

**CONSTRUCCIÓN DE MODELOS QSAR CON POSIBLE CAPACIDAD PREDICTIVA DE
LA ACTIVIDAD INSECTICIDA PARA LARVAS DE *Spodoptera frugiperda*, A PARTIR
DE PRODUCTOS DE ORIGEN NATURAL Y SINTÉTICO**

DIEGO FERNANDO ARDILA ARIZA

**UNIVERSIDAD PONTIFICIA BOLIVARIANA
ESCUELA DE INGENIERÍAS
MAESTRÍA EN CIENCIAS NATURALES Y MATEMÁTICA
MEDELLÍN
2025**

CONSTRUCCIÓN DE MODELOS QSAR CON POSIBLE CAPACIDAD PREDICTIVA DE LA ACTIVIDAD INSECTICIDA PARA LARVAS DE *Spodoptera frugiperda*, A PARTIR DE PRODUCTOS DE ORIGEN NATURAL Y SINTÉTICO

DIEGO FERNANDO ARDILA ARIZA

Trabajo de grado para optar al título de Magíster en Ciencias Naturales y Matemática

**Director:
CÉSAR AUGUSTO LUNA CÁCERES
M.Sc. en Química**

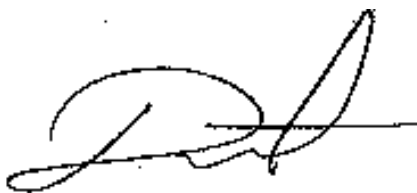
**UNIVERSIDAD PONTIFICIA BOLIVARIANA
ESCUELA DE INGENIERÍAS
MAESTRÍA EN CIENCIAS NATURALES Y MATEMÁTICA
MEDELLÍN
2025**

Mayo, 12 de 2025

Yo Diego Fernando Arzila Ariza,

“Declaro que este trabajo de grado no ha sido presentado con anterioridad para optar a un título, ya sea en igual forma o con variaciones, en ésta o en cualquiera otra universidad”.

Firma del autor:

A handwritten signature in black ink, appearing to be 'DFA', written over a horizontal line.

A la memoria de mi abuelo Genaro Ariza y mi tía María Aidé Ariza, que en la paz y regocijo del Señor descansan...

AGRADECIMIENTOS

Manifiesto mis más sinceros agradecimientos:

-Al director César Augusto Luna Cáceres, Qco. *M.Sc.* por su orientación, dedicación, comprensión y paciencia para culminar este trabajo de grado.

-A la Universidad Pontificia Bolivariana, Seccional Medellín por su apreciado apoyo incondicional, en especial al Programa de la Maestría en Ciencias Naturales y Matemática de la Escuela de Ingenieros por su acompañamiento en la culminación de los estudios.

CONTENIDO

	Pág.
RESUMEN.....	15
ABSTRACT	16
INTRODUCCIÓN.....	17
1. CONSTRUCCIÓN DE MODELOS QSAR CON POSIBLE CAPACIDAD PREDICTIVA DE LA ACTIVIDAD INSECTICIDA PARA LARVAS DE <i>Spodoptera frugiperda</i>, A PARTIR DE PRODUCTOS DE ORIGEN NATURAL Y SINTÉTICO	18
1.1 PLANTEAMIENTO DEL PROBLEMA.....	18
1.2 JUSTIFICACIÓN	19
1.3 OBJETIVOS	20
1.3.1 Objetivo general.....	20
1.3.2 Objetivos específicos.....	20
1.4 HIPÓTESIS.....	21
2. MARCO REFERENCIAL	22
2.1 MARCO CONCEPTUAL Y TEÓRICO.....	22
2.1.1 La mecánica cuántica	22
2.1.2 La teoría del funcional de densidad (DFT).....	22
2.1.3 La teoría del orbital molecular (OM).....	23
2.1.4 Optimización de la geometría molecular.....	23
2.1.5 La metodología QSAR.....	23
2.1.6 Los descriptores moleculares.....	23
2.1.7 Métodos estadísticos de regresión en el modelado QSAR	24
2.1.8 Métodos de validación interna y externa en modelos QSAR	26
2.2 ANTECEDENTES Y ESTADO DEL ARTE... ..	27
2.2.1 La acción de los insecticidas sobre los insectos plaga.....	27

2.2.2 Revisión de estudios QSAR de actividad insecticida reportados en la literatura	28
3. METODOLOGÍA	32
3.1 SELECCIÓN DE LOS COMPUESTOS CON ACTIVIDAD LARVICIDA FRENTE AL GUSANO COGOLLERO <i>Spodoptera frugiperda</i>	32
3.2 OPTIMIZACIÓN DE LA GEOMETRÍA MOLECULAR DE LOS COMPUESTOS SELECCIONADOS.....	32
3.3 CONSTRUCCIÓN DE LOS MODELOS QSAR PARA LA PREDICCIÓN DE LA ACTIVIDAD LARVICIDA.....	33
3.3.1 Cálculo de los descriptores moleculares.....	33
3.3.2 Construcción y validación de los modelos estadísticos	34
3.4 PREDICCIÓN DE LA ACTIVIDAD LARVICIDA DE MOLÉCULAS QUÍMICAS ESTRUCTURALMENTE RELACIONADAS.....	34
4. RESULTADOS Y DISCUSIÓN	36
4.1 SELECCIÓN DE COMPUESTOS CON ACTIVIDAD LARVICIDA FRENTE AL <i>Spodoptera frugiperda</i> PARA LA CONSTRUCCIÓN DE MODELOS QSAR.....	36
4.1.1 Conjunto de compuestos reportados con actividad larvícida.....	36
4.2 OPTIMIZACIÓN DE LA GEOMETRÍA MOLECULAR DE LOS COMPUESTOS SELECCIONADOS PARA EL ESTUDIO QSAR.....	43
4.3 CONSTRUCCIÓN DE MODELOS QSAR PARA LA PREDICCIÓN DE LA ACTIVIDAD LARVICIDA DE COMPUESTOS DE ORIGEN NATURAL Y SINTÉTICO FRENTE AL <i>Spodoptera frugiperda</i>	44
4.3.1 Cálculo de los descriptores moleculares para los compuestos del análisis QSAR	45
4.3.2 Modelo QSAR de actividad larvícida para derivados del eugenol	46
4.3.3 Modelo QSAR de actividad larvícida para amidas sintéticas.....	49
4.3.4 Modelo QSAR de actividad larvícida para los terpenos.....	51
4.3.5 Modelo QSAR de actividad larvícida para los flavonoides	54

4.4 VALIDACIÓN CRUZADA (CV) DE LOS MODELOS QSAR	56
4.4.1 Validación cruzada interna (CV-Int.) de los modelos de derivados de eugenol, amidas sintéticas, terpenos naturales y flavonoides.....	57
4.4.2 Validación cruzada externa (CV-Ext.) de los modelos de derivados de eugenol, amidas sintéticas, terpenos naturales y flavonoides.....	59
4.4.3 Predicción de la actividad larvicida de compuestos nuevos.....	63
5. CONCLUSIONES.....	68
6. RECOMENDACIONES	69
7. DIVULGACIÓN DE LOS RESULTADOS.....	70
8. BIBLIOGRAFÍA.....	71
ANEXOS.....	86

LISTA DE TABLAS

	Pág.
Tabla 1. Relación por dimensiones de los descriptores del paquete PaDEL.....	33
Tabla 2. ANOVA de los descriptores incluidos con el mayor peso estadístico al modelo QSAR de eugenoles para actividad larvicida	46
Tabla 3. Medidas estadísticas del modelo QSAR de derivados del eugenol para actividad larvicida.....	48
Tabla 4. ANOVA de los descriptores incluidos con el mayor peso estadístico al modelo QSAR de amidas sintéticas para actividad larvicida	49
Tabla 5. Medidas estadísticas del modelo QSAR de amidas sintéticas para actividad larvicida.....	50
Tabla 6. ANOVA de los descriptores incluidos con el mayor peso estadístico al modelo QSAR de terpenos para actividad larvicida	52
Tabla 7. Medidas estadísticas del modelo QSAR de terpenos naturales para actividad larvicida.....	53
Tabla 8. ANOVA de los descriptores incluidos con el mayor peso estadístico al modelo QSAR de flavonoides para actividad larvicida.....	54
Tabla 9. Medidas estadísticas del modelo QSAR de flavonoides para actividad larvicida	55
Tabla 10. Resultados de la actividad larvicida predicha para los compuestos de prueba, aplicando el modelo QSAR de eugenoles	60
Tabla 11. Resultados consolidados de la actividad larvicida predicha para las distintas combinaciones moleculares basadas en los modelos QSAR	67

LISTA DE FIGURAS

	Pág.
Figura 1. Ejemplo para el diseño de moléculas teóricas con base en las características estructurales definidos por los descriptores moleculares.....	35
Figura 2. Estructuras químicas de derivados de eugenol con actividad larvicida frente al <i>Spodoptera frugiperda</i>	36
Figura 3. Estructuras químicas de terpenos naturales con actividad larvicida frente al <i>Spodoptera frugiperda</i>	37
Figura 4. Estructuras químicas de amidas sintéticas con actividad larvicida frente al <i>Spodoptera frugiperda</i>	39
Figura 5. Estructuras químicas de flavonoides con actividad larvicida frente al <i>Spodoptera frugiperda</i>	40
Figura 6. Comparación de la actividad larvicida de compuestos naturales y sintéticos en diferentes unidades de dosis letal (DL_{50} y CL_{50})	41
Figura 7. Energía electrónica total de los compuestos del análisis QSAR.	43
Figura 8. Distribución de la cantidad de descriptores moleculares por categoría y dimensionalidad	45
Figura 9. Representación gráfica de la actividad larvicida experimental frente a la predicha para el modelo QSAR de derivados del eugenol.....	48
Figura 10. Representación gráfica de la actividad larvicida experimental frente a la predicha para el modelo QSAR de amidas sintéticas.	51
Figura 11. Representación gráfica de la actividad larvicida experimental frente a la predicha para el modelo QSAR de terpenos.	53
Figura 12. Representación gráfica de la actividad larvicida experimental frente a la predicha para el modelo QSAR de flavonoides.....	54
Figura 13. Representación gráfica de la actividad larvicida experimental frente a la predicha para los compuestos en la CV-LOO del modelo QSAR de eugenoles.....	57

Figura 14. Representación gráfica de la actividad larvicida experimental frente a la predicha para los compuestos en la CV-LOO del modelo QSAR de terpenos.....	58
Figura 15. Representación gráfica de la actividad larvicida experimental frente a la predicha para los compuestos en la CV-LOO del modelo QSAR de flavonoides.	58
Figura 16. Representación gráfica de la actividad larvicida experimental frente a la predicha para los compuestos en la CV-LOO del modelo QSAR de amidas	59
Figura 17. Representación gráfica de la actividad larvicida experimental frente a la predicha para el modelo QSAR de eugenoles. Incluye los compuestos de prueba.....	60
Figura 18. Representación gráfica de la actividad larvicida experimental frente a la predicha para el modelo QSAR de amidas. Incluye los compuestos de prueba.....	61
Figura 19. Representación gráfica de la actividad larvicida experimental frente a la predicha para el modelo QSAR de terpenos. Incluye los compuestos de prueba.	65
Figura 20. Representación gráfica de la actividad larvicida experimental frente a la predicha para el modelo QSAR de flavonoides. Incluye el compuesto de prueba.....	63
Figura 21. Estructuras químicas de los compuestos propuestos a partir de las combinaciones moleculares.....	65

LISTA DE ECUACIONES

	Pág.
Ecuación 1. Ecuación de Schrödinger dependiente del tiempo.....	22
Ecuación 2. Ecuación de Schrödinger independiente del tiempo.....	22
Ecuación 3. Ecuación de los orbitales moleculares	23
Ecuación 4. Regresión lineal.....	25
Ecuación 5. Regresión multilínea.....	25
Ecuación 6. Coeficiente R^2	25
Ecuación 7. Parámetro Fisher	26
Ecuación 8. Desviación estándar.....	26
Ecuación 9. Cálculo error de predicción	59

LISTA DE ANEXOS

	Pág.
Anexo 1. Matriz de las propiedades mecanocuánticas y termoquímicas de los compuestos del estudio	87
Anexo 2. ANOVA de los descriptores moleculares elegidos por el submétodo <i>Forward</i> en el intervalo $0,00 \leq \alpha \leq 0,10$ al ingresarlos en el modelo QSAR de derivados de eugenol..	88
Anexo 3. ANOVA de los descriptores teóricos mecanocuánticos y termoquímicos incluidos al análisis QSAR de los derivados de eugenol	88
Anexo 4. Matriz de los descriptores moleculares del modelo QSAR de eugenoles para actividad larvicida.	88
Anexo 5. Resultados de actividad larvicida predicha de la forma $\text{Log}_{10}(\text{DL}_{50})$ para los compuestos a partir del modelo QSAR de eugenoles.....	89
Anexo 6. ANOVA de los descriptores moleculares elegidos por el submétodo <i>Forward</i> en el intervalo $0,00 \leq \alpha \leq 0,10$ al ingresarlos en el modelo QSAR de amidas sintéticas	89
Anexo 7. ANOVA de los descriptores teóricos mecanocuánticos y termoquímicos incluidos al análisis QSAR de los derivados de eugenol	90
Anexo 8. Matriz de los descriptores moleculares del modelo QSAR de amidas sintéticas para actividad larvicida	90
Anexo 9. Resultados de actividad larvicida predicha de la forma $\text{Log}_{10}(\text{DL}_{50})$ para los compuestos a partir del modelo QSAR de amidas sintéticas.	90
Anexo 10. ANOVA de los descriptores moleculares elegidos por el submétodo <i>Forward</i> en el intervalo $0,00 \leq \alpha \leq 0,10$ al ingresarlos en el modelo QSAR de terpenos	90
Anexo 11. ANOVA de los descriptores teóricos mecanocuánticos y termoquímicos incluidos al análisis QSAR de los derivados de eugenol	91
Anexo 12. Matriz de los descriptores moleculares del modelo QSAR de terpenos para actividad larvicida	91
Anexo 13. Resultados de actividad larvicida predicha de la forma $\text{Log}_{10}(\text{DL}_{50})$ para los compuestos a partir del modelo QSAR de terpenos	92

Anexo 14. ANOVA de los descriptores moleculares elegidos por el submétodo Forward en el intervalo $0,00 \leq \alpha \leq 0,10$ al ingresarlos en el modelo QSAR de flavonoides	92
Anexo 15. ANOVA de los descriptores teóricos mecanocuánticos y termoquímicos incluidos al análisis QSAR de los derivados de eugenol	92
Anexo 16. Matriz de los descriptores moleculares del modelo QSAR de flavonoides para actividad larvica	93
Anexo 17. Resultados de actividad larvica predicha de la forma $\text{Log}_{10}(\text{CL}_{50})$ para los compuestos a partir del modelo QSAR de flavonoides	93
Anexo 18. Resultados de la actividad larvica predicha de la forma $\text{Log}_{10}(\text{DL}_{50})$ para los compuestos en la CV-LOO a partir del modelo QSAR de eugenoles.....	93
Anexo 19. Resultados de la actividad larvica predicha de la forma $\text{Log}_{10}(\text{DL}_{50})$ para los compuestos en la CV-LOO a partir del modelo QSAR de amidas sintéticas	93
Anexo 20. Resultados de la actividad larvica predicha de la forma $\text{Log}_{10}(\text{DL}_{50})$ para los compuestos en la CV-LOO a partir del modelo QSAR de terpenos	94
Anexo 21. Resultados de la actividad larvica predicha de la forma $\text{Log}_{10}(\text{CL}_{50})$ para los compuestos en la CV-LOO a partir del modelo QSAR de flavonoides	94
Anexo 22. Resultados de actividad larvica predicha de la forma $\text{Log}_{10}(\text{DL}_{50})/\text{Log}_{10}(\text{CL}_{50})$ para los compuestos de prueba a partir de los modelos QSAR para eugenoles, amidas, terpenos y flavonoides.	94
Anexo 23. Matriz de los descriptores moleculares para los compuestos propuestos (modelo QSAR de eugenoles).....	95
Anexo 24. Matriz de los descriptores moleculares para los compuestos propuestos (modelo QSAR de amidas).....	95
Anexo 25. Matriz de los descriptores moleculares para los compuestos propuestos (modelo QSAR de terpenos)	95
Anexo 26. Matriz de los descriptores moleculares para los compuestos propuestos (modelo QSAR de flavonoides).....	95
Anexo 27. ANOVA individual para la selección de los dos mejores descriptores moleculares del modelo QSAR para derivados del eugenol, amidas, terpenos y flavonoides	96

RESUMEN

A nivel mundial, la propagación acelerada por insectos plaga son uno de los principales problemas del sector agrícola en regiones tropicales y subtropicales. Las plantas de maíz (*Zea mays L.*), sorgo (*Sorghum*) y algodón (*Gossypium hirsutum*) son algunos de los cultivos más afectados (Ávila-Murillo et al., 2014; Morillo & Notz, 2001). Cada año, más de 10 millones de agricultores se ven afectados por la reducción del rendimiento de sus cultivos, ocasionados por el *Spodoptera frugiperda*. Por esta causa, en el mundo se pierden alrededor de 20,0 millones de toneladas de maíz, colocando en riesgo la seguridad alimentaria y nutricional de las personas (Bateman et al., 2018; Njuguna et al., 2021).

Por más de 60 años se ha implementado el uso de insecticidas sintéticos como organoclorados, organofosforados, carbamatos, peritroides, entre otros, para el control de vectores y plagas (He et al., 2020). El metomil y lambdacihalotrina son larvicidas de mayor uso en campañas de control del *Spodoptera frugiperda*, los cuales son tóxicos para la salud humana y los insectos; puesto que han generado resistencia a su poder insecticida (Carvalho et al., 2013; Ríos-Díez & Saldamando-Benjumea, 2011).

Para sustituir los insecticidas sintéticos se emplean nuevos compuestos provenientes de las plantas u otros recursos que sean biodegradables y menos tóxicos que los usados actualmente (Sombra et al., 2020). Así, se ha encontrado que algunos derivados de eugenol, amidas sintéticas y compuestos obtenidos de extractos vegetales y aceites esenciales como algunos terpenos, flavonoides, estilbenos y alcaloides presentan actividad larvicida contra el *Spodoptera frugiperda* (Batista-Pereira et al., 2006; Castral et al., 2011; Hernández-Carlos & Gamboa-Angulo, 2019; Romanelli et al., 2010; Romo-Asunción et al., 2015; Vargas-Méndez et al., 2019).

En esta investigación, se construyeron modelos QSAR con posible capacidad de predicción, empleando datos de actividad larvicida reportados en algunos derivados de eugenol, amidas sintéticas, terpenos y flavonoides frente *Spodoptera frugiperda*. Para tal fin, se realizaron cálculos mecanocuánticos para la optimización y obtención de propiedades electrónicas de los compuestos seleccionados, para ser usados con los descriptores moleculares, con el método DFT, la función base 6-31G(d,p) y el funcional de intercambio y correlación híbrido B3LYP. Fueron calculados los descriptores moleculares para construir los modelos empleando el método estadístico MLR, y posteriormente fueron validados por medio de la validación cruzada *leave-one-out* o interna y externa logrando obtener modelos QSAR de actividad larvicida con parámetros de los coeficientes de determinación (R^2), de validación interna (Q_{LOO}^2) y externa (R_{pred}^2) superiores a un valor de 0,70.

PALABRAS CLAVE: actividad insecticida; QSAR; *Spodoptera frugiperda*.

ABSTRACT

The accelerated spread of insect pests is one of the main problems of the agricultural sector in tropical and subtropical regions worldwide. Corn (*Zea mays L.*), sorghum (*Sorghum*), and cotton (*Gossypium hirsutum*) plants are some of the most affected crops (Ávila-Murillo et al., 2014; Morillo & Notz, 2001). Every year, more than 10 million farmers are affected by reduced crop yields caused by *Spodoptera frugiperda*. For this reason, around 400 million tons of corn are lost in the world, placing people's food and nutritional security at risk (Bateman et al., 2018; Njuguna et al., 2021).

The use of synthetic insecticides such as organochlorines, organophosphates, carbamates, pyrethroids, among others, has been implemented to control vectors and pests for more than 60 years (He et al., 2020). Methomyl and lambda-cyhalothrin are the most widely used larvicides in *Spodoptera frugiperda* control campaigns, but are also toxic to human health and insects. This *Spodoptera frugiperda* has generated resistance to the insecticidal power (Carvalho et al., 2013; Ríos-Díez & Saldamando-Benjumea, 2011).

In order to replace synthetic insecticides, new compounds are used from plants or other biodegradable resources and less toxic than those currently used (Sombra et al., 2020). Thus, it has been found that some eugenol derivatives, synthetic amides and compounds obtained from plant extracts and essential oils such as flavonoids and stilbenes, have larvicidal activity against *Spodoptera frugiperda* (Romanelli et al., 2010; Romo-Asunción et al., 2015; Vargas-Méndez et al., 2019).

In this research, QSAR models with potential predictive capability were constructed using reported larvicidal activity data for certain eugenol derivatives, synthetic amides, terpenes, and flavonoids against *Spodoptera frugiperda*. For this purpose, quantum mechanical calculations were performed to optimize and obtain electronic properties of the selected compounds, to be used with molecular descriptors, using the DFT method, the 6-31G(d,p) basis set, and the B3LYP hybrid exchange-correlation functional. Molecular descriptors were calculated to build the models using the MLR statistical method, and were later validate them through internal and external cross-validation or *leave-one-out* obtaining QSAR models of larvicidal activity with parameters of the coefficients of determination (R^2), internal (Q_{Loo}^2) and external (R_{pred}^2) validation greater than a value of 0.70.

KEYWORDS: Insecticidal activity; QSAR; *Spodoptera frugiperda*.

INTRODUCCIÓN

En la actualidad se registran altas tasas de pérdida en cultivos de maíz debido a la propagación de insectos plaga como el *Spodoptera frugiperda*. Se estima que en todo el mundo alrededor del 60% de los cultivos de maíz en fase de plántula son afectados por el *Spodoptera frugiperda* (Ávila-Martínez et al., 2023; Hernández-Trejo et al., 2018). Adicionalmente, este insecto es la principal plaga del maíz, por el cual afecta su producción hasta un 52%, ocasionando daños en los tejidos vegetales y hojas, debido a su comportamiento voraz y rápido, y su capacidad migratoria; amenazando la seguridad alimentaria de millones de personas (Lima et al., 2010; WAN et al., 2021).

Para el control de insectos plaga se han empleado diversos tipos de insecticidas sintéticos, entre los cuales prevalecen el metomil y lambdacihalotrina que a nivel mundial es empleado para controlar larvas de orden Lepidoptera (Ahissou et al., 2021; Morillo & Notz, 2001; Rios-Díez & Saldamando-Benjumea, 2011). Sin embargo, los registros sobre la variedad de casos de resistencia, por su uso indiscriminado y sin vigilancia adecuada, han mostrado el alto grado de resistencia al metomil y lambdacihalotrina de los vectores en países de Latinoamérica como México, Nicaragua, Colombia, Ecuador, Venezuela, entre otros (Gutiérrez-Moreno et al., 2019; Morillo & Notz, 2001; Reséndiz-Ramírez et al., 2017; Zenner de Polanía et al., 2009).

Una de las alternativas para reemplazar el metomil y lambdacihalotrina es emplear compuestos provenientes de aceites esenciales vegetales que presentan actividad larvicida. Esta propiedad, al igual que otras actividades biológicas, se le atribuye a los metabolitos secundarios presentes en dichos aceites (Ileer et al., 2022; Salinas-Sánchez et al., 2012). Cabe resaltar que en la literatura se reportan estudios sobre la existencia de compuestos con actividad larvicida más efectiva contra el *Spodoptera frugiperda* (Ávila-Murillo et al., 2014; Castral et al., 2011; Hernández-Carlos & Gamboa-Angulo, 2019; Vargas-Méndez et al., 2019).

Es de destacar que el desarrollo de metodologías computacionales para predecir actividades biológicas reduce considerablemente costos y esfuerzos (García-Domenech et al., 2010). Por este motivo, resulta muy útil la construcción de modelos para predecir actividades biológicas empleando la metodología Relación Cuantitativa Estructura-Actividad (QSAR) (Todeschini et al., 2009; Villaverde et al., 2020). En el presente trabajo de grado se construyeron cuatro modelos QSAR a partir de conjuntos de compuestos con actividad larvicida reportada en la literatura para predecir esta propiedad.

