Revista Investigaciones Altoandinas*, 17(3), 2.
- Villacrés, E., Palacios, C., Maldonado, L., Cuadrado, L., Peralta, E., & Quelal, M. B. (2015). Evaluación del contenido y propiedades de la fibra dietética de la quinua (*Chenopodium quinoa* Willd.).
- Voigt, K. (2003). Databases on environmental information. *Handbook of Chemoinformatics: From Data to Knowledge in 4 Volumes*, 722-742.
- Voronov, S., Pleskachiov, Y., Shitikova, A., Zargar, M., & Abdelkader, M. (2023). Diversity of the biological and proteinogenic characteristics of quinoa genotypes as a multi-purpose crop. *Agronomy*, 13(2), 279.
- 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.
- Xu, X., Luan, F., Liu, H., & Gao, Y. (2011). QSPR models for the prediction of odor threshold of aliphatic alcohols. *Journal of Computational Science & Engineering*, 2, 69-78.
- Yang, X., Xing, B., Guo, Y., Wang, S., Guo, H., Qin, P., . . . Ren, G. (2022). Rapid, accurate and simply-operated determination of laboratory-made adulteration of quinoa flour with rice flour and wheat flour by headspace gas chromatography-ion mobility spectrometry. *Lwt*, 167, 113814.
- Yang, X., Zhu, K., Guo, H., Geng, Y., Lv, W., Wang, S., . . . Ren, G. (2021). Characterization of volatile compounds in differently coloured *Chenopodium quinoa* seeds before and after cooking by headspace-gas chromatography-ion mobility spectrometry. *Food chemistry*, 348, 129086.
- Zhang, Y., Yang, N., Fray, R. G., Fisk, I., Liu, C., Li, H., & Han, Y. (2018). Characterization of volatile aroma compounds after in-vial cooking of foxtail millet porridge with gas chromatography-mass spectrometry. *Journal of Cereal Science*, 82, 8-15.