1. CONSTRUCCIÓN DE MODELOS QSAR CON POSIBLE CAPACIDAD PREDICTIVA DE LA ACTIVIDAD INSECTICIDA PARA LARVAS DE *Spodoptera frugiperda*, A PARTIR DE PRODUCTOS DE ORIGEN NATURAL Y SINTÉTICO

1.1 PLANTEAMIENTO DEL PROBLEMA

La alta propagación e infestación de los sectores agrícolas por el gusano *Spodoptera frugiperda* abarca las zonas tropicales y subtropicales del planeta, principalmente en países de Medio y Lejano Oriente, África, además de América (Pérez-Gutiérrez et al., 2011; WAN et al., 2021). La mayoría de los agricultores afectados por las pérdidas de los cultivos de maíz (*Zea mays L.*), sorgo (*Sorghum*) y algodón (*Gossypium hirsutum*) debido al *S. frugiperda*, se encuentran en el centro y oeste de África, donde aproximadamente el valor de estas pérdidas oscila entre US \$2481 millones y US \$6187 millones, y más de USD \$460 millones de pérdidas en los cultivos de maíz (*Zea mays L.*), se han registrado en países de América (Ayil-Gutiérrez et al., 2018; Phambala et al., 2020; Sisay et al., 2018).

En primer lugar, hay que señalar que aún en bajas concentraciones, el larvicida metomil y lambdacihalotrina ocasionan efectos tóxicos en animales vertebrados y en ecosistemas acuáticos. También, el metomil y lambdacihalotrina afectan la salud de las personas y se acumulan con facilidad en los suelos al ser absorbidos por la vegetación (Djouaka et al., 2018; Lin et al., 2020; Mohi El-Din et al., 2014; Van Scoy et al., 2013). Estos compuestos, se empezaron a restringir como medida para reducir la resistencia de los insectos plaga que infestan los cultivos agrícolas; principalmente el cultivo de maíz (*Zea mays L.*) (Morillo & Notz, 2001; Ríos-Díez & Saldamando-Benjumea, 2011). Por tal razón, se hace necesaria la búsqueda de nuevos compuestos de origen natural y sintético que presenten actividad larvicida contra estos insectos plaga.

Se han reportado numerosos metabolitos secundarios que se encuentran en los aceites esenciales de las plantas. Por ejemplo, los terpenos presentan actividades biológicas como insecticidas, fungicidas, antialimentarias, repelentes, ovicidas, larvicidas, entre otras (Ayil-Gutiérrez et al., 2018; Leyva et al., 2009; Lima et al., 2020). Algunos terpenos como timol, pineno y citral son relativamente inocuos en el ser humano; por lo cual, la implementación de aceites esenciales y sus compuestos mayoritarios es una alternativa para obtener nuevos compuestos con actividad larvicida (Ávila-Murillo et al., 2014; Geldof et al., 1992; Hernández-Carlos & Gamboa-Angulo, 2019).

En este orden de ideas, la determinación de la actividad larvicida de todos los posibles metabolitos secundarios resulta una tarea costosa y de mucho tiempo; puesto que se requiere realizar un número considerable de bioensayos, estudiar diversos tipos de plantas y plantear rutas sintéticas para desarrollar nuevos compuestos activos (Cherkasov et al., 2014). Por consiguiente, la construcción de modelos estadísticos para predecir la actividad larvicida aplicando la metodología QSAR es una alternativa factible para identificar compuestos que podrían tener esta actividad; puesto que su finalidad es estimar razonablemente el potencial activo presente en las moléculas, sin remitir a la síntesis, caracterización y prueba de amplios conjuntos de moléculas químicas (Gadaleta et al., 2016; Walker et al., 2003). Adicionalmente con un modelo QSAR se puede predecir el comportamiento de la actividad, disminuyendo el número de moléculas a evaluar

experimentalmente y recayendo en menos ensayos a realizar (Giaginis et al., 2014; Muratov et al., 2020).

En esta investigación, se propuso construir modelos QSAR para predecir la actividad insecticida de compuestos sintéticos nuevos y compuestos mayoritarios de aceites esenciales frente a larvas de *Spodoptera frugiperda*.

1.2 JUSTIFICACIÓN

Se estima que cada año, se pierden entre 8,3 y 20,6 millones de toneladas de maíz (*Zea mays L.*) en el mundo debido a la infestación del gusano *Spodoptera frugiperda* y la falta de un control adecuado (Phambala et al., 2020; Sisay et al., 2018). Factores como los cambios climáticos, la colonización de nuevos hábitats y el establecimiento en ellos favorecen la propagación de insectos plaga, incrementando así las pérdidas en los cultivos agrícolas (Baloch et al., 2020; Early et al., 2018).

En 83 países tropicales, más de 100 millones de agricultores enfrentan pérdidas económicas a causa de esta plaga, especialmente en regiones como India, Etiopía, Kenia, Ghana, Zambia, China, Nepal y Japón, donde reside el 80 % de los agricultores afectados (Bateman et al., 2018; HE et al., 2021; Zhao et al., 2020). En América, *S. frugiperda* ha sido reportada en más de 13 países, incluidos Brasil, Colombia, Ecuador, Argentina, Estados Unidos y México, lo que evidencia su rápida expansión geográfica (Figueroa-Brito et al., 2011; Hardke et al., 2015; Siddhartha et al., 2019; Soberón Risco et al., 2012). En Colombia, esta plaga está presente en gran parte del territorio, con mayor incidencia en los Llanos Orientales, el Eje Cafetero, las regiones costeras, los Santanderes, el Amazonas y el Chocó (Zenner de Polanía et al., 2009).

La Organización Mundial de la Salud (OMS) y la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO) han promovido el uso de insecticidas sintéticos para combatir plagas y vectores, especialmente en áreas urbanas y selváticas (FAO & WHO, 2016; Jepson et al., 2020). Sin embargo, el uso indiscriminado de estos productos ha generado resistencia en insectos, así como riesgos ecológicos y para la salud humana (Lizarazo et al., 2008; Ordóñez-García et al., 2020; Trumper, 2014; Vásquez et al., 2006).

En América, países como Venezuela, México, Brasil y naciones de la región Caribe reportan altos niveles de resistencia al metomil y a la lambdacihalotrina en *S. frugiperda* (Carvalho et al., 2013; Lira et al., 2020; Morillo & Notz, 2001). En Colombia, esta resistencia se ha identificado en regiones como Cundinamarca, Valle del Cauca, Santander, Tolima, Huila, Antioquia y La Guajira (Barrera et al., 2011; Saldamando & Vélez-Arango, 2010; Santos Amaya et al., 2009; Zenner de Polanía & Borrero Fonseca, 1996).

Frente a esta problemática, se han investigado nuevos insecticidas, tanto naturales como sintéticos, siendo los primeros una alternativa prometedora (Harve & Kamath, 2004), aunque presentan limitaciones económicas y logísticas (García-Domenech et al., 2010; Kausar & Falcao, 2018). En este contexto, la metodología QSAR (relación cuantitativa estructura-actividad) resulta útil para reducir costos y optimizar el desarrollo de compuestos,

permitiendo estimar actividades biológicas mediante modelos estadísticos (Dambolena et al., 2016; Gramatica, 2020; Kar et al., 2017; Villaverde et al., 2020). QSAR es una herramienta útil para predecir actividades biológicas a partir de modelos estadísticos que permiten evaluar parámetros como el coeficiente de determinación (R^2), Fisher (F), desviación estándar (s) y coeficiente de predicción (Q^2) (Eriksson et al., 2003; Todeschini et al., 1995). Además, esta metodología predice actividades biológicas a partir de modelos matemáticos elaborados, como lo señalan Liu & Long, 2009; Vukovic et al., 2019 quienes resaltan su aplicabilidad para estimar valores de actividad con precisión.

Estudios QSAR han demostrado su eficacia para predecir actividades larvicidas en compuestos frente a insectos plaga. Por ejemplo, Romanelli et al. (2010) diseñaron un modelo predictivo con derivados de flavonoides contra *S. frugiperda*, utilizando el descriptor molecular Mor28p para evaluar la relación del peso molecular con los momentos dipolares. El modelo logró un $R^2 = 0,882$, $Q^2 = 0,812$ y valores estadísticamente significativos ($F = 27,898$; $p = 0,00$). Da Silva et al. (2015) desarrollaron un modelo basado en tiosemicarbazonas para *Aedes aegypti*, empleando descriptores moleculares como el momento dipolar (μ) y el coeficiente de partición octanol/agua (LogP), obteniendo parámetros estadísticos de $R^2 = 0,97$ y $F = 54,61$. Finalmente, Wang et al. (2012) aplicaron QSAR a isoxazolininas y podofilotoxinas para *Mythimna separata*, integrando descriptores moleculares como GATS4e, EEig06x, RDF080v, Mor09v y H-052, logrando un modelo con $R^2 = 0,861$, $Q^2 = 0,797$ y $s = 0,0592$.

En definitiva, la metodología QSAR permite construir modelos con los que se puede predecir la actividad insecticida. En este caso, se diseñaron modelos estadísticos que permiten predecir la actividad larvicida de compuestos puros frente al insecto *Spodoptera frugiperda*, a partir de datos experimentales reportados en la literatura.

1.3 OBJETIVOS

1.3.1 Objetivo general:

Construir modelos QSAR para la actividad larvicida de terpenos, flavonoides, amidas sintéticas y derivados de eugenol frente al gusano cogollero *Spodoptera frugiperda*.

1.3.2 Objetivos específicos:

- Identificar las familias de compuestos moleculares con actividad larvicida frente al *Spodoptera frugiperda* para la construcción de modelos QSAR para cada una de ellas.
- Proponer modelos QSAR para cada familia encontrada en la literatura, empleando la metodología estadística de la regresión lineal múltiple (MLR).
- Evaluar la capacidad predictiva de los modelos QSAR mediante validación cruzada *leave-one-out* y validación externa, y proponer nuevas moléculas con potencial actividad larvicida basadas en las características estructurales de compuestos naturales y sintéticos.

1.4 HIPÓTESIS

La metodología QSAR es una herramienta válida para desarrollar modelos estadísticos que resulten capaces de cuantificar relaciones entre la estructura química y la actividad de insecticidas de origen natural y sintético, a través de la estrategia del análisis de regresión lineal múltiple (MLR) aplicada a diferentes bases de datos de propiedades de interés investigativo extraídas de la literatura. Para ello, se emplearán los descriptores moleculares que surjan del análisis de cientos de descriptores teóricos, obtenidos de programas computacionales de acceso libre. Así, los modelos QSAR podrán ser utilizados para la predicción de actividad insecticida de nuevos compuestos sintetizados en el laboratorio y carentes de datos experimentales; además de asistir a la búsqueda de estructuras químicas con valores favorables de actividad insecticida para el control del gusano cogollero *Spodoptera frugiperda*.

2. MARCO REFERENCIAL

2.1 MARCO TEÓRICO Y CONCEPTUAL

La metodología QSAR se fundamenta en la estadística multivariable aplicada a descriptores moleculares, obtenidos a partir de la estructura química optimizada mediante métodos mecanocuánticos (Ambure et al., 2019; Katritzky et al., 1994). Entre estos descriptores destacan la energía electrónica total, el momento dipolar y la energía de los orbitales HOMO y LUMO (Jensen, 1999; Stone & Jonathan, 1993).

El principal desafío en QSAR es la selección de descriptores que permitan construir modelos predictivos eficientes (Golbraikh & Tropsha, 2002; Randić & Basak, 1999). Se emplean métodos estadísticos como la regresión multilineal (MLR) (Konovalov et al., 2008; Schüürmann et al., 2008) y es esencial optimizar la geometría molecular para obtener estructuras estables y calcular descriptores relevantes (Rinnan et al., 2010).

2.1.1 La mecánica cuántica. La mecánica cuántica permite predecir propiedades fisicoquímicas y estructurales a partir de la ecuación de Schrödinger, tanto dependiente (Ecuación 1) como independiente del tiempo (Ecuación 2) (Atkins & Friedman, 2011; Essin & Griffiths, 2006; Sakura & Napolitano, 1994).

Dentro de la química cuántica, la ecuación de Schrödinger tiene dos perspectivas de estudio: dependiente del tiempo (Ecuación 1) e independiente (Ecuación 2).

Ecuación 1. Ecuación de Schrödinger dependiente del tiempo.

$$-\frac{\hbar}{i} \frac{\partial \Psi(q, t)}{\partial t} = -\frac{\hbar^2 \partial^2 \Psi(q, t)}{2m \partial^2 q} + V(q) \Psi(q, t) \quad (1)$$

Ecuación 2. Ecuación de Schrödinger independiente del tiempo.

$$-\frac{\hbar^2 \partial^2 \psi(x)}{2m \partial^2 x} + V(x) \psi(x) = E \psi(x) \quad (2)$$

Existen diversos métodos en la mecánica cuántica que permiten describir la energía electrónica de un sistema mediante funcionales, entre los cuales destacan la Teoría del Funcional de Densidad (DFT) y la aproximación Hartree-Fock (HF), ampliamente utilizados para estudiar sistemas electrónicos (Friesner, 2005; Manzhos, 2020; Schleich et al., 2013).

2.1.2 La teoría del funcional de densidad (DFT). El método DFT es un enfoque de la mecánica cuántica que permite obtener la estructura electrónica de las moléculas a partir de la densidad electrónica en lugar de la función de onda (Ψ_f), optimizando así los cálculos de energía de un sistema (Krishnan et al., 1980). A diferencia de los métodos *ab initio* como Hartree-Fock (HF), que no utilizan parámetros semiempíricos, la DFT incorpora funcionales de intercambio y correlación aproximados para estimar la energía electrónica, lo que reduce los tiempos de simulación sin perder precisión (Altürk et al., 2017).

Algunos métodos DFT utilizan conjuntos base para describir los electrones, como 6-31G(d,p), STO-3G y def2-QZVP, facilitando cálculos eficientes en programas como Gaussian, ORCA y GAMESS (Caldeweyher & Brandenburg, 2018; Weigend & Ahlrichs, 2005). Dentro de los funcionales más utilizados en DFT se encuentra B3LYP, que combina términos de energía de intercambio y correlación (Becke, 1993; Paier et al., 2007).

2.1.3 La teoría del orbital molecular (OM). Ésta describe las moléculas de acuerdo a la combinación lineal de los orbitales atómicos, para obtener la distribución de los electrones con el fin de establecer la estructura de un compuesto; la interacción de los orbitales se requiere para comprender la unión entre los átomos (Hall, 1950).

De acuerdo con la teoría, las moléculas están constituidas por grupos de orbitales moleculares que caracterizan una función de onda (Ψ_f), a la cual se le atribuye una combinación lineal de un conjunto de funciones base (X_i). En la misma función de onda, los coeficientes de normalización (c_{ij}) aportan la contribución de cada función base al orbital molecular (Ψ_f), y se calculan a partir de sustituciones en la ecuación de Schrödinger (Ecuación 3) (Atkins & Friedman, 2011; Hall, 1991; Høst et al., 2008).

Ecuación 3. Ecuación de los orbitales moleculares.

$$\Psi_f = \sum_{i=1}^n (c_{ij})(X_i) \quad (3)$$

Ahora bien, las interacciones de mayor importancia se dan entre dos orbitales que sean próximos en energía y orbitales de otra entidad (Huang et al., 2017; Masan et al., 2021). A éstos se les denomina orbitales frontera y su clasificación corresponde a: orbital molecular ocupado de mayor energía (HOMO) y orbital molecular desocupado de menor energía (LUMO) (De Lile et al., 2020).

2.1.4 Optimización de la geometría molecular. Este proceso se basa en la búsqueda de la conformación más estable sobre la superficie de energía potencial (PES) (Kim et al., 2015). A través de iteraciones, se obtiene la geometría con menor energía (Vreven et al., 2006). Este procedimiento es crucial en estudios QSAR, pues influye en la estimación de descriptores moleculares y en la calidad estadística de los modelos (Rinnan et al., 2010).

2.1.5 La metodología QSAR. El concepto de QSAR surgió en 1872 con Crum-Brown y Fraser, evolucionando con Hansch y Fujita, quienes establecieron bases matemáticas para su desarrollo (Hansch, 1976). Actualmente, QSAR se emplea en química, toxicología y farmacología para predecir la actividad de compuestos antes de su síntesis (Cherkasov et al., 2014; Muratov et al., 2020).

La hipótesis central es que la actividad de una molécula está determinada por su estructura química (Golbraikh & Tropsha, 2002). Los modelos QSAR cuantifican esta relación para estimar la actividad de nuevas sustancias (Kwon et al., 2019; Muhammad et al., 2018).

2.1.6 Los descriptores moleculares. Por más de 50 años son considerados representaciones numéricas de la estructura química de una molécula, permitiendo establecer relaciones con su actividad biológica, toxicidad y propiedades fisicoquímicas

(Guha & Willighagen, 2012). Se dividen en teóricos (calculados matemáticamente) y experimentales (derivados de mediciones en laboratorio) (Todeschini & Consonni, 2000). Los descriptores moleculares se pueden clasificar de acuerdo a su complejidad o método de cálculo, de la siguiente manera (Fujita, 1997; Hansch et al., 2001; Martin, 2012; Winkler, 2002):

a) Los descriptores teóricos: Se obtienen de la estructura química y se clasifican por su nivel de información.

-0D y 1D (constitucionales y de conteo): Incluyen el peso molecular, número de enlaces y grupos funcionales.

-2D (topológicos y electrotopológicos): Describen la conectividad atómica mediante índices como Randić y Hosoya.

-3D (geométricos y electroestáticos): Consideran el volumen, área de superficie y distribución de carga electrónica.

-4D (basados en interacciones): Analizan interacciones fármaco-receptor con modelos como los métodos de Análisis Comparativo del Campo Molecular CoMFA y el Análisis Comparativo de Índices de Similaridad Molecular (CoMSIA).

b) Los descriptores experimentales: Se obtienen de mediciones en el laboratorio.

-Hidrofóbicos: LogP (lipofilidad).

-Electrónicos: Coeficiente de Hammett (σ), polarizabilidad.

-Estéricos: Volumen molecular, constante estérica de Taft (E_s).

-Termodinámicos: La constante de disociación (pK_a), entalpía (ΔH).

Actualmente, se dispone de miles de conceptos y teorías sobre descriptores moleculares asequibles en la literatura (Rácz et al., 2019). Sin embargo, un problema principal a solucionar en la metodología QSAR es la selección apropiada de un grupo representativo y reducido de descriptores, para construir un modelo estadístico con la capacidad de explicar y predecir lo mejor posible a la actividad de estudio (Neves et al., 2018; Shahlaei, 2013; Tandon et al., 2019).

2.1.7 Métodos estadísticos de regresión en el modelado QSAR. Los estudios QSAR utilizan métodos estadísticos para correlacionar las propiedades químicas de una molécula con su actividad biológica o fisicoquímica. Para ello, se aplican técnicas de regresión que permiten desarrollar ecuaciones matemáticas que describan esta relación (Cronin, 2010; Eriksson et al., 2003; Golbraikh et al., 2017).

Los métodos de regresión lineal son ampliamente utilizados en QSAR, ya que proporcionan modelos interpretables y reproducibles. Entre ellos, el análisis de regresión multilínea (MLR) es uno de los más empleados debido a su simplicidad y eficiencia para establecer correlaciones entre descriptores moleculares y la actividad biológica (Pirhadi et al., 2015; Saxena & Prathipati, 2003; Yee & Wei, 2012).

Por el grado de sencillez, reproducibilidad y aplicación, el método MLR es ampliamente empleado para la selección de las mejores correlaciones entre los datos de actividad y los descriptores moleculares determinados para un compuesto o sistema químico (Yee & Wei, 2012). Un modelo de regresión lineal consiste en predecir la variable dependiente Y

(actividad) a partir de la variable independiente X_1 (descriptor) con su coeficiente de regresión (a_1) y una constante aleatoria (a_0) como se indica en la Ecuación 4 (Harrell, 2001; Roy et al., 2008):

Ecuación 4. Regresión lineal.

$$\begin{aligned} \text{Actividad} &= f(\text{descriptor molecular}) \\ Y &= a_0 + a_1X_1 \quad (4) \end{aligned}$$

Por otro lado, las soluciones basadas en la estrategia MLR, mostradas en la Ecuación 5 emplean más de una variable independiente para calcular de forma lógica el efecto final estimado (Eriksson et al., 2003). Así pues, el método MLR establece una relación lineal entre la actividad biológica (Y) y dos o más descriptores moleculares (X_i) (Gramatica, 2007; Yap et al., 2007).

Ecuación 5. Regresión multilineal.

$$\begin{aligned} \text{Actividad} &= f(\text{descriptores moleculares}) \\ Y &= a_0 + a_1X_1 + a_2X_2 + a_3X_3 \dots + a_nX_n \quad (5) \end{aligned}$$

En el método MLR, cada variable independiente tiene una contribución delimitada y de forma directa, que depende del valor del coeficiente de regresión y su signo algebraico respectivo. Cada coeficiente deberá ser significativo con $p < 0,05$; el cual se puede verificar mediante una prueba *t-student* (Andrade, 2019; Darlington & Hayes, 2016; Roy et al., 2015b).

La calidad estadística de un modelo MLR-QSAR se evalúa a partir de ciertos parámetros como se describe a continuación:

a) El coeficiente de correlación múltiple (R^2). Este parámetro determina la calidad de ajuste de un modelo; es decir, la similitud que existe entre los datos observados y los predichos por el modelo (Cheng et al., 2014). Este coeficiente se define a través de la Ecuación 6. Adicionalmente, un valor alto del coeficiente R^2 no es indicativo de que el modelo se ajusta bien a los datos analizados (Veerasamy et al., 2011).

Ecuación 6. Coeficiente R^2 .

$$R^2 = 1 - \frac{\sum(Y_{obs} - Y_{pred})^2}{\sum(Y_{obs} - \bar{Y}_{obs})^2} \quad (6)$$

De acuerdo con la ecuación anterior, Y_{obs} representa valores de respuesta observada, mientras que Y_{pred} corresponde a datos de respuesta calculada con el modelo; por último, \bar{Y}_{obs} hace referencia a los valores promedio de la respuesta observada (Roy et al., 2015b). Para la interpretación de este parámetro, se debe considerar que la correlación R^2 toma valores desde 0,0 a 1,0 (porcentajes), e indica la variación en la variable dependiente, explicada por la ecuación de regresión; así un valor cercano a 0,0 muestra que la ecuación no explica la variable dependiente (Leach & Gillet, 2007).

b) La varianza de Fisher (F). La forma de evaluar y juzgar la importancia de los coeficientes de regresión es utilizando el parámetro/prueba Fisher (Ecuación 7). Éste relaciona el significado de la regresión con el valor medio de las desviaciones.

Ecuación 7. Parámetro F.

$$F = \frac{\frac{\sum(Y_{pred} - \bar{Y})^2}{d}}{\frac{\sum(Y_{obs} - Y_{pred})^2}{N - d - 1}} \quad (7)$$

Además, el valor de F presenta dos grados de libertad: d y N-d-1 como se indica en la ecuación anterior. Para el caso de los modelos QSAR, el parámetro d corresponde a los descriptores moleculares del estudio y la expresión N-d-1 permite calcular los grados de libertad. Cabe mencionar que valores altos de Fisher (F) corresponden a mayores niveles de significancia de los coeficientes de regresión (Roy et al., 2015b).

c) La desviación estándar (s). Para obtener un buen modelo, la estimación del error estándar de la variable dependiente debe ser lo más bajo posible (Roy et al., 2015b). La desviación es calculada utilizando la Ecuación 8 como se señala enseguida:

Ecuación 8. Desviación estándar.

$$s = \sqrt{\frac{\sum(Y_{obs} - Y_{pred})^2}{N - d - 1}} \quad (8)$$

De otro lado, se debe destacar que la desviación estándar es una medida absoluta de la calidad del ajuste de los modelos estadísticos. Idealmente, esta medida debería ser cero, pero en situaciones experimentales no es así (Vanderkar et al., 2007).

Con todo lo anterior, los parámetros de ajuste de los datos no verifican una buena predicción de un modelo; puesto que son parámetros de calidad estadística del mismo (Gramatica, 2012). Por consiguiente, la razón central es aplicar métodos de validación en los modelos QSAR, para evaluar su capacidad predictiva con conjuntos moleculares no ensayados o sin datos experimentales (Tropsha et al., 2003; Veerasamy et al., 2011). A continuación, se describe la fundamentación teórica de los métodos de validación que se emplearán en esta investigación.

2.1.8 Métodos de validación interna y externa en modelos QSAR. Desde que la validación de los modelos es considerada una etapa definitiva para su aceptación, varias estrategias de validación han sido útiles para controlar la correlación aleatoria de los descriptores moleculares, para así evitar el ajuste excesivo de los datos, y poder evaluar la veracidad, capacidad predictiva y precisión del modelo QSAR (Gramatica, 2007).

Entre los métodos más comunes de validación de modelos MLR-QSAR, en primer lugar, se destaca la validación cruzada (CV) de tipo *leave-one-out* “dejando uno fuera” (CV-LOO) o validación interna (CV-Int.), y en segundo lugar, la validación externa (CV-Ext.) (Chirico & Gramatica, 2011; Konovalov et al., 2008). Así, la CV-Int. consiste en eliminar un compuesto

del modelo y aplicar la ecuación para predecir la actividad del compuesto; de esta forma será medible su calidad y bondad de ajuste (Golbraikh et al., 2003; Wold, 1978). Este procedimiento se realiza tantas veces como el número de compuestos que estén incluidos en el estudio QSAR. En cuanto a la CV-Ext., ésta confirma la capacidad predictiva, aplicabilidad y estabilidad del modelo utilizando un nuevo grupo de moléculas no empleadas dentro del análisis QSAR (Muhammad et al., 2018; Roy, 2007).

A través de la validación, la calidad, ajuste y predictibilidad de un modelo se puede evaluar en términos de varias medidas estadísticas como los coeficientes de determinación en validación interna (Q_{LOO}^2) y validación externa (R_{pred}^2) (Gramatica & Sangion, 2016). Un modelo QSAR es considerado aceptable para predecir si $Q_{LOO}^2 > 0,5$ y $R_{pred}^2 > 0,6$ (Golbraikh & Tropsha, 2002). Además, los valores del error estándar, pueden usarse como medida de la precisión y capacidad predictiva de un modelo QSAR, con tal que el error estándar del modelo estadístico no sea mucho mayor que los errores de los datos biológicos (Wood et al., 2013).

2.2 ANTECEDENTES Y ESTADO DEL ARTE

El gusano cogollero *Spodoptera frugiperda* es una de las plagas más devastadoras en cultivos agrícolas clave como el maíz, sorgo, algodón, arroz y pastos. Su alta capacidad de reproducción y adaptación a diversos ecosistemas ha generado significativas pérdidas económicas (Ávila-Murillo et al., 2014; Morillo & Notz, 2001). A pesar de los esfuerzos mediante insecticidas sintéticos y métodos de fumigación, la plaga ha desarrollado resistencia a múltiples compuestos químicos, reduciendo la eficacia de estos controles (Reséndiz-Ramírez et al., 2017).

Dado este panorama, la investigación se ha centrado en el desarrollo de insecticidas naturales y en la aplicación de modelos computacionales como QSAR, que permiten predecir la actividad de nuevas moléculas con potencial insecticida (Chaudhry et al., 2007; Paluch et al., 2009). Estos modelos representan una alternativa prometedora para el diseño de compuestos eficientes, menos costosos y con menor impacto ambiental (Duchowicz et al., 2009; Hansch et al., 1974).

2.2.1 La acción de los insecticidas sobre los insectos plaga. Los insecticidas sintéticos han sido ampliamente utilizados en la agricultura desde mediados del siglo XX, con clases como los organoclorados (DDT), organofosforados, carbamatos y piretroides. Sin embargo, la resistencia a estos compuestos ha sido documentada en múltiples especies de insectos plaga, incluido *S. frugiperda* (Barrera et al., 2011; Zenner de Polanía et al., 2009).

En Colombia, la resistencia a metomil, clorpirifos y cipermetrina ha sido reportada desde 1989 en cultivos de La Guajira, extendiéndose luego a otras regiones del país como Cundinamarca, Tolima, Antioquia y Santander (Ríos-Díez & Saldamando-Benjumea, 2011). Esta problemática ha llevado a la búsqueda de insecticidas biológicos, como los derivados de *Bacillus thuringiensis* y productos fermentativos de *Saccharopolyspora spinosa*. Sin embargo, su uso prolongado también ha resultado en el desarrollo de resistencia (Ávila-Murillo et al., 2014; Cruz et al., 2014; Rutte et al., 2020).

Como alternativa, se ha promovido el uso de extractos vegetales con propiedades insecticidas. Los aceites esenciales de familias como *Euphorbiaceae*, *Fabaceae*, *Caricaceae* y *Meliaceae* han demostrado actividad contra diversas plagas, presentando baja toxicidad y alta biodegradabilidad (Pérez-Gutiérrez et al., 2011; Santiago et al., 2008; Santos et al., 2010). Sin embargo, la necesidad de encontrar compuestos con eficacia probada ha impulsado el uso de modelos computacionales como QSAR para predecir su actividad antes de realizar ensayos experimentales.

2.2.2 Revisión de estudios QSAR de actividad insecticida reportados en la literatura.

La metodología QSAR permite establecer relaciones matemáticas entre la estructura química de un compuesto y su actividad biológica, facilitando la predicción de su efectividad como insecticida (Perkins et al., 2003; Tropsha & Golbraikh, 2007).

Para la mayoría de estudios que se citan a continuación, las investigaciones basadas en la metodología QSAR hacen referencia a la actividad insecticida de diversos compuestos de origen natural y sintético:

a) Romanelli et al. (2010) realizaron un estudio en el que sintetizaron 26 derivados de flavonoides y aplicaron la metodología QSAR para evaluar su actividad insecticida contra *Spodoptera frugiperda*. Utilizaron la regresión lineal múltiple (MLR) para desarrollar un modelo predictivo basado en el descriptor molecular Mor28p, que está relacionado con la polarización electrónica de la molécula. La ecuación de predicción obtenida fue: $\text{Log}_{10}(\text{TL}_{50}) = 1,807 - 3,843 \text{ Mor28p}$. El modelo presentó valores estadísticos sólidos: coeficiente de determinación $R^2 = 0,882$; coeficiente de predicción $Q^2 = 0,812$; estadístico Fisher $F = 27,898$; desviación estándar $s = 0,093$ y una significancia estadística $p = 0,00$; lo que indica una alta capacidad predictiva.

Para evaluar la validez del modelo, los autores realizaron una validación cruzada externa con 16 flavonoides que no fueron incluidos en la construcción del modelo QSAR inicial. Compararon los valores experimentales con los predichos para tres flavonoides de referencia: flavona, 7-hidroxi flavona y 6-bromoflavona, obteniendo diferencias mínimas entre los valores calculados y los observados. Los compuestos luteolina y apigenina fueron identificados como los más activos, ya que presentaron los valores más bajos de TL_{50} , indicando una mayor toxicidad contra *S. frugiperda*.

b) Wang et al. (2012) evaluaron 33 isoxazolininas y derivados de podofilotoxina con actividad larvívica frente a *Mythimna separata*. Desarrollaron un modelo QSAR utilizando el método MOBY DIGS y emplearon regresión lineal múltiple (MLR) combinada con algoritmos genéticos (GA) para seleccionar los mejores descriptores. La ecuación de predicción obtenida fue: $\text{LOF} = - 2,843 + 1,511 \text{ GATS4e} + 0,505 \text{ EEig06x} - 0,062 \text{ RDF080v} - 0,307 \text{ Mor09v} + 0,081 \text{ H-052}$. Donde GATS4e representa la electronegatividad de Sanderson, RDF080v la distribución radial atómica, y Mor09v el volumen de Van der Waals. Los resultados mostraron una alta capacidad predictiva con $R^2 = 0,861$; $Q^2 = 0,797$ y $s = 0,0592$.

Para optimizar la geometría molecular de los compuestos, utilizaron el método AM1, lo que permitió mejorar la precisión de los cálculos y evaluar la relación entre la estructura electrónica y la toxicidad. Los resultados revelaron que cuatro derivados de podofilotoxina, identificados como 5e, 9c, 11g y 11h, mostraron la mayor actividad insecticida, con diferencias mínimas entre los valores experimentales y predichos. Esto sugiere que estos

compuestos pueden servir como base para el desarrollo de nuevos insecticidas con alta efectividad larvívora.

c) Nakagawa et al. (2001) construyeron siete modelos QSAR basados en 46 compuestos derivados de dibenzolhidrazinas con actividad larvívora contra *Leptinotarsa decemlineata*. Aplicaron análisis de regresión lineal múltiple (MLR) con selección de descriptores mediante el método *Forward*, obteniendo la siguiente ecuación para el mejor modelo: $pDL_{50} = 3,429(\pm 1,760) (\text{LogP}) - 0,536(\pm 0,272) (\text{LogP})^2 - 1,838(\pm 0,261) (I_{\text{para}}) + 0,550(\pm 2,766)$. Los parámetros estadísticos mostraron una alta precisión: $R^2 = 0,965$; $F_{2,8} = 112,72$ y $s = 0,281$, lo que indica una alta precisión en la predicción de la actividad insecticida.

Los resultados mostraron que los compuestos con sustituyentes halogenados (flúor, bromo y yodo) presentaron la mayor actividad larvívora, con valores experimentales y predichos altamente correlacionados. Esto sugiere que la inclusión de estos grupos funcionales en la estructura de las dibenzolhidrazinas mejora su efecto insecticida.

d) Grodnitzky & Coats (2002) estudiaron 20 monoterpenos acíclicos contra *Musca domestica* y desarrollaron dos modelos QSAR. En el primer modelo, utilizaron descriptores geométricos y electrónicos, obteniendo la ecuación: $\text{Log}(1/DL_{50}) = -30,7(\pm 4,9) + 15,1(\pm 2,4) [\text{E-State}] + 213,8(\pm 36) [\text{GETAWAY}] - 105,8(\pm 17,6) [\text{Interaction}]$. El modelo mostró valores predictivos elevados en el coeficiente de correlación $R^2 = 0,86$; coeficiente de predicción $Q^2 = 0,72$; en un Fisher estadístico $F = 32,59$ y un error de estimación residual $s = 0,11$. En su segundo modelo, basado en el compuesto timol y sus derivados, identificaron que la densidad electrónica de Mulliken era el descriptor más importante, con la ecuación: $\text{Log}(1/DL_{50}) = 65,3(\pm 7,9) - 11,6(\pm 1,4) [\text{Mulliken population}]$. Los parámetros estadísticos confirmaron su precisión son: $R^2 = 0,90$; $Q^2 = 0,84$; $F = 68,52$ y $s = 0,08$.

Para optimizar la geometría molecular de los compuestos, utilizaron el método semiempírico AM1 en CAChe Work System, mientras que la densidad electrónica fue calculada con GAMESS utilizando la función de base 6-31G(d). Los monoterpenos más activos fueron carvomenten-4-il pivalato, carvil pivalato y limoneno, con valores experimentales y predichos altamente correlacionados. En el modelo específico de timol, los derivados isopropil éter y etil éter de timol fueron los más efectivos.

e) Zhao et al. (2012) sintetizaron y evaluaron 21 derivados de 4-Fenil-acil-sustituido-3-(2,5-dimetilfenil)-4-hidroxiespiro[4,5]dec-3-eno-2,8-diona con actividad insecticida contra *Aphis fabae* y *Tetranychus cinnabarinus*. Utilizaron DFT con la función de base 6-31G(d) para la optimización geométrica y regresión lineal múltiple (MLR) en SPSS 16.0 para construir el modelo QSAR, permitiendo predecir la actividad insecticida en función de los descriptores mecanocuánticos y fisicoquímicos más relevantes. La ecuación de predicción obtenida fue: $\%I = 23,619 + 89,081 \text{HOMO} + 8,125 \text{LUMO} + 12,218 \text{Q}_N - 1,136 \text{ClogP} + 0,081 \text{MR} - 0,009 \text{V}_W - 0,326 \text{DM} - 0,008 \text{E}_T$. El modelo presentó valores estadísticos robustos, con un coeficiente de correlación $R = 0,937$; un estadístico Fisher $F = 10,734$; una desviación estándar $s = 0,218$ y una significancia estadística $p = 0,0002$, indicando su alta capacidad predictiva.

Los compuestos más activos fueron 8-I-b, 8-I-g y 8-II-a, con valores experimentales y predichos altamente consistentes. Estos resultados confirmaron la importancia de las propiedades electrónicas e hidrofóbicas en la actividad insecticida, destacando la utilidad del modelo QSAR en la optimización del diseño de nuevos insecticidas.

f) dos Santos et al. (2018) aplicaron la metodología QSAR a un conjunto de 55 monoterpenos con actividad larvicida contra *Aedes aegypti*, empleando regresión de mínimos cuadrados parciales (PLS) para construir un modelo predictivo basado en descriptores moleculares de importancia estructural, hidrofóbica y electrónica. Para la optimización geométrica de los compuestos, utilizaron el método PM6, con refinamiento adicional mediante DFT-B3LYP/6-31G+(d) en GAUSSIAN09. La ecuación de predicción obtenida fue: $pIC_{50} = -0.6019 (\pm 0.3788) MATS7e - 1.44696 (\pm 10.9534) (VE1sign_A) (2) - 0.1094 (\pm 0.0994) (nCt) + 0.2936 (\pm 0.1607) (MLogP) + 0.3398 (\pm 0.3256) (R8u) + 2,3196 (\pm 0,3577)$. El modelo mostró alta capacidad predictiva, con coeficiente de determinación $R^2 = 0.75$; validación interna $Q^2 = 0.68$; validación externa $Q^2 = 0.85$ ($K^b = 10$) y error estándar $s = 0.284$.

Los compuestos más activos fueron 7, 8, 14 y 15, con valores de pIC_{50} entre 3.64 y 3.85, lo que indica una elevada toxicidad contra las larvas de *Aedes aegypti*. Los resultados sugieren que los monoterpenos con anillos aromáticos y sustituyentes halogenados poseen una mayor eficacia insecticida, lo que refuerza la utilidad del modelo QSAR en la identificación de compuestos con alto potencial larvicida y bajo impacto ambiental.

g) Soares et al. (2020) realizaron un análisis QSAR sobre 42 monoterpenos con actividad larvicida frente a *Reticulitermes chinensis* y *Drosophila melanogaster*, utilizando regresión lineal múltiple (MLR) para construir modelos predictivos. Para la optimización geométrica de los compuestos, emplearon el software GROMACS 5.0, mientras que la selección de descriptores se llevó a cabo en Mobydigs 1.1. Las ecuaciones de predicción obtenidas para cada especie fueron:

Para *Drosophila melanogaster*: $pCL_{50} = 0.22 (\pm 0.14) X2v - 0.35 (\pm 0.12) Mor21u + 0.28 (\pm 0.13) nAROH + 0.26 (\pm 0.11) H-049 + 0,35 (\pm 0.13) BLTD48 + 6.81 (\pm 0.10)$.

Para *Reticulitermes chinensis*: $pCL_{50} = -1.13 (\pm 0.10) MATS2m - 0.68 (\pm 0.12) ESpm09u + 0.74 (\pm 0.12) JGI5 + 6.66 (\pm 0.09)$.

Los valores estadísticos obtenidos confirmaron la precisión de los modelos, con $R^2 = 0,81$; $s = 0,298$; $Q^2 = 0,72$; $S_{PRESS} = 0.296$ y $F = 22.3$ para *Drosophila melanogaster*, $R^2 = 0,80$; $s = 0,529$; $Q^2 = 0,77$; $S_{PRESS} = 0.490$ y $F = 41.2$ para *Reticulitermes chinensis*.

Los monoterpenos más activos fueron pulegona, citronelal y carvacrol, con valores de pCL_{50} elevados, lo que indica una alta toxicidad larvicida. Estos resultados sugieren que los compuestos con estructuras aromáticas funcionalizadas y ciertas propiedades hidrofóbicas presentan mayor actividad insecticida, lo que demuestra la eficacia de QSAR en la identificación de compuestos prometedores para el control de plagas.

Con esta revisión de estudios QSAR, se puede concluir que existen modelos estadísticos para predecir con precisión la actividad insecticida de compuestos de diversas familias químicas. Además, la construcción y aplicación de un modelo QSAR resulta útil cuando las pruebas y medidas experimentales son arduas y de alto costo; puesto que provee información importante para orientar el diseño racional, rápido y selectivo de nuevas moléculas con actividad esperada, basándose en la información química, biológica y física de los compuestos existentes y ensayados (Dearden, 2016; Muhammad et al., 2018). Por esta razón, la metodología QSAR constituye una herramienta computacional imprescindible

en la búsqueda de nuevas estructuras químicas con valores propicios de actividad insecticida sobre el principal insecto plaga de los cultivos de maíz, *Spodoptera frugiperda*. En esta investigación, se propone construir modelos QSAR para predecir la actividad insecticida de compuestos de origen natural y sintético frente a larvas de *Spodoptera frugiperda*.

3. METODOLOGÍA

En la presente propuesta de investigación, se planteó construir modelos QSAR con posible capacidad de predecir la actividad larvica de compuestos de origen natural y sintético. El diseño de estos modelos empezó con el cálculo de los descriptores moleculares teóricos para los compuestos de cada familia química seleccionada de la literatura. Seguido, se aplicó un tratamiento estadístico con el fin de encontrar la relación cuantitativa de la estructura química de los compuestos con la actividad larvica y posteriormente se evaluó su eficiencia; y se finalizó con el planteamiento de posibles sistemas moleculares que puedan ser eficientes en el proceso insecticida frente al *Spodoptera frugiperda*.

3.1 SELECCIÓN DE LOS COMPUESTOS CON ACTIVIDAD LARVICIDA FRENTE AL GUSANO COGOLLERO *Spodoptera frugiperda*

Se realizó una búsqueda de compuestos de origen natural y sintético con actividad larvica, expresada en términos de la dosis/concentración letal media (DL_{50}/CL_{50}) en diversas bases de datos de índole científica, tales como Scopus, ScienceDirect, PubMed, entre otras. La búsqueda se orientó a familias químicas de terpenos, flavonoides, amidas sintéticas y derivados de eugenol, principalmente (Ávila-Murillo et al., 2014; Castral et al., 2011; Estrela et al., 2003; Hernández-Carlos & Gamboa-Angulo, 2019; Vargas-Méndez et al., 2019).

De acuerdo con investigaciones reportadas en la literatura, las moléculas mencionadas anteriormente son potencialmente activas, dado que en sus estructuras químicas poseen sustituyentes dadores y aceptores de electrones, anillos aromáticos, cadenas alifáticas y/o cíclicas, las cuales pueden ser determinantes en la actividad larvica (Ávila-Murillo et al., 2014; Castral et al., 2011; Estrela et al., 2003; Hernández-Carlos & Gamboa-Angulo, 2019; Vargas-Méndez et al., 2019). Así pues, estas familias químicas fueron seleccionadas para posteriormente realizar una revisión bibliográfica; puesto que están conformadas por compuestos provenientes de extractos vegetales y aceites esenciales (terpenos y flavonoides) y de rutas sintéticas alternativas basadas en química verde (amidas y derivados de eugenol).

3.2 OPTIMIZACIÓN DE LA GEOMETRÍA MOLECULAR DE LOS COMPUESTOS SELECCIONADOS

En primer lugar, se construyeron las estructuras químicas de las moléculas seleccionadas usando el programa Avogadro 4.8 (Avogadro, 2018; Hanwell et al., 2012), con lo cual se obtuvo una primera distribución espacial de los átomos. En segundo lugar y partiendo de dicha distribución, se realizaron los cálculos mecanocuánticos necesarios para optimizar la geometría molecular, llevando a los átomos a su distribución espacial más estable. El nivel de teoría empleado fue el B3LYP/6-31G(d,p) (Becke, 1993; Bochevarov et al., 2013; Mousavi et al., 2013; Pople et al., 1996), por ser el nivel más recurrente en este tipo de familias moleculares y con el cual se obtiene buenos resultados; puesto que permite un buen tratamiento de la correlación electrónica, y conlleva a una mejor descripción de los sistemas moleculares (DeLong et al., 2017; Kar et al., 2010; Nikolic & Agababa, 2009;

Rasulev et al., 2007; Rinnan et al., 2010; Wan et al., 2004). Las simulaciones se realizaron en el programa ORCA 4.2 (Neese et al., 2020; ORCA, 2017).

Así mismo, esta simulación molecular mecanocuántica permite obtener, además de la optimización geométrica, los siguientes descriptores mecanocuánticos: la energía electrónica total, la energía del orbital molecular más alto ocupado (E_{HOMO}) y el más bajo desocupado (E_{LUMO}), la diferencia de energía entre los orbitales moleculares HOMO-LUMO (Gap) y el momento dipolar total (Todeschini & Consonni, 2009).

3.3 CONSTRUCCIÓN DE LOS MODELOS QSAR PARA LA PREDICCIÓN DE LA ACTIVIDAD LARVICIDA

El diseño de los modelos QSAR para la propuesta de investigación se realizó en tres etapas: en la primera, se calcularon los descriptores moleculares teóricos a partir de la geometría optimizada de cada molécula; en la segunda etapa se desarrolló el tratamiento estadístico para obtener los modelos estadísticos de predicción; y la tercera etapa correspondió a la validación cruzada interna y externa; además de la aplicación del modelo para predecir la actividad larvicida de compuestos que no tienen datos experimentales.

3.3.1 Cálculo de los descriptores moleculares. Con el programa de acceso libre PaDEL Descriptor, se obtuvieron diversos descriptores moleculares, a partir de la geometría molecular optimizada de cada compuesto (Yap, 2010). Los descriptores obtenidos, se pueden clasificar como electrónicos, geométricos, constitucionales, mecanocuánticos, topológicos, electrotopológicos, entre otros. PaDEL permite calcular un total de 1875 descriptores moleculares como se muestra en la Tabla 1 (Moriwaki et al., 2018).

Por otro lado y teniendo en cuenta, que los compuestos seleccionados presentan varios heteroátomos y grupos funcionales; debido que éstos afectan los niveles y distribuciones electrónicas de las moléculas (Lin et al., 2020), no se logró plantear el descriptor teórico que explicara el comportamiento electrónico a través de un análisis poblacional natural, debido a la complejidad de asociación entre los valores calculados para las cargas atómicas parciales de las moléculas de estudio. Estos resultados se iban a determinar con el programa gratuito JANPA 2.01 (JANPA, 2020; Nikolaienko et al., 2014).

Tabla 1. Relación por dimensiones de los descriptores del paquete PaDEL.

Descriptores unidimensionales (1D)	Descriptores bidimensionales (2D)	Descriptores tridimensionales (3D)
Descriptores constitucionales (120)	<ul style="list-style-type: none"> - Descriptores de autocorrelación (346) - Descriptores de Basak (42) - Descriptores BCUT (6) - Descriptores de Burden (96) - Descriptores de conectividad (56) 	<ul style="list-style-type: none"> - Descriptores CPSA (29) - Descriptores RDF (210) <ul style="list-style-type: none"> - Descriptores geométricos (21) - Descriptores WHIM (91) <ul style="list-style-type: none"> - Descriptores de autocorrelación 3D (80)

	<ul style="list-style-type: none"> - Descriptores de estado electrotopológico (489) - Descriptores de Kappa (3) - Descriptores de propiedad molecular (15) <ul style="list-style-type: none"> - Descriptores quimicocuánticos (6) - Descriptores topológicos (265) 	
--	--	--

Nota. En la Tabla se muestra el listado de descriptores moleculares calculados por PaDEL, relacionados en familias y dimensiones. Adaptada de la biblioteca de descriptores moleculares, ChemDes, 2021 (<https://bittylink.com/8e4>) CC BY-NC-SA 4.0.

Con base en lo anterior, se obtuvieron las matrices que contienen estas variables (anteriormente llamados descriptores) para cada uno de los conjuntos moleculares, incluyendo los datos reportados de actividad larvicida expresados en dosis/concentración letal media (DL_{50}/CL_{50}) de cada compuesto.

3.3.2 Construcción y validación de los modelos estadísticos. Para cada familia de compuestos declarados en el primer apartado metodológico, se construyeron modelos QSAR, el cual se obtienen mediante una regresión multilineal (MLR) empleando como variables independientes los descriptores y como dependiente la actividad larvicida en su forma logarítmica ($\log_{10}DL_{50}/CL_{50}$). Para ello, los descriptores moleculares se correlacionaron linealmente con los datos de actividad larvicida (como variables experimentales) para el *S. frugiperda* reportados en la literatura (Todeschini et al., 1995; Yap, 2014).

Con base en el análisis MLR, se tuvo en cuenta las medidas estadísticas de los coeficientes de correlación múltiple (R^2), la significancia estadística (p), el error estándar residual (EER), la varianza Fisher (F), justificado en un análisis de varianza (ANOVA), entre otros (Eriksson et al., 2003; Veerasamy et al., 2011). El tratamiento estadístico se desarrolló con la interfaz RStudio 4.0 del programa libre R, utilizando el paquete *Data Analysis and Graphics Data and Functions* (DAAG) (Nagaraja, 2014; RStudio, 2021). Como resultado, se obtuvieron modelos estadísticos con predicción de la actividad larvicida.

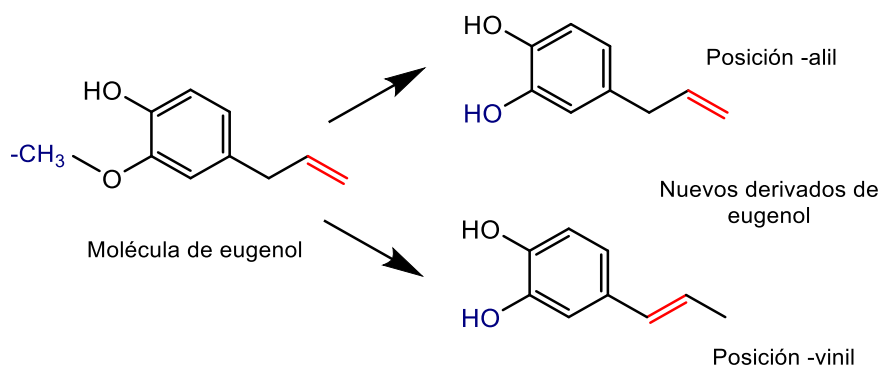
Posteriormente, los modelos QSAR se probaron mediante dos validaciones; por un lado, la validación cruzada interna, mediante el coeficiente Q_{LOO}^2 para evaluar su capacidad predictiva (Cramer et al., 1988; Muhammad et al., 2018). Y por otro la validación externa, para la cual se utilizó un conjunto de compuestos de prueba o *test set* con actividad larvicida reportada en la literatura (no incluidos en la construcción de los modelos QSAR) a través del coeficiente R_{pred}^2 (Gramatica & Sangion, 2016; Kolossov & Stanforth, 2010).

3.4 PREDICCIÓN DE LA ACTIVIDAD LARVICIDA DE MOLÉCULAS QUÍMICAS ESTRUCTURALMENTE RELACIONADAS

Con base en el apartado anterior, la validación externa es una de las etapas más importantes en el desarrollo de modelos QSAR, ya que con esta estrategia se confirma la robustez y la capacidad predictiva de los modelos (Veerasamy et al., 2011). De manera que se diseñaron nuevas moléculas teóricas a partir de la modificación de los sustituyentes entre los compuestos de las diferentes familias seleccionadas, con el objeto de predecir la actividad larvicida empleando los modelos propuestos en el estudio.

Para diseñar estas moléculas, se seleccionaron las características estructurales de cada familia, a través de los descriptores moleculares *n*-ésimos (dm_n) más representativos de cada modelo QSAR. Para ello, se seleccionaron los 2 descriptores moleculares teóricos de mayor peso o significancia estadística ($p < 0,001$), los cuales fueron corroborados mediante el análisis de varianza (ANOVA) (Bugeac et al., 2021; Filipowska et al., 2016; RácZ et al., 2018). Teniendo en cuenta lo anterior, se realizaron las respectivas combinaciones de las características estructurales que brindan mayor aporte a la actividad larvicida, de la siguiente manera: combinación A (dm_1 , dm_2 , dm_3 y dm_4); combinación B (dm_1 , dm_2 , dm_5 y dm_6); combinación C (dm_1 , dm_2 , dm_7 y dm_8); combinación D (dm_3 , dm_4 , dm_5 y dm_6); combinación E (dm_3 , dm_4 , dm_7 y dm_8); combinación F (dm_5 , dm_6 , dm_7 y dm_8). Cabe señalar que, los descriptores dm_1 y dm_2 pertenecen al modelo QSAR 1; dm_3 y dm_4 al modelo QSAR 2; dm_5 y dm_6 al modelo QSAR 3; por último, el dm_7 y dm_8 al modelo QSAR 4. Como resultado, se propusieron 2 moléculas por cada combinación, para un total de 12 moléculas teóricas y posteriormente, se calcularon sus actividades implementado los modelos QSAR diseñados. A continuación, se planteó un ejemplo de cómo se diseñarían estas moléculas; en este caso, a partir de la familia de derivados de eugenol (Figura 1).

Figura 1. Ejemplo para el diseño de moléculas teóricas con base en las características estructurales definidos por los descriptores moleculares.



A partir de la molécula del eugenol, se propusieron sustituciones estructurales en el grupo metoxi ($-OCH_3$) por grupos hidroxilo ($-OH$); ya que incide en la densidad electrónica y la polarización molecular, la cual se explica por el descriptor RDF50p (descriptor que relaciona la distribución radial y ponderada por polarización relativa). Por otra parte, se planteó la transposición del doble enlace alil a posición vinil, el cual promueve deslocalización electrónica e incide en el estado electrotopológico de las moléculas; esta característica se explica por el descriptor SHAvin (sumatoria de átomos de hidrógeno sobre cada carbono en los enlaces vinil enlazados con los carbonos aromáticos). La molécula del eugenol fue tomada de la investigación de Vargas-Méndez et al., 2019.

4. RESULTADOS Y DISCUSIÓN

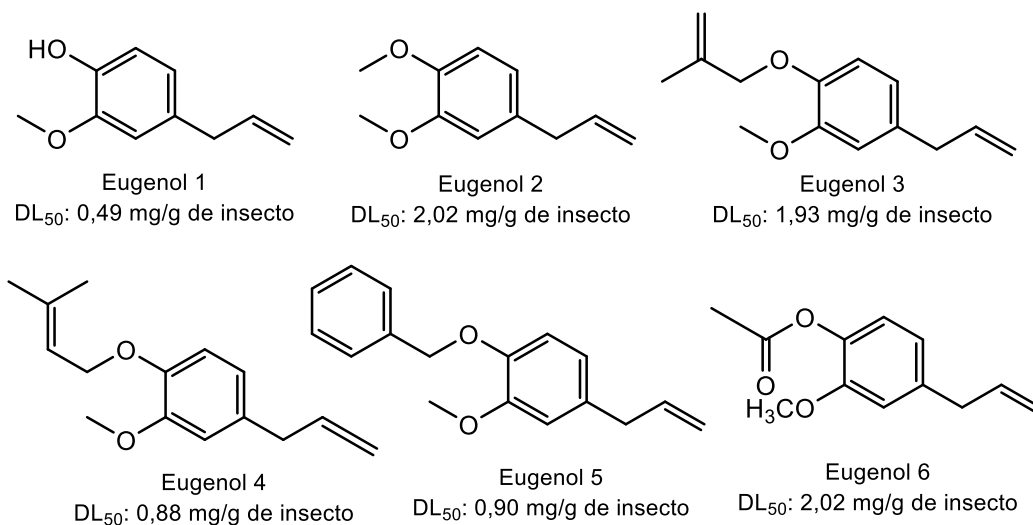
La construcción de los modelos QSAR de actividad insecticida frente a larvas de *Spodoptera frugiperda* se realizó en cuatro etapas: en la primera, se seleccionaron las familias de compuestos con actividad larvicida reportada en la literatura; en la segunda, se optimizaron los compuestos para el estudio QSAR; en la tercera se realizó el tratamiento estadístico para obtener el modelo; y la cuarta etapa es la validación cruzada interna y externa; además de la aplicación del modelo para predecir la actividad larvicida de compuestos que no tienen datos experimentales.

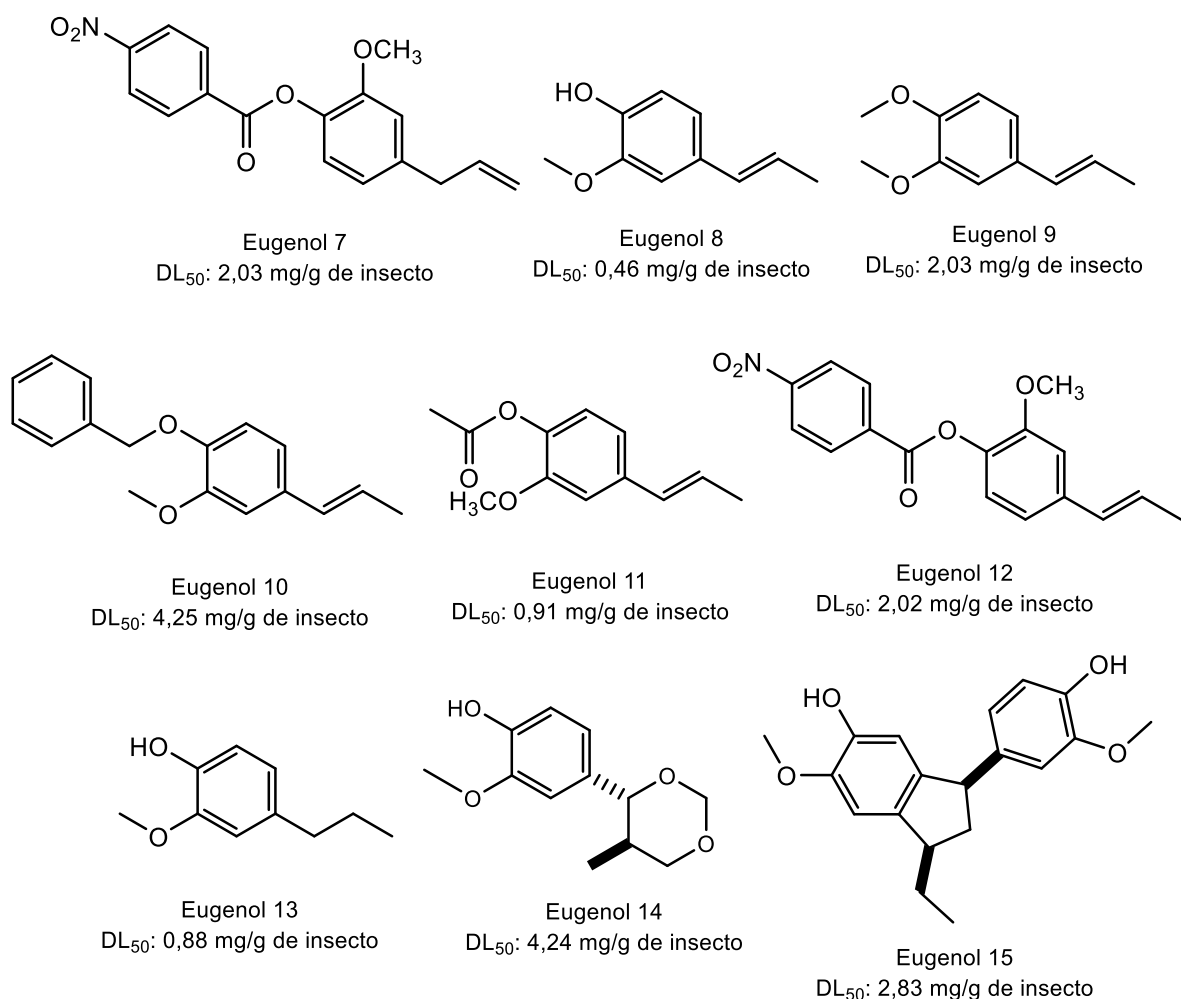
4.1 SELECCIÓN DE COMPUESTOS CON ACTIVIDAD LARVICIDA FRENTE AL *Spodoptera frugiperda* PARA LA CONSTRUCCIÓN DE MODELOS QSAR

4.1.1 Conjunto de compuestos reportados con actividad larvicida. De los compuestos con actividad larvicida frente a *Spodoptera frugiperda* reportados en la literatura, se seleccionaron 44 compuestos. Estos fueron clasificados en cuatro familias químicas: terpenos naturales, derivados del eugenol, amidas sintéticas y flavonoides. A continuación, se describe brevemente cada una de estas familias:

Derivados de Eugenol: Para Vargas-Méndez et al. (2019), la actividad de estos derivados (Figura 2) se centra en la aromaticidad de los anillos y a los sustituyentes dadores de electrones, principalmente los grupos hidroxilo (-OH) y metoxi (-OCH₃), ya que éstos tienden a ser más tóxicos contra las larvas. Estas características, al igual que la densidad electrónica y la planaridad de los anillos resultan eficaces en la actividad de estos compuestos (Santos et al., 2010).

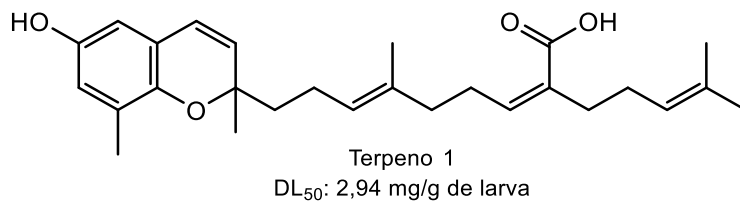
Figura 2. Estructuras químicas de derivados de eugenol con actividad larvicida frente al *Spodoptera frugiperda*.

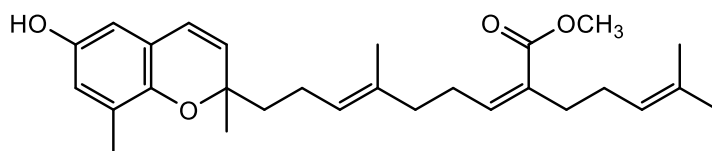




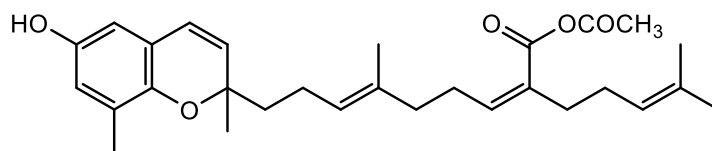
Terpenos naturales: Los compuestos seleccionados (Figura 3), son recopilados por Hernández-Carlos & Gamboa-Angulo (2019) a partir de otras investigaciones reportadas por otros autores. Se ha demostrado que el carácter lipofílico de las cadenas de carbono e hidrógeno (C-H) presentes en estructuras alifáticas o cíclicas es un factor clave para que estos compuestos se adhieran a las membranas superficiales de la epidermis de los insectos (tegumento), provocando su muerte (Govindarajan et al., 2012).

Figura 3. Estructuras químicas de terpenos naturales con actividad larvicida frente al *Spodoptera frugiperda*.

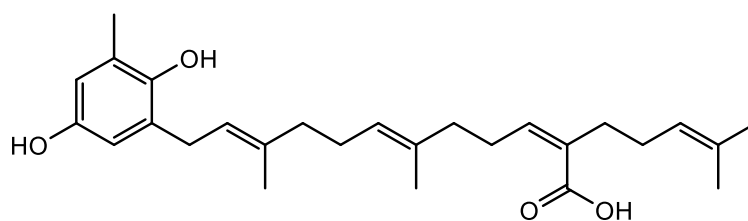




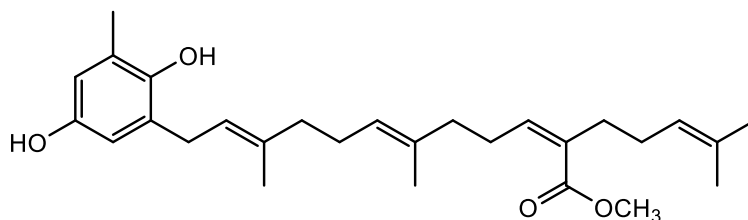
Terpeno 2
DL₅₀: 15,52 mg/g de larva



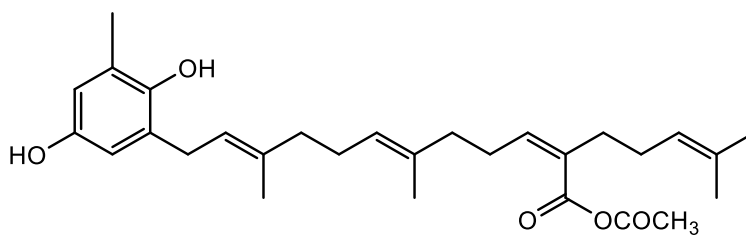
Terpeno 3
DL₅₀: 3,89 mg/g de larva



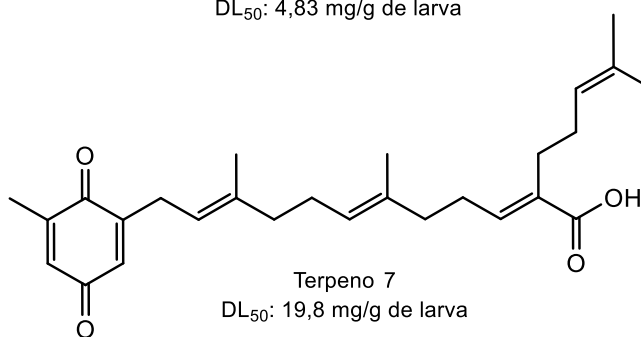
Terpeno 4
DL₅₀: 10,17 mg/g de larva



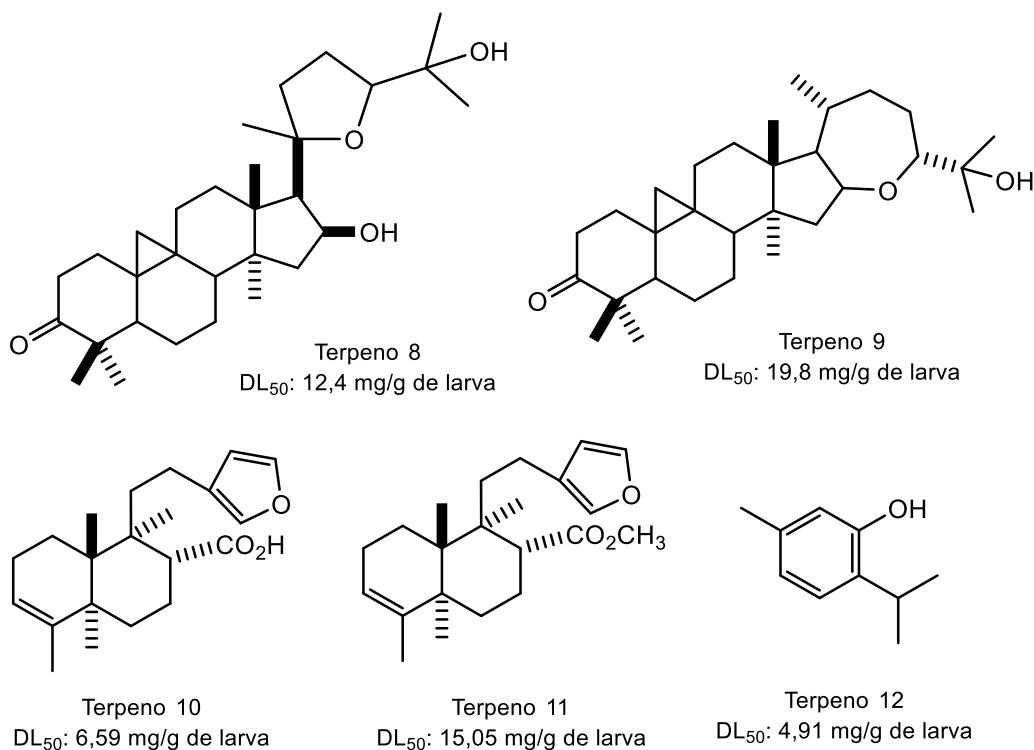
Terpeno 5
DL₅₀: 14,89 mg/g de larva



Terpeno 6
DL₅₀: 4,83 mg/g de larva

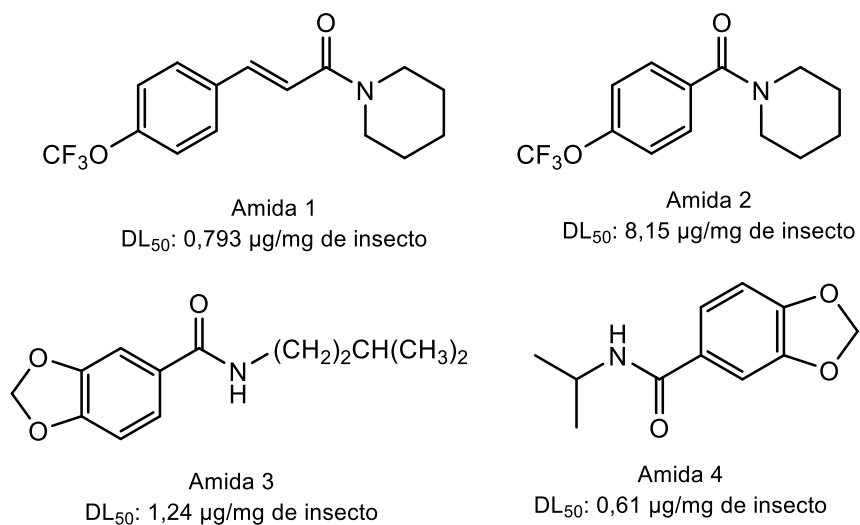


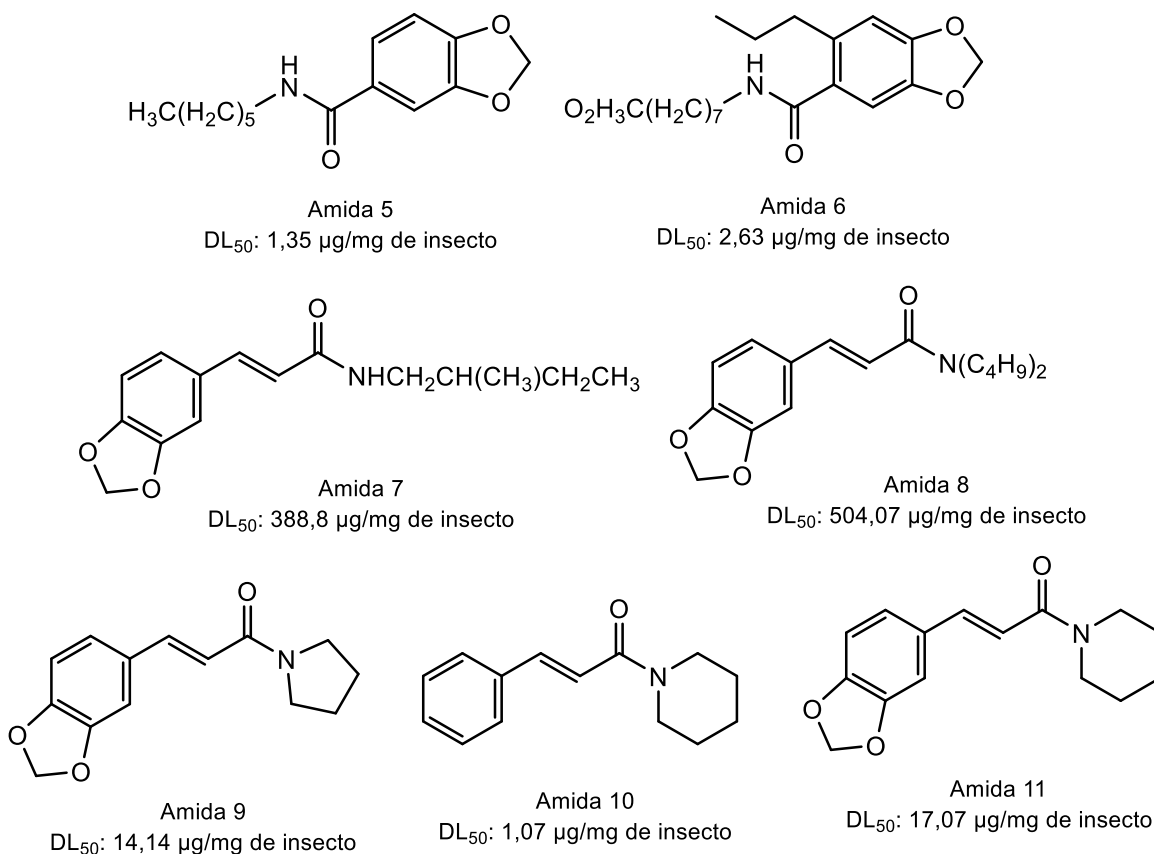
Terpeno 7
DL₅₀: 19,8 mg/g de larva



Amidas sintéticas: Del conjunto de 11 amidas sintéticas (Figura 4) analizados por Batista-Pereira et al. (2006), Castral et al. (2011) y Estrela et al. (2003) se le atribuye su actividad larvica al grupo acilo de la amida (RCONR’R’), ya que éste tiende a ser más tóxico contra las larvas. Es de señalar que la influencia de los sustituyentes amida en la actividad de los compuestos depende de las posiciones tanto en el anillo aromático como en las cadenas carbonadas abiertas y cerradas (Soberón Risco et al., 2012).

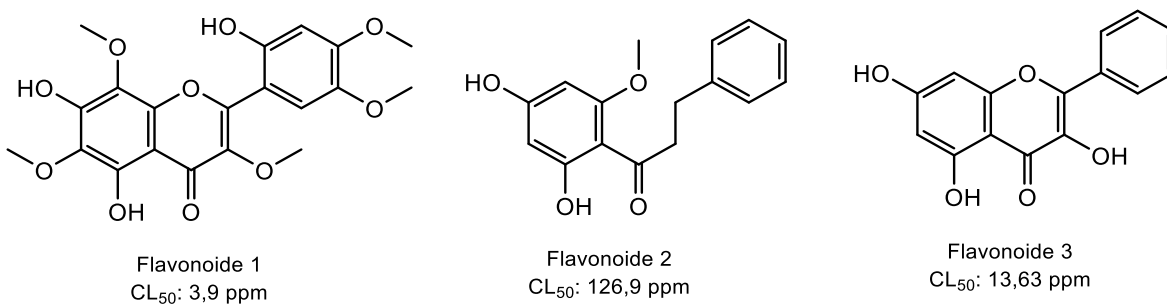
Figura 4. Estructuras químicas de amidas sintéticas con actividad larvica frente al *Spodoptera frugiperda*.

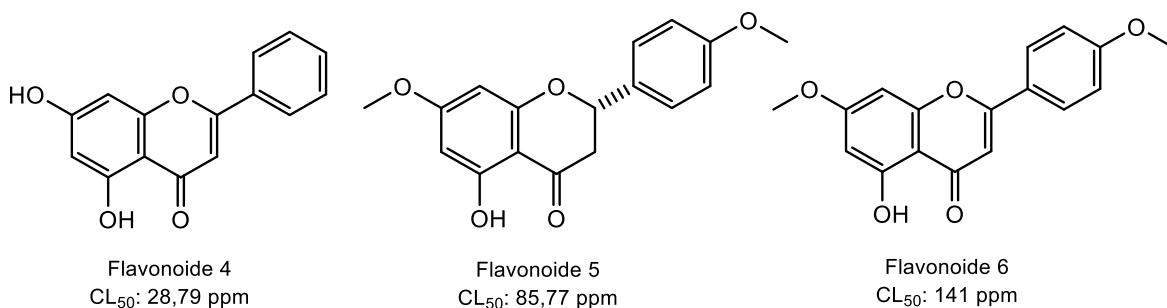




Flavonoides: Los flavonoides son otra familia de compuestos seleccionados (Figura 5), también recopilado en la revisión Hernández-Carlos & Gamboa-Angulo (2019) y Ávila-Murillo et al. (2014) su actividad se centra principalmente en la presencia de los grupos hidroxilo (-OH) y fenólicos, y a mayor número de estos grupos son más activos. Cabe resaltar que, con la presencia de sustituyentes voluminosos como metoxi (-OCH₃) (flavonoide 1) en la posición de la cromona (estructura de las flavonas) genera que se incremente la actividad (Diaz Napal & Palacios, 2015).

Figura 5. Estructuras químicas de flavonoides con actividad larvicida frente al *Spodoptera frugiperda*.

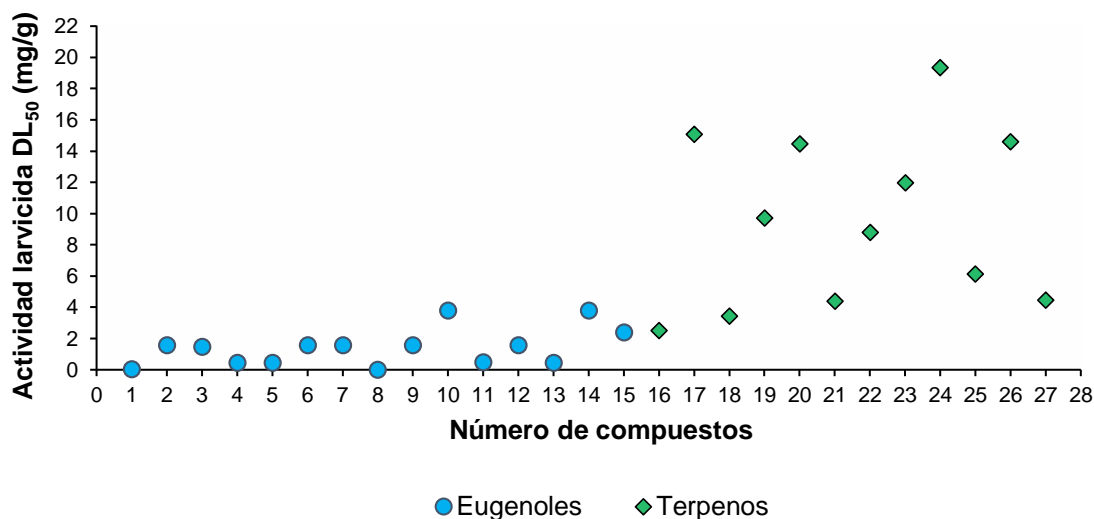




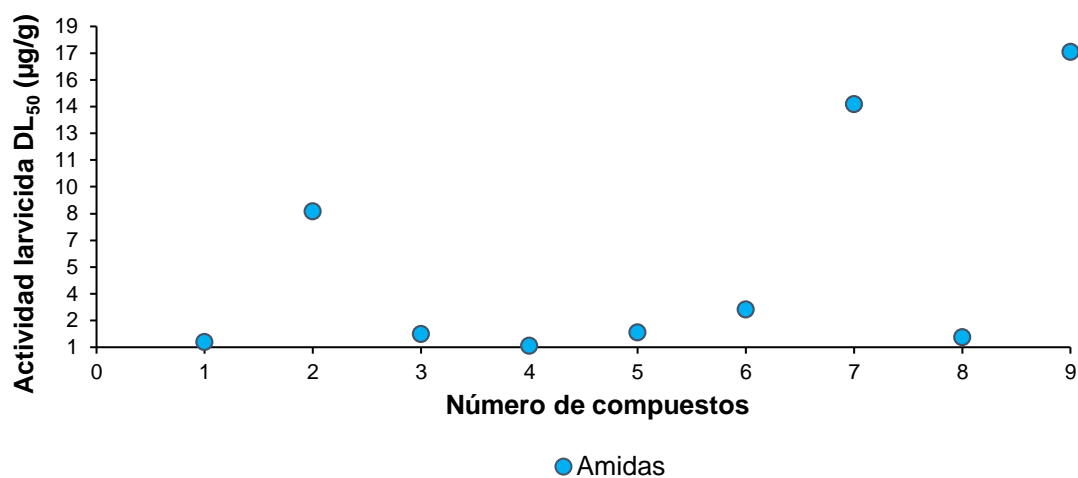
Además, se graficó la relación entre la actividad larvica vs los compuestos (Figura 6) con el propósito de analizar el impacto de los diferentes sustituyentes en dicha actividad. En la Figura 6a, se observa los derivados de eugenol y terpenos analizados, los cuales presentan una actividad larvica en un rango de 1,93 a 4,91 mg/g de insecto. Por otro lado, las amidas sintéticas muestran valores que oscilan entre 0,793 y 2,63 µg/g de insecto para la mayoría de los compuestos (Figura 6b). Cabe destacar que, para facilitar la visualización de las tendencias generales, se excluyeron de esta figura las amidas sintéticas 7 y 8, que mostraron actividades superiores a 350 µg/g de insecto. Finalmente, en el caso de los flavonoides, la actividad larvica varía entre 3,9 y 28,79 ppm (Figura 6c). No obstante, los compuestos de este grupo presentan valores variables, sin una tendencia clara e identificable.

Figura 6. Comparación de la actividad larvica de compuestos naturales y sintéticos en diferentes unidades de dosis letal (DL₅₀ y CL₅₀).

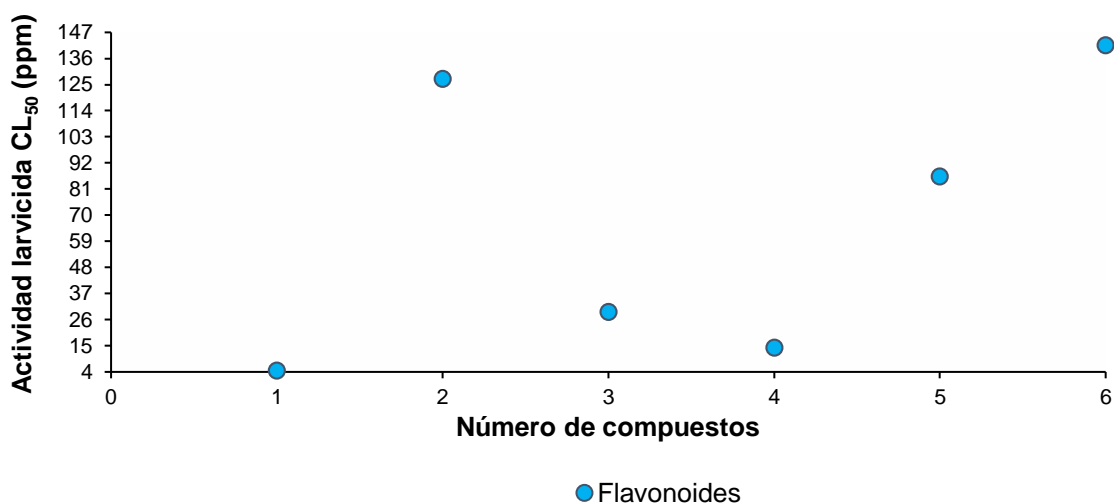
a) Figura para eugenoles y terpenos.



b) Figura para Amidas: Los datos de actividad de los compuestos 7 y 8 se eliminaron para reducir la escala Y (actividad larvica).



c) Figura para flavonoides.



Dicho lo anterior, los terpenos 17, 19, 20, 23 a 26, se encuentran por encima de 4,91 mg/g de insecto; en segundo lugar, las amidas 2, 7 y 9, se ubican por encima de 2,63 µg/g de insecto; en tercer lugar, los flavonoides 2, 5 y 6, se establecen por encima de 28,79 ppm. Éstos son los compuestos menos activos de estas familias, junto con las amidas sintéticas que se eliminaron. Cabe señalar que el grupo de los derivados de eugenol es el más activo por presentar los valores de DL₅₀ más bajos entre 0,49 y 2,83 mg/g de insecto. Así, el eugenol 1 (0,49 mg/g de insecto) es el compuesto más activo del estudio y la amida 8 (504,07 µg/g de insecto) es el compuesto menos activo.

En este trabajo se propuso seleccionar diferentes familias de compuestos para construir modelos estadísticos QSAR. Esta diversificación facilitará la aplicación de modelos a la búsqueda de compuestos de diversas estructuras químicas con actividad larvívica frente al *Spodoptera frugiperda*.

4.2 OPTIMIZACIÓN DE LA GEOMETRÍA MOLECULAR DE LOS COMPUESTOS SELECCIONADOS PARA EL ESTUDIO QSAR

A partir de la optimización de las geometrías moleculares, se obtuvieron los descriptores mecanocuánticos de la energía electrónica total, la energía de los OM's HOMO y LUMO, la diferencia de los OM's HOMO-LUMO, el momento dipolar, la entropía y la energía libre de Gibbs para emplearlos en la construcción del modelo QSAR (ver Anexo 1); y así mismo, se realizó el cálculo de las frecuencias vibracionales.

El éxito en el cálculo de frecuencias vibracionales, al obtener únicamente valores positivos, valida que las geometrías optimizadas representan mínimos de energía estables en la PES, sin modos vibracionales imaginarios (Lewars, 2016; Puzyn et al., 2010). Esto asegura la confiabilidad de los descriptores mecanocuánticos utilizados para los modelos QSAR y respalda la solidez en la estabilidad de las estructuras moleculares analizadas.

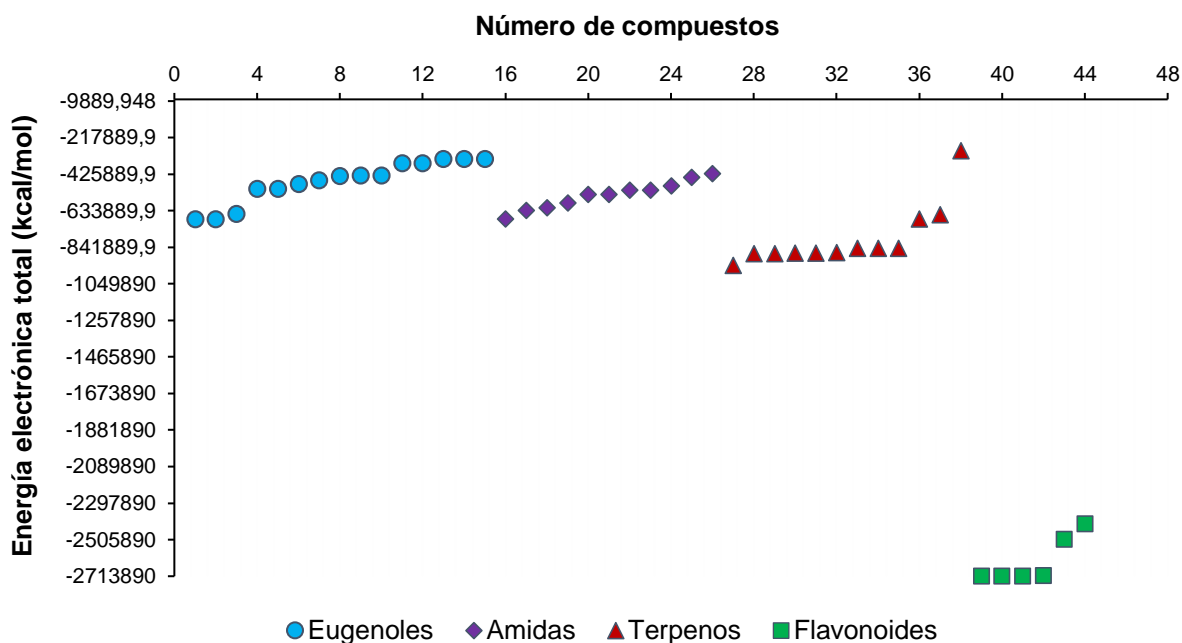
Por otra parte, el análisis Relación Estructura-Actividad (SAR) de los descriptores mecanocuánticos se llevó a cabo; sin embargo, debido a que los resultados obtenidos no demostraron ser concluyentes como variables independientes en el análisis de regresión multilínea de los modelos QSAR, no se incluyeron en el texto del trabajo. Si bien se identificaron tendencias relacionadas con las diferencias estructurales y energéticas de los compuestos estudiados, la diversidad molecular influyó en la dispersión de los datos, lo que limitó su utilidad para construir modelos QSAR consistentes. Al examinar cada familia por separado, se identificaron estructuras particulares que explican parcialmente dicha dispersión (Figuras 4 a 5).

En primer lugar, los eugenoles se caracterizan por poseer anillos aromáticos sustituidos con grupos funcionales oxigenados. La variabilidad en la posición y naturaleza de estos sustituyentes influye en la homogeneidad de los descriptores mecanocuánticos. Por su parte, los terpenos presentan una marcada diversidad estructural al combinar sistemas aromáticos y alicíclicos, cadenas carbonadas de distintas longitudes y diversos grupos oxigenados, lo que incrementa la complejidad analítica.

Así mismo, las amidas contribuyen a esta heterogeneidad estructural al estar conformadas por anillos aromáticos y cadenas alquílicas cortas, además de incorporar heteroátomos como flúor (F), nitrógeno (N) y oxígeno (O), los cuales impactan de forma significativa en los descriptores electrónicos y geométricos. Finalmente, los flavonoides, con estructuras altamente complejas formadas por múltiples anillos aromáticos interconectados y grupos funcionales oxigenados diversos, representan el mayor desafío al momento de interpretar y correlacionar los descriptores con la actividad larvívica observada.

Por otra parte, en la Figura 7 se observan los valores de la energía electrónica total de cada compuesto en un rango entre -291440 y -2713889 kcal/mol y se presentaron tendencias de forma decreciente para cada uno de los conjuntos moleculares del estudio.

Figura 7. Energía electrónica total de los compuestos del análisis QSAR.



Teniendo en cuenta que los compuestos son tan diferentes, respecto al número y los tipos de átomos era de esperar este comportamiento (Figura 7). En primer lugar, se puede observar que el grupo de los derivados de eugenol, amidas sintéticas y terpenos son los compuestos de mayor energía y, en segundo lugar, los flavonoides son los compuestos de menor energía. Sin embargo, los descriptores mecanocuánticos se conservaron para la construcción de los modelos, dado su potencial para captar información clave sobre las propiedades estructurales y energéticas de los compuestos.

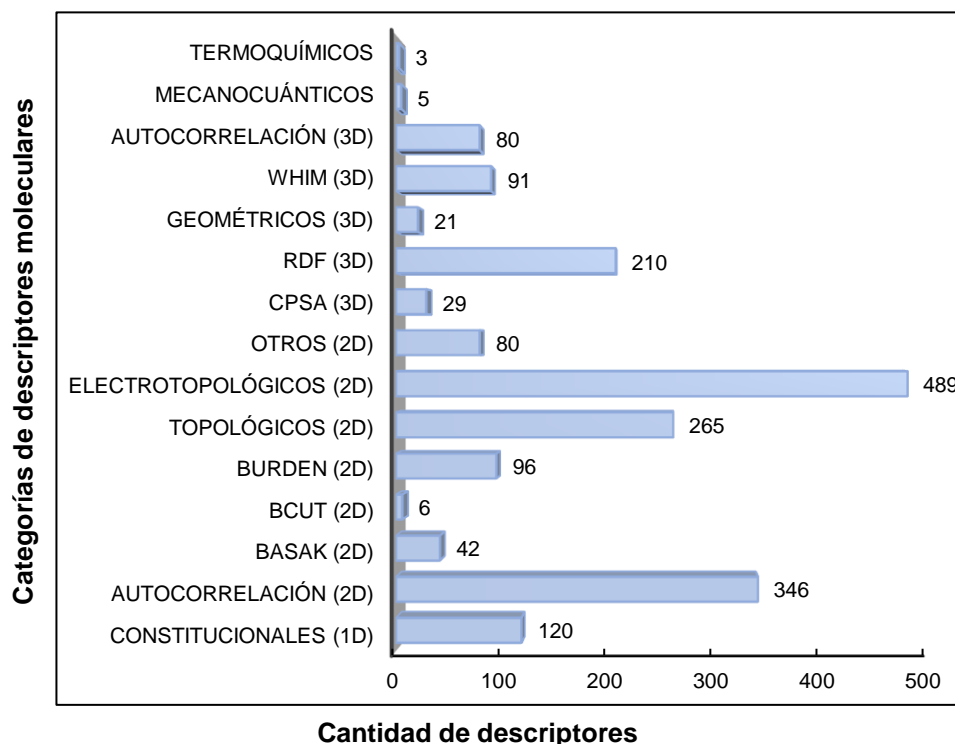
Una vez definidos los conjuntos moleculares de eugenoles, terpenos, amidas y flavonoides, y los datos experimentales de actividad larvicida, se procedió a diseñar los modelos de predicción como se describe a continuación.

4.3 CONSTRUCCIÓN DE MODELOS QSAR PARA LA PREDICCIÓN DE LA ACTIVIDAD LARVICIDA DE COMPUESTOS DE ORIGEN NATURAL Y SINTÉTICO FRENTE AL *Spodoptera frugiperda*

El estudio QSAR se desarrolló empleando la metodología reportada en el apartado 3.3, la cual consistió en el cálculo de los descriptores moleculares obtenidos a partir del programa libre PaDEL. Una vez seleccionado el conjunto de los descriptores moleculares con el método *Forward* empleando significancias estadísticas ($p < 0,01$), se realizó el tratamiento estadístico de regresión lineal múltiple (MLR); luego, la validación de los modelos por medio del método *leave-one-out* o validación interna y externa; por último, la predicción de actividad larvicida de un grupo de compuestos naturales. Cabe señalar que para la construcción de los modelos QSAR se utilizaron 37 datos de actividad larvicida, de los cuales 13 se emplearon para el modelo de los eugenoles, 9 en el modelo de amidas sintéticas, 10 para el modelo de terpenos y 5 se emplearon en la construcción del modelo de flavonoides.

4.3.1 Cálculo de los descriptores moleculares para los compuestos del análisis QSAR. A partir de la estructura optimizada de los compuestos se calcularon los 1875 descriptores moleculares, incluyendo: los descriptores mecanocuánticos de la energía electrónica total, la energía de los OMs HOMO y LUMO, la diferencia entre los OMs HOMO-LUMO, el momento dipolar, y los termoquímicos como son la energía libre de Gibbs, la entropía, la entalpía, entre otros para desarrollar el análisis estadístico MLR (Figura 8).

Figura 8. Distribución de la cantidad de descriptores moleculares por categoría y dimensionalidad.



Según la Figura 9, las categorías de descriptores 2D, particularmente los topológicos, los estados electrotopológicos y los parámetros de autocorrelación, destacan por su abundancia, lo que indica que la caracterización molecular desde una perspectiva bidimensional es predominante. En contraste, los descriptores 3D son menos numerosos, mientras que los 1D tienen una representación intermedia, situándose por debajo de las categorías más destacadas. Las categorías con un menor número de descriptores podrían estar asociadas a enfoques más específicos o áreas menos desarrolladas dentro de la metodología QSAR.

No obstante, la inclusión de todos los descriptores es fundamental, ya que cada uno codifica aspectos relevantes de la complejidad molecular. La finalidad del modelado QSAR es identificar combinaciones de parámetros teóricos que expliquen de manera óptima la actividad larvicida. Esta diversidad en la información estructural constituye la base para construir modelos estadísticos robustos, predictivos y generalizables (Muhammad et al., 2018; Rácz et al., 2019).

4.3.2 Modelo QSAR de actividad larvica para derivados del eugenol. Con los descriptores seleccionados, se llevó a cabo el tratamiento estadístico, utilizando el análisis MLR. Teniendo en cuenta la multicolinealidad, que es la intercorrelación que existe entre los descriptores de un modelo evidenciando aspectos similares de la actividad (Muratov et al., 2020; Roy et al., 2008), se aplicó el submétodo *Forward* del análisis MLR. Este método consistió en seleccionar aquellas variables independientes (descriptores moleculares) con mayor significancia estadística (p) e introducirlas secuencialmente en el modelo QSAR (Draper & Smith, 1998; Pirhadi et al., 2015).

De este modo, se seleccionaron los descriptores con niveles de significancia (α) en el intervalo $0,00 \leq \alpha \leq 0,10$, obteniendo 12 descriptores de los 1875 descriptores seleccionados inicialmente (Anexo 2), de los cuales 6 presentaron α de 0,00 (nssO, MATS3m, Entropía, GATS3v, ATSC6e y AATSC8e). Con estos 6 descriptores se construyó el siguiente modelo estadístico (Tabla 2). La significancia estadística observada en α de 0,00 asegura la robustez y validez de los resultados obtenidos. De otro lado, los demás descriptores con $\alpha \geq 0,0118$ no aportan mayor peso estadístico con relación a la distribución de Fisher (F) (Devore, 2007; Grisoni et al., 2018).

Tabla 2. ANOVA de los descriptores incluidos con el mayor peso estadístico al modelo QSAR de eugenoles para actividad larvica.

Descriptores moleculares	R ²	EER	F	p
nssO	0,4032	0,2407	9,108	0,0117
Log₁₀(DL₅₀) = -0,4845 + 0,3780 [nssO]				
nssO, MATS3m	0,5492	0,2092	8,31	0,007481
Log₁₀(DL₅₀) = -0,5860 + 0,4486 [nssO] + 2,3997 [MATS3m]				
nssO, MATS3m, Entropía	0,6605	0,1815	8,782	0,004881
Log₁₀(DL₅₀) = -0,5865 + 0,4326 [nssO] + 2,4134 [MATS3m] - 4,851x10⁻⁴ [Entropía]				
nssO, MATS3m, Entropía, GATS3v	0,8383	0,1253	16,55	0,0006172
Log₁₀(DL₅₀) = -6,4991 + 0,3366 [nssO] + 9,3887 [MATS3m] - 7,263x10⁻⁴ [Entropía] + 5,6209 [GATS3v]				
nssO, MATS3m, Entropía, GATS3v, ATSC6e	0,9714	0,05272	82,4	4,587x10 ⁻⁶
Log₁₀(DL₅₀) = -9,0670 + 0,4043 [nssO] + 12,4173 [MATS3m] - 7,829x10⁻⁴ [Entropía] + 7,8698 [GATS3v] - 0,2682 [ATSC6e]				
nssO, MATS3m, Entropía, GATS3v, ATSC6e, AATSC8e	0,9957	0,02044	463,6	9,936x10 ⁻⁸
Log₁₀(DL₅₀) = -7,957 + 2,905x10⁻¹ [nssO] + 1,139x10¹ [MATS3m] - 7,735x10⁻⁴ [Entropía] + 7,131 [GATS3v] - 1,608x10⁻¹ [ATSC6e] + 1,066x10¹ [AATSC8e]				

Como se puede apreciar en la Tabla 2, se resume la construcción del modelo QSAR para predecir la actividad larvica de los eugenoles analizados, mostrando cómo la inclusión de cada descriptor molecular mejora progresivamente la calidad del modelo. En primer lugar, al incluir en el modelo el descriptor nssO, presenta un coeficiente de correlación (R²) moderado de 0,4032, un error estándar residual (EER) de 0,2407 y un coeficiente estadístico Fisher (F) de 9,108. Esto indica una capacidad predictiva inicial aceptable pero limitada.

Con la inclusión de más descriptores, como MATS3m, Entropía y GATS3v, el modelo mejora sustancialmente, evidenciándose en el incremento de R^2 de 0,4032 a 0,9957. De manera paralela, el EER disminuye considerablemente, desde 0,2407 hasta 0,02044, indicando una mayor precisión en las predicciones realizadas por el modelo estadístico. La distribución Fisher (F), que evalúa la significancia global del modelo, presenta un comportamiento fluctuante, disminuyendo en los intervalos iniciales (de 9,108 a 8,31) pero aumentando drásticamente en el modelo con la inclusión del descriptor AATSC8e. Este incremento notable en el F de 463,6 sugiere una contribución significativa de este descriptor al modelo QSAR, consolidando su utilidad para describir la relación entre las propiedades moleculares y la actividad larvica.

Por otro lado, los descriptores mecanocuánticos no fueron incluidos en los modelos finales debido a su baja correlación y falta de niveles de significancia estadística (p) ($\alpha \geq 0,10$), indicando que no aportan valor en la predicción de la actividad larvica (Anexo 3). Sin embargo, los resultados reflejan que el modelo QSAR con seis descriptores moleculares (nssO, MATS3m, Entropía, GATS3v, ATSC6e y AATSC8e) es más robusto, preciso y estadísticamente significativo. Su capacidad predictiva es respaldada por un alto coeficiente de correlación ($R^2 = 0,9957$) y un bajo EER (0,02044) y una significancia estadística considerablemente alta ($p = 9,936 \times 10^{-8}$). Esto evidencia que este modelo QSAR es una herramienta confiable para la predicción de la actividad larvica de los compuestos analizados.

Cabe destacar que el descriptor molecular AATSC8e es aquel que mejor describe la actividad larvica en el modelo QSAR de eugenoles por su alto coeficiente de F, seguido de los descriptores moleculares ATSC6e y GATS3v. Así, el nssO es un descriptor 2D de estado electrotopológico, que está basado en el recuento de estados electrotopológicos para átomos de oxígeno (Xia et al., 2018); el MATS3m es un descriptor clasificado como 2D-Autocorrelación, y está relacionado con la autocorrelación de Moran en retraso 3, basado en la ponderación por masa atómica (Xia et al., 2019). La Entropía es un descriptor termoquímico asociado a los estados probables de un sistema molecular, que puede evolucionar al ser sometido a perturbaciones. El aumento del desorden es una característica propia de la fase gaseosa (Dehmer et al., 2009); el GATS3v es un descriptor 2D-Autocorrelación, basado en la autocorrelación de Geary por retraso 3 y ponderado por el volumen de Van Der Waals (Arthur et al., 2022); el ATSC6e es un descriptor clasificado como 2D-Autocorrelación, y está relacionado con la autocorrelación Broto-Moreau en desfase 6 en función logarítmica, basado en la ponderación por la electronegatividad (Rezić et al., 2022). Por otra parte, el AATSC8e es un descriptor del grupo de Autocorrelación, basado en la centralización media Broto-Moreau por retraso 8 y ponderado a través de la electronegatividad de Sanderson (Pingaew et al., 2021). En el Anexo 4 se aprecian los cálculos de los descriptores moleculares.

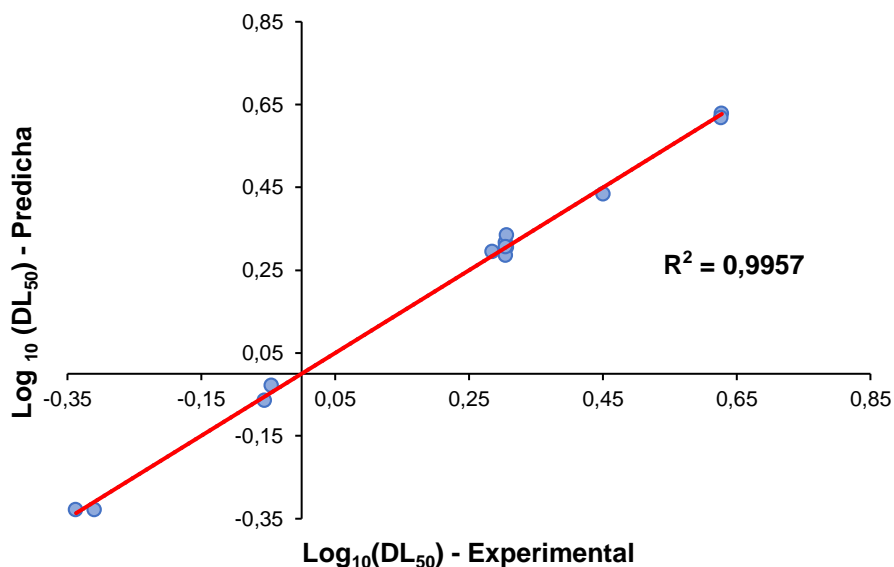
La Tabla 3 presenta el modelo QSAR consolidado para predecir la actividad larvica de derivados del eugenol. Los resultados obtenidos evidencian la alta precisión y significancia estadística del modelo (Veerasingam et al., 2011). Los principales parámetros estadísticos son: un valor de F calculado (F_{cal}) de 463,6; un coeficiente R de 0,9978 y un coeficiente de R^2 de 0,9957; lo que indica que el modelo explica el 99,57% de la variabilidad en los datos experimentales. Además, el EER es de 0,02044; respaldando la estabilidad del modelo, y la significancia p de $9,936 \times 10^{-8}$, lo que confirma su capacidad predictiva y fiabilidad.

Tabla 3. Medidas estadísticas del modelo QSAR de derivados del eugenol para actividad larvícida.

Parámetros estadísticos		
F = 463,6	R = 0,9978	R ² = 0,9957
GL = 6	EER = 0,02044	p = 9,936x10 ⁻⁸
Ecuación matemática de predicción de la actividad		
$\text{Log}_{10}(\text{DL}_{50}) = -7,957(\pm 1,845) + 2,905 \times 10^{-1}(\pm 0,107) [\text{nssO}] + 1,139 \times 10^1(\pm 2,150) [\text{MATS3m}] - 7,735 \times 10^{-4}(\pm 1,500 \times 10^{-3}) [\text{Entropía}] + 7,131(\pm 1,628) [\text{GATS3v}] - 1,608 \times 10^{-1}(\pm 0,116) [\text{ATSC6e}] + 1,066 \times 10^1(\pm 8,192) [\text{AATSC8e}]$		

A partir de la ecuación matemática obtenida en la Tabla 3, los coeficientes presentan intervalos de confianza al 95% (\pm), lo que asegura la precisión del ajuste del modelo QSAR. Por ejemplo, el descriptor nssO tiene un valor de $2,905 \times 10^{-1} \pm 0,107$, mostrando una baja dispersión y alta estabilidad en contribuir el carácter predictivo al modelo. En contraste, descriptores como el AATSC8e presentan mayor error estándar ($1,066 \times 10^1 \pm 8,192$), lo que sugiere menor precisión en su aporte a la predicción. Sin embargo, el análisis estadístico respalda la fiabilidad del modelo, ya que el valor calculado de F (463,6) supera ampliamente el valor tabulado F_{tab} (2,9153) con grados de libertad 6 y 13, respectivamente. Este resultado, junto con un bajo EER y un alto coeficiente R², confirma que los descriptores moleculares seleccionados tienen un impacto significativo en la predicción de la actividad larvícida (Leach & Gillet, 2007; Triola, 2004).

Figura 9. Representación gráfica de la actividad larvícida experimental frente a la predicha para el modelo QSAR de derivados del eugenol.



Ahora bien, al graficar los valores de $\text{Log}_{10}(\text{DL}_{50})$ experimental Vs $\text{Log}_{10}(\text{DL}_{50})$ se observa que el modelo de predicción presentó buena correlación entre los datos experimentales y predichos de actividad larvícida (Anexo 5), demostrando la capacidad de predicción del

modelo (Figura 9) (Gadaleta et al., 2016; Golbraikh et al., 2017). Sin embargo, se presentaron algunos valores negativos de $\text{Log}_{10}(\text{DL}_{50})$ para DL_{50} entre 0,46 y 0,90 mg/g de insecto; puesto que son datos de actividad mayores que cero y menores que uno. Cabe destacar que la capacidad de predicción del modelo QSAR de eugenoles obtenido al 95% de confianza es buena porque evidencia medidas de Fisher y una significancia estadística muy altas.

4.3.3 Modelo QSAR de actividad larvicida para amidas sintéticas. La construcción del modelo estadístico se realizó con la metodología 3.3. Realizando una correlación lineal de los descriptores moleculares Vs la actividad larvicida, se obtuvieron 1875 descriptores moleculares para desarrollar el análisis estadístico MLR. Con los descriptores seleccionados se llevó a cabo el tratamiento estadístico, utilizando el análisis mencionado anteriormente. Para tratar la multicolinealidad presentada, se aplicó el submétodo *Forward* del análisis MLR (Muratov et al., 2020; Pirhadi et al., 2015).

De este modo, se seleccionaron los descriptores con α en el intervalo $0,00 \leq \alpha \leq 0,10$, obteniendo 2 descriptores de los 1875 descriptores seleccionados (Anexo 6), de los cuales presentaron α de 0,00 el ALogP y ZMIC1. Estos descriptores, al ser los únicos que alcanzaron el nivel de significancia más alto, sugieren una fuerte asociación con la actividad larvicida. De otra parte, los demás descriptores con $\alpha \geq 1,143 \times 10^{-5}$ no aportan mayor peso estadístico con respecto a la distribución de Fisher-Snedecor (F), por lo que no son significativos para predecir dicha actividad larvicida (Devore, 2007; Grisoni et al., 2018). En la Tabla 4, se muestra el análisis ANOVA con los dos descriptores seleccionados inicialmente, el cual se observa que, al ingresar el segundo descriptor al modelo, los coeficientes F y p disminuyen restando significancia estadística, aunque los valores de R^2 y EER no variaran considerablemente.

Tabla 4. ANOVA de los descriptores incluidos con el mayor peso estadístico al modelo QSAR de amidas sintéticas para actividad larvicida.

Descriptores moleculares	R^2	EER	F	p
ALogP	0,9374	0,216	120,8	$1,143 \times 10^{-5}$
$\text{Log}_{10}(\text{DL}_{50}) = 0,77719 + 1,01276 [\text{ALogP}]$				
ALogP, ZMIC1	0,9317	0,2258	55,53	$1,347 \times 10^{-4}$
$\text{Log}_{10}(\text{DL}_{50}) = 0,34526 + 0,99877 [\text{ALogP}] + 0,01093 [\text{ZMIC1}]$				

Como se muestra en los descriptores obtenidos (Tabla 4), no se encontraron los descriptores mecanocuánticos ni termoquímicos en el modelo QSAR debido a que no presentaron una correlación adecuada ni una significancia estadística relevante con la actividad larvicida (Anexo 7). Esto indica que dichos descriptores teóricos no aportan información suficiente para predecir con precisión la variable de interés.

Adicionalmente, se observa que al ingresar el segundo descriptor al modelo QSAR, se produce una disminución en el parámetro de F (de 120,8 a 55,53) y en la correlación R^2 baja (de 0,9374 a 0,9317), mientras que el EER aumenta ligeramente (de 0,216 a 0,2258). Estos cambios reflejan que la inclusión del segundo descriptor no contribuye de manera significativa a mejorar el modelo en comparación con el primer descriptor molecular

(ALogP). A pesar de lo anterior, el modelo QSAR muestra una buena capacidad predictiva de la actividad larvícida, respaldada por el valor del coeficiente R^2 , que permanece elevado.

Para este caso, el descriptor molecular ALogP es aquel que posee mayor peso estadístico en el modelo de actividad larvícida por su alto coeficiente de F, seguido del descriptor ZMIC1, el cual no se consideró para construir el modelo QSAR. Así pues, la variable ALogP, que es un descriptor molecular constitucional está basado en el cálculo LogP aditivo atómico y los valores molares de refractividad descritos por Ghose y Crippen (Toppur & Jaims, 2021). En el Anexo 8 se muestran los cálculos de los descriptores.

La Tabla 5 presenta el modelo QSAR consolidado para la predicción de la actividad larvícida de amidas sintéticas. Los resultados obtenidos demostraron que el modelo es estadísticamente robusto y preciso, lo cual se refleja en los parámetros estadísticos calculados: el coeficiente de determinación (R^2) fue de 0,9374; indicando que el modelo explica el 93,74% de la variabilidad de los datos experimentales; el error estándar residual (EER) fue bajo, con un valor de 0,216; el valor de F calculado (F_{cal}) alcanzó 120,8; superando ampliamente el valor tabulado ($F_{tab} = 3,2927$) para los grados de libertad considerados ($GL = 7$). Estos resultados, acompañados de una significancia estadística (p de $1,143 \times 10^{-5}$), respaldan la fiabilidad del modelo. Así mismo, los intervalos de confianza al 95% (\pm) para los coeficientes de la ecuación matemática muestran un margen estrecho (intervalos $\leq \pm 1,0$); lo que refuerza la precisión del modelo estadístico. Dichos intervalos reflejan que la variabilidad de los coeficientes es baja y que los valores estimados son confiables (Leach & Gillet, 2007; Triola, 2004).

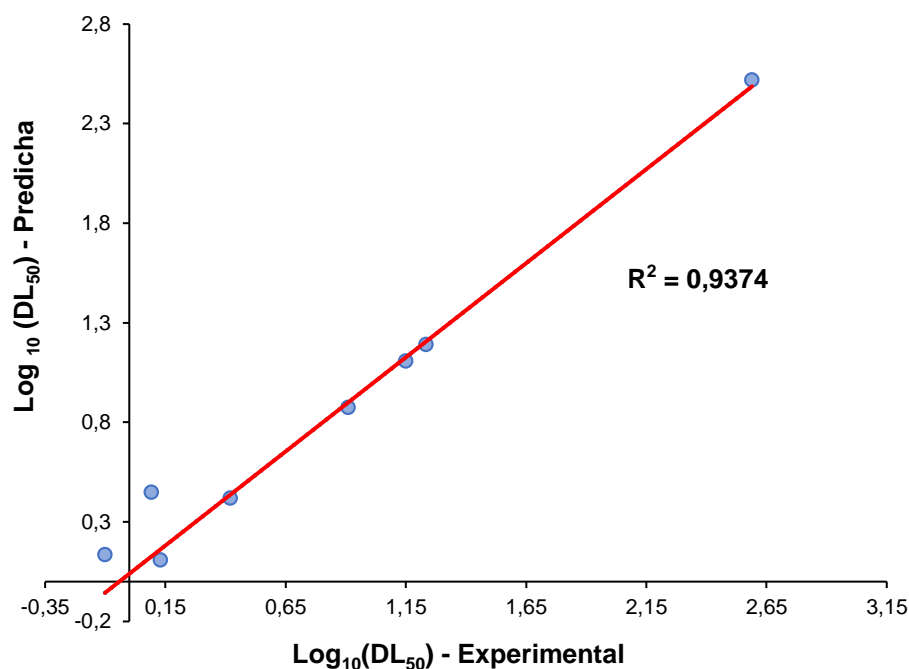
Tabla 5. Medidas estadísticas del modelo QSAR de amidas sintéticas para actividad larvícida.

Parámetros estadísticos		
F = 120,8	R = 0,9681	$R^2 = 0,9374$
GL = 7	EER = 0,216	$p = 1,143 \times 10^{-5}$
Ecuación matemática de predicción de la actividad		
$\text{Log}_{10}(\text{DL}_{50}) = 0,77719(\pm 0,341) + 1,01276(\pm 0,436) [\text{ALogP}]$		

Por otra parte, la ecuación matemática refleja la influencia directa del descriptor molecular ALogP en la actividad larvícida, donde el signo positivo del coeficiente indica que valores mayores de ALogP incrementan la actividad predicha de los compuestos. La combinación de un alto valor de R^2 , un EER bajo y una significancia estadística alta confirma que el modelo QSAR presenta un ajuste óptimo y una buena capacidad predictiva. Además, el valor de F calculado, considerablemente mayor que el tabulado, respalda la validez del modelo en la predicción de los valores de la dosis letal media (DL_{50}). Finalmente, se observó una buena correlación entre los datos experimentales y los valores predichos de la actividad larvícida (Anexo 9), lo que valida la precisión y robustez del modelo estadístico.

Ahora bien, se observa que el modelo de predicción para amidas sintéticas presentó buena correlación entre los datos experimentales y predichos de la actividad larvícida (Anexo 9), demostrando la capacidad de predicción del modelo (Figura 10).

Figura 10. Representación gráfica de la actividad larvica experimental frente a la predicha para el modelo QSAR de amidas sintéticas.



En la gráfica anterior se observan menos valores negativos de $\text{Log}_{10}(\text{DL}_{50})$ en comparación con el modelo QSAR de eugenoles. En cambio, en el modelo QSAR solo se presentó un valor de $\text{Log}_{10}(\text{DL}_{50})$ negativo para DL_{50} de 1,07 $\mu\text{g}/\text{mg}$ de larva porque es un dato mayor que cero y cercano a uno, entonces el logaritmo da un valor negativo o se dice que es un logaritmo negativo.

4.3.4 Modelo QSAR de actividad larvica para los terpenos. La construcción del modelo de actividad larvica se realizó con la metodología 3.3. Realizando una correlación lineal de los descriptores moleculares Vs la actividad larvica, se obtuvieron 1875 descriptores moleculares para desarrollar el análisis estadístico MLR. Con los descriptores moleculares seleccionados se llevó a cabo el tratamiento estadístico, utilizando el análisis MLR. Para tratar la multicolinealidad presentada, se aplicó el submétodo *Forward* del análisis MLR (Muratov et al., 2020; Pirhadi et al., 2015).

Así pues, se seleccionaron los descriptores con α en el intervalo $0,00 \leq \alpha \leq 0,10$, obteniendo 9 descriptores de los 1875 descriptores seleccionados (Anexo 10), de los cuales 5 presentaron α de 0,00 (GATS2m, MLFER_A, E1e, RDF140v, TDB2u). Este resultado refleja una relación estadísticamente garantizada entre estos descriptores y la actividad larvica, ya que un valor de nivel de significancia $\alpha = 0,00$ implica que la probabilidad de que la relación sea atribuida al azar es nula. En este sentido, estos descriptores representan las variables más relevantes y explicativas de la actividad larvica, lo que refuerza la confianza en la validez del modelo. De otro lado, los demás descriptores con $\alpha \geq 0,001465$ no aportan mayor peso estadístico con respecto a la distribución de Fisher-Snedecor (F). Con estos 5 descriptores se construyó el siguiente modelo QSAR de actividad larvica (Tabla 6).

Tabla 6. ANOVA de los descriptores incluidos con el mayor peso estadístico al modelo QSAR de terpenos para actividad larvica.

Descriptores moleculares	R ²	EER	F	p
GATS2m	0,7046	0,1546	22,46	0,001465
Log₁₀(DL₅₀) = -4,921 + 6,915 [GATS2m]				
GATS2m, MLFER_A	0,8254	0,1189	22,27	0,0009227
Log₁₀(DL₅₀) = -4,85417 + 7,02091 [GATS2m] - 0,2355 [MLFER_A]				
GATS2m, MLFER_A, E1e	0,8641	0,1049	20,08	0,001569
Log₁₀(DL₅₀) = -4,9439 + 6,6292 [GATS2m] - 0,3554 [MLFER_A] + 0,9336 [E1e]				
GATS2m, MLFER_A, E1e, RDF140v	0,9426	0,06815	37,96	0,0006208
Log₁₀(DL₅₀) = -4,6401 + 6,1414 [GATS2m] - 0,3474 [MLFER_A] + 1,2732 [E1e] - 0,03092 [RDF140v]				
GATS2m, MLFER_A, E1e, RDF140v, TDB2u	0,9996	0,00576	4390	1,452x10 ⁻⁷
Log₁₀(DL₅₀) = -1,310x10¹ + 5,072 [GATS2m] - 3,684x10⁻¹ [MLFER_A] + 2,201[E1e] - 4,346x10⁻² [RDF140v] + 4,053 [TDB2u]				

Como se observa en los descriptores seleccionados en la Tabla 6, no se encontraron los descriptores mecanocuánticos ni termoquímicos. Esto se debe a que dichos descriptores no mostraron una correlación significativa con la actividad larvica, lo que los hace estadísticamente irrelevantes para este estudio, evitando redundancias (Anexo 11). Adicionalmente, se evidencia que, al incluir sucesivamente los descriptores más relevantes, la calidad del modelo QSAR mejora notablemente, asegurando más eficiencia y precisión. A medida que se agregan los tres primeros descriptores (GATS2m, MLFER_A y E1e), el coeficiente de determinación (R²) aumenta progresivamente de 0,7046 a 0,9426; indicando una mejor correlación con la actividad larvica. Paralelamente, el error estándar residual (EER) disminuye de 0,1546 a 0,06813; reflejando una mayor precisión del modelo.

La incorporación del descriptor RDF140v y, finalmente, TDB2u fortalece aún más el modelo estadístico, elevando el R² hasta 0,9996 y reduciendo el EER a 0,00576. Este comportamiento indica que cada descriptor adicional aporta significativamente al ajuste del modelo y respalda su relevancia en la predicción de la actividad larvica. El análisis ANOVA refleja valores altos de distribución de Fisher (F) y de significancia estadística (p) menores a 0,05 para cada paso del modelo, validando la significancia estadística de los descriptores seleccionados. Por ejemplo, al incluir el descriptor TDB2u, se alcanza un valor F de 4390, con una significancia p de 1,452x10⁻⁷, demostrando su impacto significativo en el ajuste del modelo. Este descriptor molecular es el que presenta mayor peso estadístico, seguido del descriptor RDF140v.

Para comprender mejor la contribución de los descriptores seleccionados, es importante destacar sus características principales. Así, el descriptor GATS2m, que es un descriptor 2D-Autocorrelación, basado en la ponderación por masas atómicas relacionado con el factor GATS por retraso 2 (Put et al., 2003); el MLFER_A es un descriptor clasificado 2D-MLFER, y está relacionado con la acidez global o sumatoria del enlace de hidrógeno de un soluto (Abdufatai et al., 2020); el descriptor E1e es un descriptor clasificado como 3D-WHIM, y está relacionado con la accesibilidad direccional por componentes primarios,

basado en la ponderación por electronegatividades relativas de Sanderson (Bijlsma et al., 2017); el descriptor RDF140v, que es un descriptor 3D-RDF, basado en una función de distribución radial en función 140 relacionado con la ponderación relativa del volumen de Van Der Waals (Arthur et al., 2016). Por otro lado, el descriptor TDB2u es un descriptor del grupo 3D-Autocorrelación, basado en la autocorrelación de la distancia topológica por retraso 2 no ponderado (Kobayashi et al., 2020). En el Anexo 12 se muestran los cálculos de los descriptores moleculares.

Los parámetros estadísticos del modelo QSAR, presentados en la Tabla 7, incluyen intervalos al 95% (\pm) de confianza para cada descriptor molecular y el intercepto, lo que refuerza la confiabilidad de los coeficientes estimados. Estos intervalos indican la estabilidad y precisión en las estimaciones, ya que refleja una baja variabilidad o incertidumbre en los datos (intervalos $\leq \pm 2,0$). Además, el hecho de que los intervalos de confianza no incluyan el valor cero para ninguno de los descriptores seleccionados confirma su significancia estadística. Esto asegura que cada descriptor molecular tiene una influencia significativa y relevante en la predicción de la actividad larvica, eliminando la posibilidad de que su contribución al modelo sea aleatoria (Leach & Gillet, 2007; Triola, 2004).

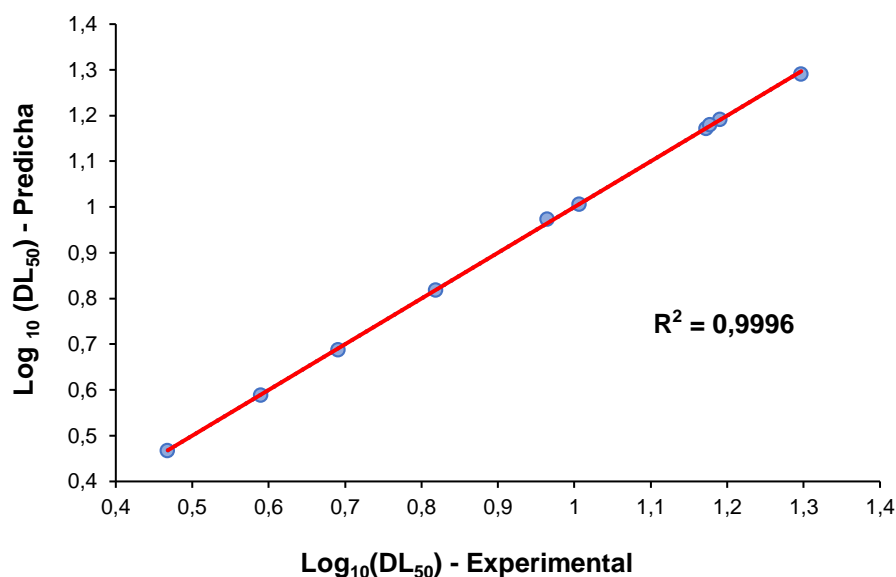
Tabla 7. Medidas estadísticas del modelo QSAR de terpenos naturales para actividad larvica.

Parámetros estadísticos		
F = 4390	R = 0,9998	R ² = 0,9996
GL = 4	EER = 0,00576	p = 1,452x10 ⁻⁷
Ecuación matemática de predicción de la actividad		
$\text{Log}_{10}(\text{DL}_{50}) = -1,310 \times 10^1 (\pm 1,800) + 5,072 (\pm 0,390) [\text{GATS2m}] - 3,684 \times 10^{-1} (\pm 0,032) [\text{MLFER_A}] + 2,201 (\pm 0,260) [\text{E1e}] - 4,346 \times 10^{-2} (\pm 5,00 \times 10^{-3}) [\text{RDF140v}] + 4,053 (\pm 0,853) [\text{TDB2u}]$		

La ecuación obtenida presenta un coeficiente de R² de 0,9996; un EER de 0,00576 y un valor F calculado de 4390. El alto valor de R² indica que el modelo explica casi el 100% de la variabilidad en la actividad larvica, mientras que el bajo EER demuestra una alta precisión en las predicciones. El análisis de la distribución de Fisher refuerza esta conclusión: el valor de p de 0,00 muestra una alta significancia estadística, y la comparación entre F_{cal} y el valor de Fisher tabulado F_{tab} = 3,4780 confirma que F_{cal} >> F_{tab} para GL1 = 4 y GL = 10 (Leach & Gillet, 2007; Triola, 2004).

Ahora bien, se observa que el modelo de predicción para terpenos presentó buena correlación entre los datos experimentales y predichos de actividad larvica (Anexo 13), demostrando la capacidad de predicción del modelo (Figura 11). Cabe destacar que la capacidad de predicción del modelo QSAR para terpenos obtenido al 95% de confianza es buena porque evidencia medidas de Fisher y una significancia estadística muy altas.

Figura 11. Representación gráfica de la actividad larvica experimental frente a la predicha para el modelo QSAR de terpenos.



4.3.5 Modelo QSAR de actividad larvica para flavonoides. La construcción del modelo estadístico se realizó con la metodología 3.3. Realizando una correlación lineal de los descriptores moleculares Vs la actividad larvica, se obtuvieron 1875 descriptores moleculares para desarrollar el análisis estadístico MLR. Con los descriptores seleccionados se llevó a cabo el tratamiento estadístico, utilizando el análisis dicho anteriormente. Para tratar la multicolinealidad presentada, se aplicó el submétodo *Forward* del análisis MLR (Muratov et al., 2020; Pirhadi et al., 2015).

Por tanto, se seleccionaron los descriptores con α en el intervalo $0,00 \leq \alpha \leq 0,10$, obteniendo 2 descriptores de los 1875 descriptores seleccionados (Anexo 14), de los cuales presentaron α de 0,00: el ATSC3s y RotBFrac. Este criterio de selección garantiza la significancia estadística al asegurar que únicamente se consideran relevantes aquellos descriptores con una probabilidad de error inferior al 10%, lo que minimiza la posibilidad de presentar colinealidades entre ellos. De otra parte, los demás descriptores con $\alpha \geq 3,224 \times 10^{-5}$ no aportan mayor peso estadístico significativo con respecto a la distribución de Fisher-Snedecor (F) (Devore, 2007; Grisoni et al., 2018). En la Tabla 8, se muestra el análisis ANOVA con los dos descriptores seleccionados inicialmente, el cual se observa que, al ingresar el segundo descriptor al modelo, los coeficientes F y p disminuyen restando significancia estadística, aunque los valores de los coeficientes R^2 y EER no varían considerablemente.

Tabla 8. ANOVA de los descriptores incluidos con el mayor peso estadístico al modelo QSAR de flavonoides para actividad larvica.

Descriptores moleculares	R^2	EER	F	p
ATSC3s	0,9976	0,03369	1670	$3,224 \times 10^{-5}$
$\text{Log}_{10}(\text{CL}_{50}) = 1,150702 - 0,041060 [\text{ATSC3s}]$				
ATSC3s, RotBFrac	0,9967	0,03984	597,2	0,001672
$\text{Log}_{10}(\text{CL}_{50}) = 1,165140 - 0,041140 [\text{ATSC3s}] - 0,105937 [\text{RotBFrac}]$				

Como se muestra en los descriptores obtenidos (Tabla 8), los descriptores mecanocuánticos y termoquímicos no demostraron una buena correlación ni significancia estadística con la actividad larvícida, lo que indica que su capacidad para explicar o predecir esta propiedad es limitada (Anexo 15). En términos estadísticos, estos descriptores no contribuyen de manera significativa al modelo QSAR debido a sus bajos valores de F y de niveles de significancia. Adicionalmente, al incorporar un segundo descriptor al modelo QSAR, se observa una disminución en el parámetro F (de 1670 a 597,2), mientras que la correlación R^2 experimenta un leve descenso (de 0,9976 a 0,9967) y el error estándar residual (EER) aumenta ligeramente (de 0,03369 a 0,03984). Esto evidencia que la inclusión de descriptores adicionales, como RotBFrac, no aporta mejoras significativas al modelo en comparación con el descriptor ATSC3s, el cual ya posee un peso estadístico considerable por sí solo.

El descriptor ATSC3s destaca como el más relevante en el análisis debido a su coeficiente F elevado y su fuerte correlación con la actividad larvícida, lo que lo posiciona como una variable clave para la construcción del modelo QSAR. Esto contrasta con la limitada contribución de los descriptores termoquímicos y mecanocuánticos, cuya falta de significancia estadística sugiere que la actividad larvícida está más influenciada por características moleculares específicas que por propiedades generales derivadas de cálculos teóricos. En este contexto, la variable ATSC3s, que es un descriptor 2D-Autocorrelación está basado en la centralización Broto-Moreau (retraso 3) relacionada con la ponderación del primer estado electrotopológico atómico (E-state) (Xia et al., 2019). En el Anexo 16 se muestran los cálculos de los descriptores moleculares.

Los parámetros estadísticos del modelo QSAR para los flavonoides, calculados con un intervalo de confianza del 95% (\pm) para el descriptor molecular y el intercepto (Tabla 9), muestran una ecuación con un coeficiente de determinación (R^2) de 0,9976, lo que indica que el modelo explica el 99,76% de la variabilidad de los datos. Este modelo presenta una distribución F_{cal} de 1670 para 3 grados de libertad (GL) y un error estándar residual (EER) de 0,03369. Además, la significancia estadística del modelo con $p = 3,2242 \times 10^{-5}$ confirma que la confiabilidad del modelo (Leach & Gillet, 2007; Triola, 2004).

Tabla 9. Medidas estadísticas del modelo QSAR de flavonoides para actividad larvícida.

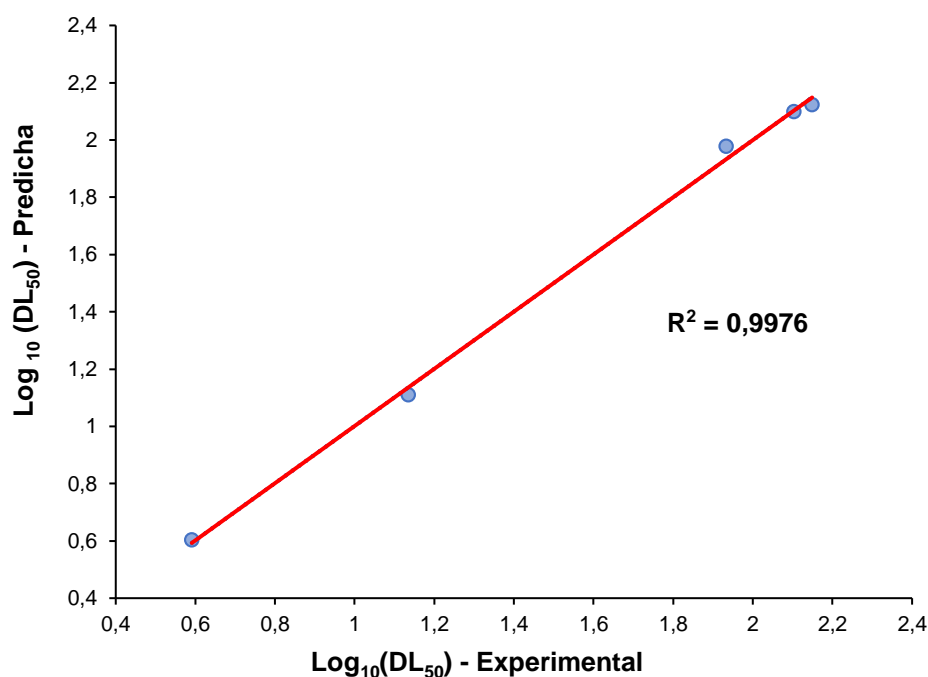
Parámetros estadísticos		
F = 1670	R = 0,9982	$R^2 = 0,9976$
GL = 3	EER = 0,03369	$p = 3,2242 \times 10^{-5}$
Ecuación matemática de predicción de la actividad		
$\text{Log}_{10}(\text{CL}_{50}) = 1,150702(\pm 0,117) - 0,041060(\pm 7,200 \times 10^{-3})$ [ATSC3s]		

El análisis de los intervalos de confianza respalda la precisión de los coeficientes estimados. Para el intercepto, el valor obtenido es $1,150702 \pm 0,117$; lo que indica, con un 95% de confianza, que el valor del intercepto se encuentra dentro de este rango, evidenciando la fiabilidad y precisión de la estimación. Así mismo, el coeficiente asociado al descriptor molecular ATSC3s es $-0,041060 \pm 7,200 \times 10^{-3}$ reflejando un bajo grado de incertidumbre en la estimación del efecto de este descriptor sobre la actividad larvícida. Estos intervalos estrechos, junto con el hecho de que el $F_{cal} \gg F_{tab}$ ($F_{tab} = 5,4095$ entre $GL_1 = 3$ y $GL_2 = 5$),

el elevado $R^2 > 0,90$ y el bajo EER confirman que el modelo QSAR no solo es estadísticamente significativo, sino también confiable y preciso para predecir la actividad larvica de las moléculas analizadas (Leach & Gillet, 2007; Triola, 2004).

Ahora bien, se observa que el modelo de predicción para flavonoides presentó buena correlación entre los datos experimentales y predichos de la actividad larvica (Anexo 17), demostrando la capacidad de predicción del modelo (Figura 12). Cabe destacar que la capacidad de predicción del modelo estadístico obtenido al 95% de confianza es buena porque evidencia medidas de Fisher y una significancia estadística muy altas.

Figura 12. Representación gráfica de la actividad larvica experimental frente a la predicha para el modelo QSAR de flavonoides.



En lo que sigue, se realizó la respectiva validación cruzada interna (CV-Int.) y la validación cruzada externa (CV-Ext.) de los modelos QSAR de derivados de eugenol, amidas sintéticas, terpenos naturales y flavonoides; puesto que son modelos con calidad estadística y buena asociación lineal significativa entre las variables de predicción. En definitiva, esta etapa permitió corroborar que los modelos son válidos estadísticamente.

4.4 VALIDACIÓN CRUZADA (CV) DE LOS MODELOS QSAR

La validación es la parte esencial en el desarrollo de un modelo estadísticamente robusto, válido y predictivo porque la utilidad de un modelo QSAR es, principalmente, aplicarlo a la predicción de la actividad de nuevos compuestos. De esta manera, los modelos construidos fueron validados empleando las dos formas propuestas en esta investigación, tal como se explica en los siguientes apartados (Veerasamy et al., 2011; Wold & Eriksson, 1995).

4.4.1 Validación cruzada interna (CV-Int.) de los modelos de derivados de eugenol, amidas sintéticas, terpenos naturales y flavonoides. La validación interna se realizó con la metodología 3.3. A partir del conjunto de moléculas vinculado a los modelos QSAR, se desarrolló esta validación, la cual consistió en excluir cada compuesto de los modelos y predecir su actividad larvícida. Así, para el modelo de eugenoles se tuvo un Q_{LOO}^2 de 0,9800, en segundo lugar, en el modelo de amidas sintéticas un Q_{LOO}^2 de 0,8796, en tercer lugar, para el modelo de terpenos un Q_{LOO}^2 de 0,9933. De igual forma, para el modelo de flavonoides se obtuvo un coeficiente de predicción Q_{LOO}^2 de 0,9957. Estos modelos presentaron $Q_{LOO}^2 > 0,50$, por lo tanto, son modelos que se consideran predictivos (Gramatica, 2007; Tropsha et al., 2003). Por otra parte, en las Figuras 18 a 21 se muestran las representaciones gráficas de la CV-Int. obtenidas para los modelos y en los Anexos 18 a 21 se encuentran los datos de actividad larvícida predichos en la CV-Int. de los modelos QSAR.

En la Figura 13 se observa que la dispersión de los datos de actividad larvícida para el modelo de eugenoles varía levemente; puesto que la diferencia del porcentaje de varianza es de 1,57% respecto al coeficiente de correlación R^2 de 0,9957. Desde luego, la tendencia lineal del modelo QSAR (Figura 9) se conserva. Así mismo, en la Figuras 14 y 15 se observa que la dispersión de los datos de actividad larvícida no varía, puesto que la tendencia lineal de los modelos estadísticos para terpenos y flavonoides no presenta cambios (Figura 10 y 11); sin embargo, la dispersión de los datos de actividad larvícida para el modelo de amidas sintéticas no se mantuvo constante (Figura 16), dado que la diferencia en el porcentaje de varianza fue del 5,78% con respecto al coeficiente R^2 de 0,9374 (Figura 10). Con las representaciones gráficas de la CV-Int., se demuestra la capacidad predictiva de los modelos QSAR.

Figura 13. Representación gráfica de la actividad larvícida experimental frente a la predicha para los compuestos en la CV-LOO del modelo QSAR de eugenoles.

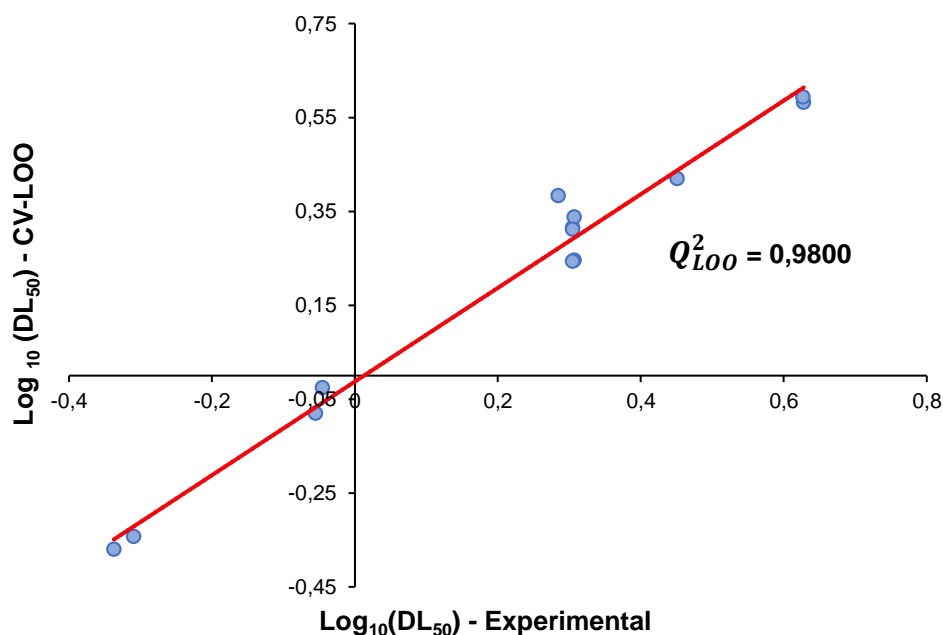


Figura 14. Representación gráfica de la actividad larvicida experimental frente a la predicha para los compuestos en la CV-LOO del modelo QSAR de terpenos.

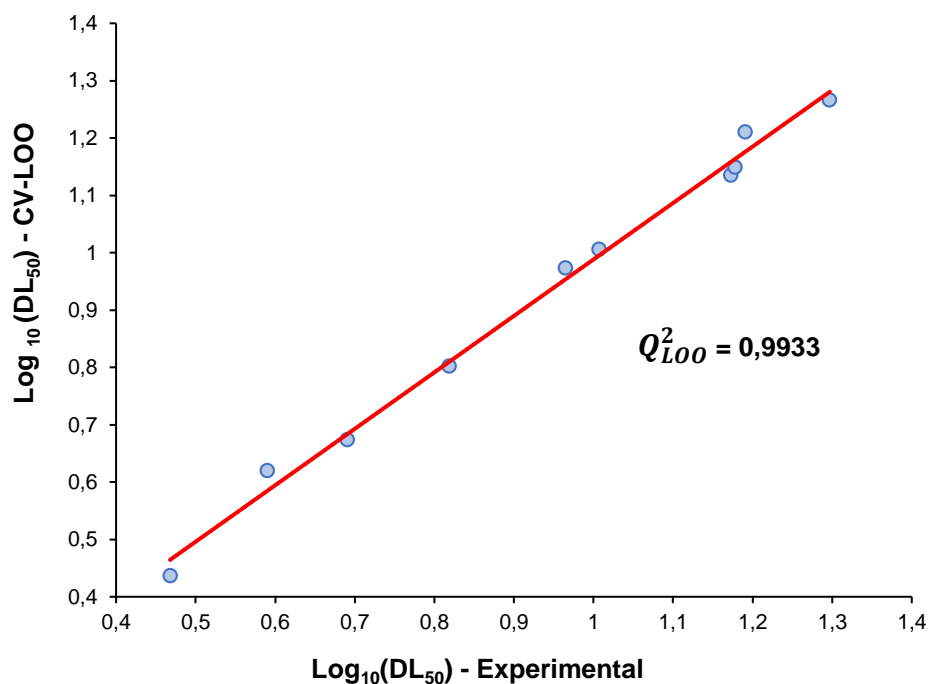


Figura 15. Representación gráfica de la actividad larvicida experimental frente a la predicha para los compuestos en la CV-LOO del modelo QSAR de flavonoides.

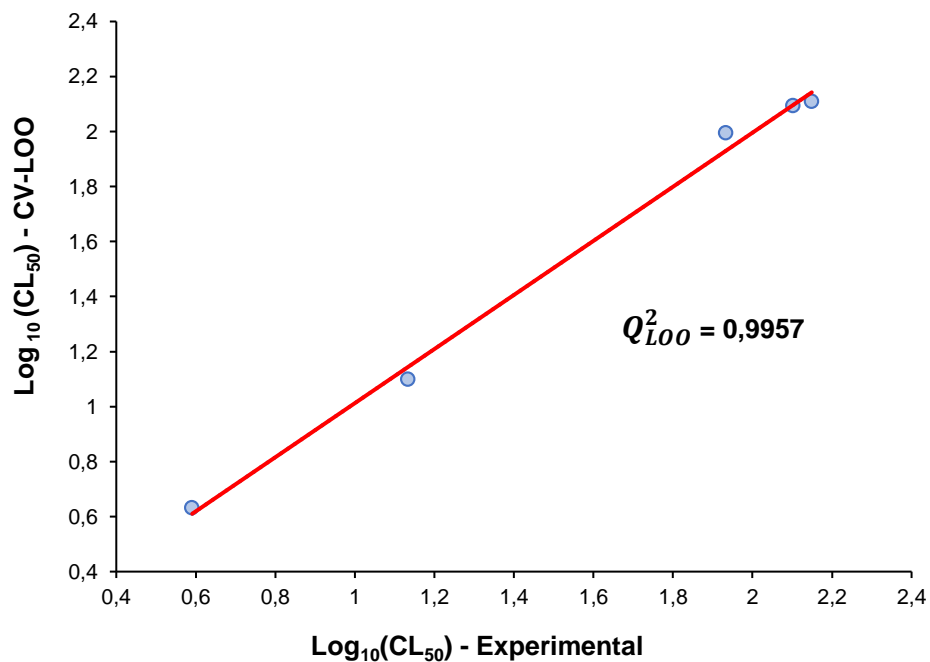
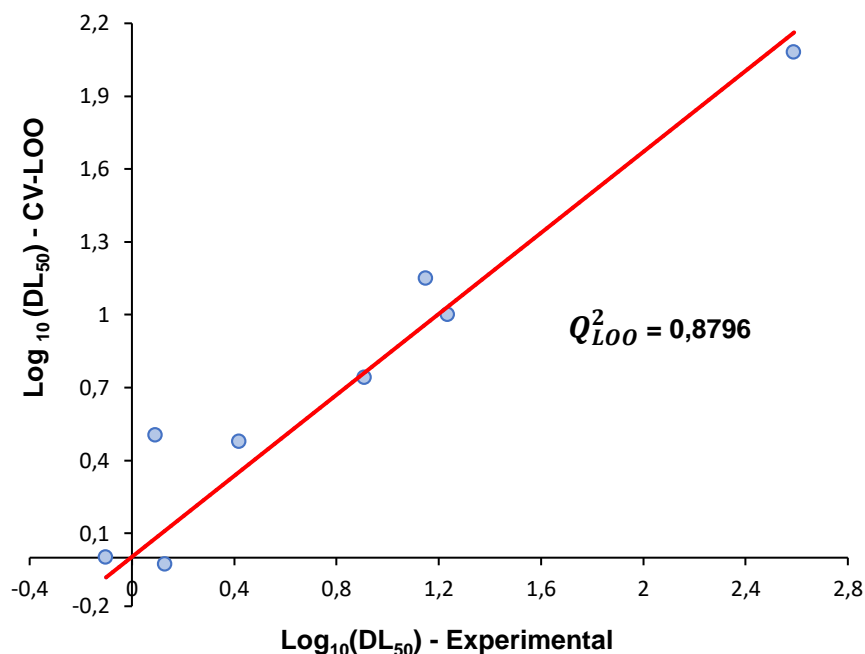


Figura 16. Representación gráfica de la actividad larvicida experimental frente a la predicha para los compuestos en la CV-LOO del modelo QSAR de amidas.



4.4.2 Validación cruzada externa (CV-Ext.) de los modelos de derivados de eugenol, amidas sintéticas, terpenos naturales y flavonoides. Los modelos QSAR de actividad larvicida de estudio fueron validados con un conjunto de prueba conformado por siete (7) compuestos, ilustrados en las Figuras 3 a 6; estos hacen parte de compuestos de origen natural (terpenos 6 y 8 - flavonoide 4), y de origen sintético (eugenoles 6 y 11 - amidas 4 y 8) reportados en la literatura y que presentan actividad larvicida contra el *Spodoptera frugiperda* (Ávila-Murillo et al., 2014; Castral et al., 2011; Estrela et al., 2003; Hernández-Carlos & Gamboa-Angulo, 2019; Vargas-Méndez et al., 2019). Los resultados de la CV-Ext. se muestran en la Tabla 10. Se debe agregar que el porcentaje de error de predicción fue utilizado como criterio para evaluar los modelos QSAR (Ecuación 9). Así pues, se consideró que los compuestos con porcentajes de error mayores al 50% son de baja predicción.

Ecuación 9. Cálculo error de predicción.

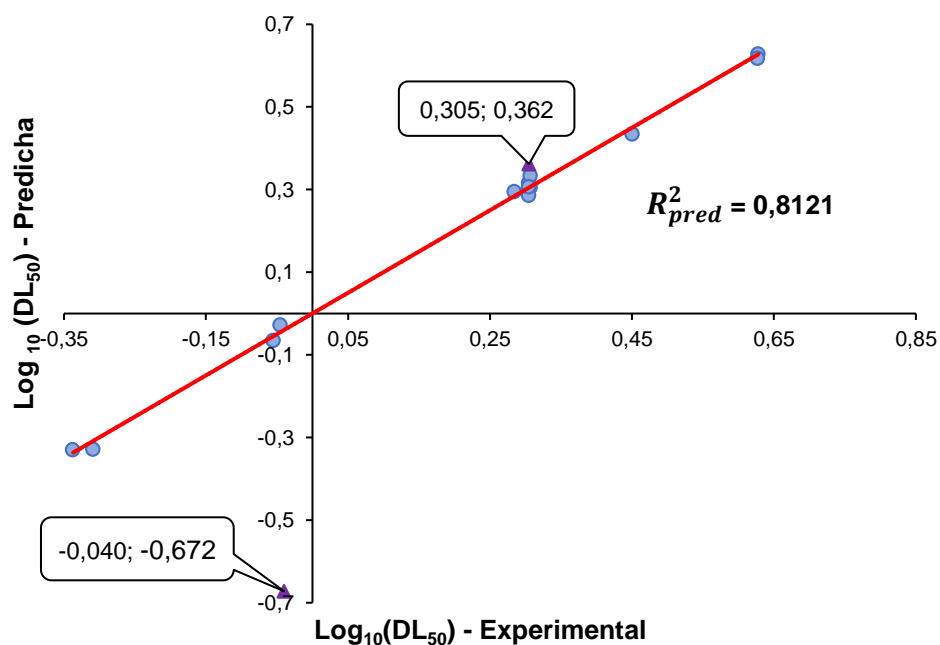
$$\%Error\ de\ predicción = \frac{|valor\ real - valor\ aproximado|}{valor\ real} * 100$$

En la Tabla 10 se presentan los resultados obtenidos de predicción de la variable DL₅₀ (mg/g de insecto) calculada para los compuestos de prueba de eugenoles. Como se observa, el eugenol 6 presentó un porcentaje de error inferior al 50% y para el eugenol 11 se obtuvo un porcentaje de error mayor al 50%; sin embargo, el eugenol 6 deja ver un carácter predictivo aceptable del modelo estadístico. En el Anexo 4 se registran los cálculos de los descriptores moleculares de los eugenoles de prueba y en el Anexo 22, sus datos de actividad larvicida predicha de forma logarítmica.

Tabla 10. Resultados de la actividad larvicida predicha para los compuestos de prueba, aplicando el modelo QSAR de eugenoles, amidas, terpenos y flavonoides.

Compuesto de prueba (Val. Ext.)	DL ₅₀ /CL ₅₀ (unidades) experimental	Log ₁₀ (DL ₅₀ /CL ₅₀) experimental	DL ₅₀ /CL ₅₀ (unidades) predicha	Log ₁₀ (DL ₅₀ /CL ₅₀) Predicha	Error de predicción (%)
Eugenol 6	2,02 mg/g de insecto	0,305	2,30 mg/g de insecto	0,362	13,86
Eugenol 11	0,91 mg/g de insecto	-0,040	0,21 mg/g de insecto	-0,672	76,92
Amida 4	0,61 µg/mg de larva	-0,215	0,785 µg/mg de larva	-0,105	28,69
Amida 8	504,07 µg/mg de larva	2,702	424,62 µg/mg de larva	2,628	15,76
Terpeno 6	4,83 ppm	0,684	6,098 ppm	0,785	26,26
Terpeno 8	12,4 ppm	1,093	18,685 ppm	1,271	50,68
Flavonoide 4	28,79 ppm	1,459	34,27 ppm	1,535	19,03

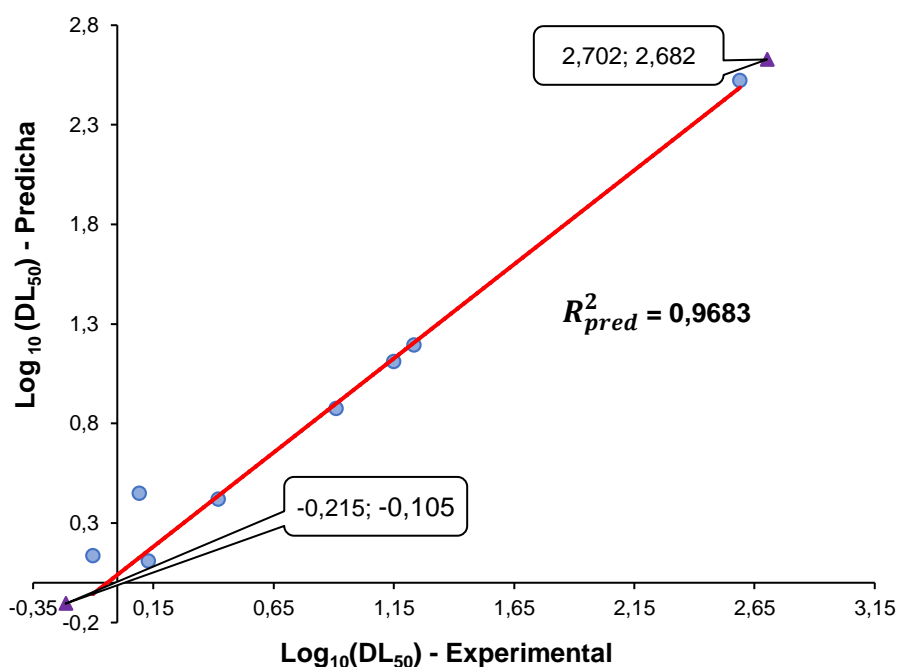
Figura 17. Representación gráfica de la actividad larvicida experimental frente a la predicha para el modelo QSAR de eugenoles. Incluye los compuestos de prueba.



La Figura 17 ilustra la correlación entre la actividad larvicida experimental y la predicha para los eugenoles, incluyendo los compuestos de prueba (eugenol 6 y 11). A pesar del resultado obtenido en el eugenol 11, el modelo QSAR presenta un buen carácter predictivo porque los valores predichos obtenidos para los compuestos de prueba se dispersan en el dominio químico del modelo, aunque no se acercan a sus valores experimentales. Con lo anterior, se confirmó la capacidad de predicción del modelo estadístico, dado que se obtuvo un coeficiente de correlación para la predicción R^2_{pred} de 0,8121.

En cuanto a las amidas, los resultados obtenidos también se presentan en la Tabla 10. Allí se obtuvieron porcentajes de error menores al 50% para las dos amidas problema. Cabe destacar que la amida 8 fue el de mejor predicción. En el Anexo 8 se muestran los cálculos de los descriptores para los compuestos y en el Anexo 22, sus datos de actividad larvicida predicha de forma logarítmica.

Figura 18. Representación gráfica de la actividad larvicida experimental frente a la predicha para el modelo QSAR de amidas. Incluye los compuestos de prueba.



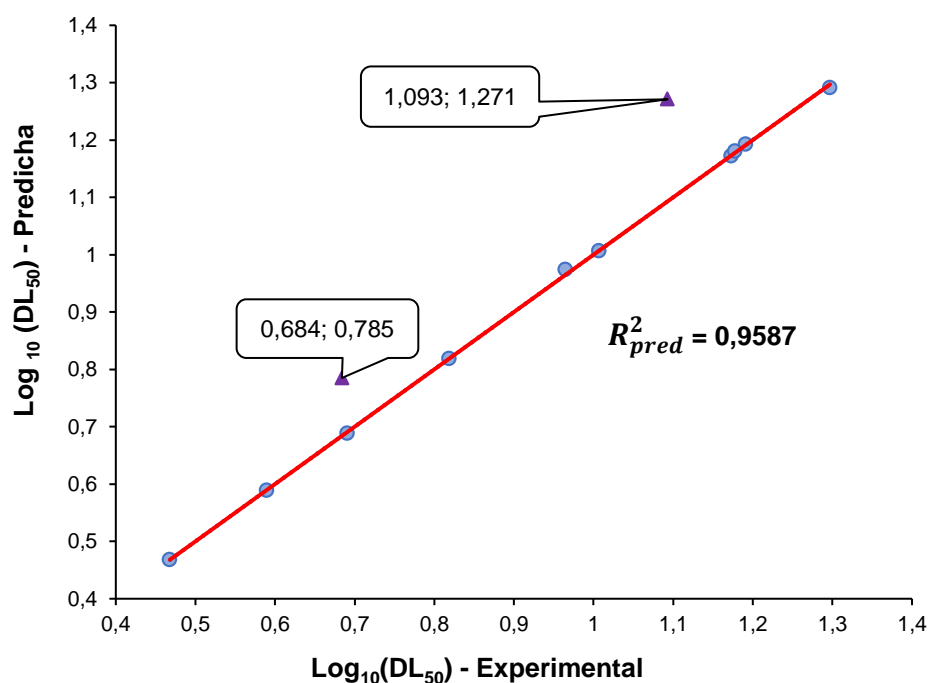
Con base en los resultados de la Tabla 10, los valores experimentales de actividad larvicida de los compuestos de prueba, presentaron una buena correlación con los valores predichos (Figura 18). Esto confirmó la capacidad de predicción del modelo QSAR de amidas; puesto que se obtuvo un R^2_{pred} de 0,9683. Como se puede observar en la Figura 18, los compuestos de prueba (amida 4 y 8) presentan predicciones destacadas de actividad larvicida y se dispersan en el dominio químico del modelo.

De forma similar, en la Tabla 10 se incluye los resultados obtenidos para los terpenos. Como se puede observar, el terpeno 8 presentó un porcentaje de error mayor al 50% y el terpeno 6 obtuvo un porcentaje inferior al 50%; desde luego, estos terpenos dejan ver un carácter

predictivo considerable del modelo QSAR. En el Anexo 12 se registran los cálculos de los descriptores moleculares de los terpenos de prueba y en el Anexo 22, sus datos de actividad larvícida predicha de forma logarítmica.

Por otro lado, en la Figura 19 se muestra la correlación entre los datos experimentales y predichos, con un coeficiente R_{pred}^2 de 0,9587, lo que confirma la capacidad predictiva del modelo QSAR. Como se puede observar en la Figura, los compuestos de prueba (terpeno 6 y 8) presentan predicciones destacadas de actividad larvícida y se dispersan en el dominio químico del modelo, aunque no se encontraran sobre la línea de tendencia.

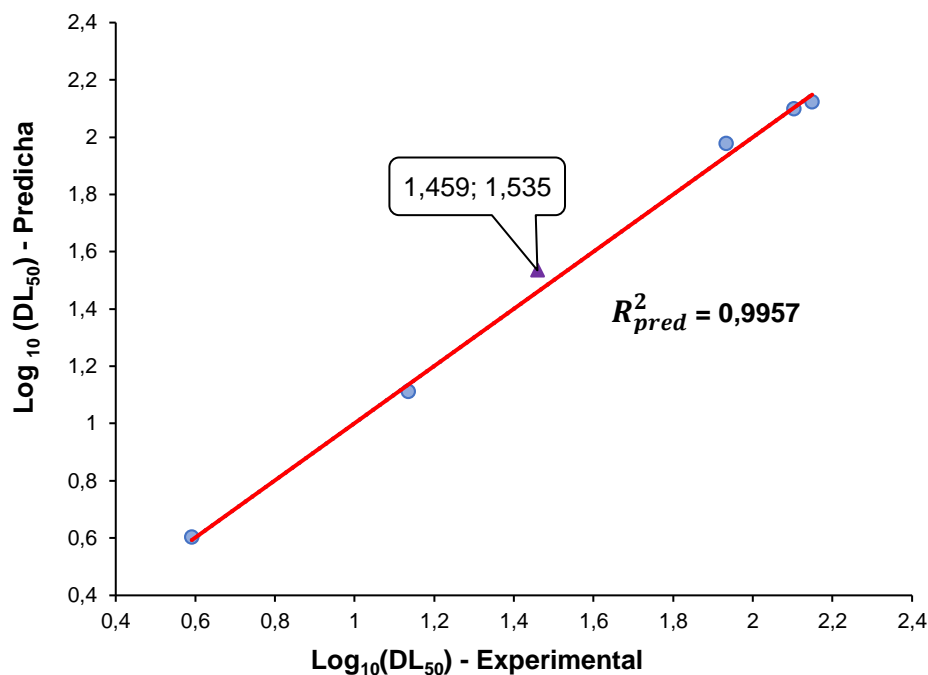
Figura 19. Representación gráfica de la actividad larvícida experimental frente a la predicha para el modelo QSAR de terpenos. Incluye los compuestos de prueba.



Para los flavonoides, la Tabla 10 muestra que el compuesto de prueba presentó un error de predicción por debajo del 50%. Esto evidencia que el modelo estadístico tiene una buena capacidad predictiva, y siendo un modelo QSAR construido con sólo 5 compuestos. En el Anexo 16 se muestran los cálculos de los descriptores para los compuestos y en el Anexo 22, sus datos de actividad larvícida predicha de forma logarítmica.

En la Figura 20, se observa que el valor experimental de actividad larvícida del compuesto de prueba, presentó una buena correlación con el valor predicho. Esto confirmó la capacidad de predicción del modelo QSAR de flavonoides; puesto que se obtuvo un R_{pred}^2 de 0,9957. Como se puede observar en la figura, el compuesto de prueba (flavonoide 4) presenta una predicción destacada de actividad larvícida y se dispersa en el dominio químico del modelo.

Figura 20. Representación gráfica de la actividad larvicida experimental frente a la predicha para el modelo QSAR de flavonoides. Incluye el compuesto de prueba.



Por último, en la discusión de los resultados se incluyó el diseño de un conjunto nuevo de moléculas con características estructurales combinadas (entre derivados de eugenol, amidas sintéticas, terpenos naturales y flavonoides) con el fin de evaluar y corroborar la capacidad predictiva de los modelos QSAR.

4.4.3 Predicción de la actividad larvicida de compuestos nuevos. Las cuatro familias químicas (eugenoles, terpenos, flavonoides y amidas) fueron empleadas para construir 12 moléculas teóricas como se ilustran en la Figura 21. A continuación se explica el diseño de estos compuestos; teniendo en cuenta las características estructurales asociadas con los mejores descriptores moleculares obtenidos de los modelos QSAR:

a) Construcción de los compuestos AM1 y AM2: A partir de la molécula del eugenol 7 (Figura 2) y de la amida 3 (Figura 4) se construyó la combinación AM1; y la combinación AM2 se diseñó con el eugenol 10 (Figura 2) y la amida 11 (Figura 4). Con estos compuestos, se propusieron sustituciones estructurales en los grupos amida (-RCONH₂) por grupos oxigenados; ya que inciden en la electronegatividad centralizada de los átomos más pesados, la cual se considera mediante los descriptores moleculares ATSC6e (electronegatividad ponderada) y AATSC8 (electronegatividad ponderada y central de Sanderson). Por otro lado, la sustitución de estas cadenas ramificadas entre los grupos funcionales, es considerada por el descriptor ALogP, el cual está relacionado con la capacidad atómica y aditiva por su constitución.

b) Construcción de los compuestos BM1 y BM2: La combinación BM1 se planteó a partir de las moléculas eugenol 14 (Figura 2) y terpeno 4 (Figura 3); en el caso de la combinación BM2 fue diseñada con base en el eugenol 13 (Figura 2) y el terpeno 1 (Figura 3). Ahora

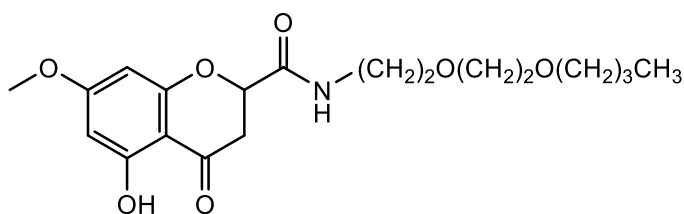
bien, se plantearon sustituciones de características aromáticas y cadenas alifáticas, las cuales influyen en la electronegatividad centralizada y ponderada (descriptores ATSC6e y AATSC8) de los átomos de carbono (C), hidrógeno (H) y oxígeno (O). Además, estas sustituciones estructurales promueven diferentes distribuciones radiales de volumen de Van Der Waals, explicada por el descriptor RDF140v y la correlación en distancia topológica entre sus átomos, sustentada por el descriptor TDB2u.

c) Construcción de los compuestos CM1 y CM2: Para la construcción de la combinación molecular CM1 se emplearon los compuestos de eugenol 2 (Figura 2) y flavonoide 1 (Figura 5) y en la combinación CM2 se utilizaron las moléculas de eugenol 12 (Figura 2) y flavonoide 3 (Figura 5). Teniendo en cuenta estas moléculas, se propusieron sustituciones estructurales en el grupo metoxi (-OCH₃) y aromático (R-C₆H₅) por cadenas cíclicas; ya que inciden en la electronegatividad centralizada y ponderada, la cual se explica por los descriptores ATS6e y AATSC8. Por otra parte, las sustituciones estructurales realizadas en cada molécula generan que el estado electrotopológico de sus átomos aumente por la deslocalización electrónica; esta característica se explica por el descriptor ATS3s (autocorrelación centralizada por el primer estado electrotopológico).

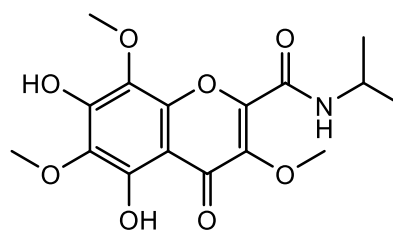
d) Construcción de los compuestos DM1 y DM2: A partir del compuesto amida 7 (Figura 4) y el terpeno 5 (Figura 3) se construyó la combinación DM1; y la combinación DM2 se diseñó con la amida 10 (Figura 4) y el terpeno 7 (Figura 3). Con estos compuestos, se propusieron sustituciones en los grupos amida (-RCONH₂) y grupo aromático (R-C₆H₅) por cadenas alifáticas con presencia de funciones oxigenadas (función carboxilo, -RCOOH y función éster, -RCOOR-), ya que inciden en la constitución aditiva atómica, explicada por el descriptor ALogP. Adicionalmente, la sustitución de las cadenas alifáticas a las estructuras cíclicas favorece la distribución radial y el volumen de Van Der Waals, explicado por el descriptor RDF140v; y por ende la autocorrelación de la distancia topológica de los átomos, sustentado por el descriptor molecular TDB2u.

e) Construcción de los compuestos EM1 y EM2: La combinación molecular EM1 se planteó a partir del compuesto amida 6 (Figura 4) y flavonoide 5 (Figura 5); en el caso de la combinación EM2 fue diseñada con base en la amida 4 (Figura 4) y el flavonoide 1 (Figura 5). Ahora bien, se plantearon sustituciones de características aromáticas (R-C₆H₅) por grupos amida (-RCONH₂), las cuales influyen en la correlación del estado electrotopológico centralizado por parte de los átomos de oxígeno (O) y nitrógeno (N), explicado por el descriptor ATS3s. Por su parte, el descriptor ALogP explica la constitución del volumen aditivo de las cadenas ramificadas de los grupos amida unidas a la base estructural de los flavonoides.

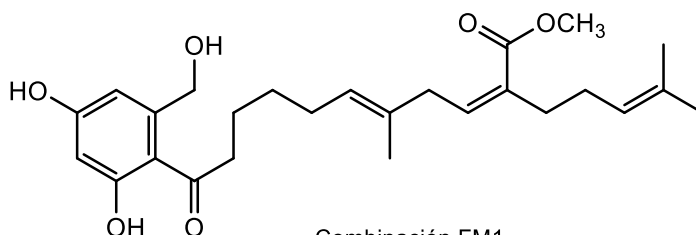
f) Construcción de los compuestos FM1 y FM2: Para el diseño de la combinación FM1 se emplearon las moléculas de terpeno 2 (Figura 3) y flavonoide 2 (Figura 5) y en la combinación FM2 se utilizaron los compuestos de terpeno 3 (Figura 3) y flavonoide 6 (Figura 5). Teniendo en cuenta estas moléculas, se propusieron sustituciones estructurales en el grupo aromático (R-C₆H₅) por cadenas alifáticas oxigenadas; ya que inciden en la distribución radial y el volumen de Van Der Waals, el cual es explicado por el descriptor RDF140v; así mismo por la correlación en la distancia topológica que existe entre los átomos de carbono (C) y oxígeno (O); esta característica se explica por el descriptor TDB2u. Adicionalmente, la presencia de cadenas alifáticas y cíclicas con la presencia de átomos de oxígeno (O) promueve la centralización y la deslocalización de los electrones en el primer



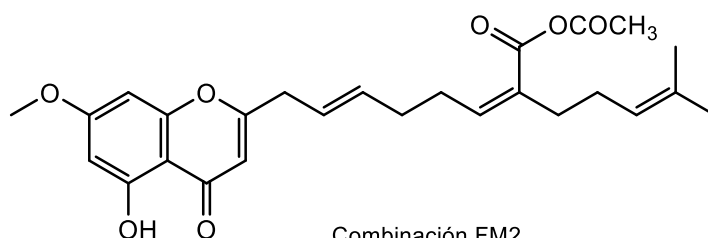
Combinación EM1



Combinación EM2



Combinación FM1



Combinación FM2

En cuanto a las predicciones de estos compuestos nuevos (Tabla 11), se puede analizar que las combinaciones moleculares DM1 y DM2 presentan similitud estructural y sus actividades larvicidas son próximas entre sí. En el caso de las moléculas de las combinaciones AM1 a CM2, se indaga que el comportamiento de la actividad varía de acuerdo con la presencia de cadenas cíclicas y grupos oxigenados; en cambio para las combinaciones EM1 a FM2, la actividad tiende a disminuir por la presencia de grupos oxigenados cercanos a cadenas alicíclicas. En los Anexos 23 a 26 se muestran los cálculos de los descriptores para las moléculas propuestas.

Uno de los compuestos más activos fue la combinación CM2 puesto que se requiere menor DL_{50} para matar las larvas de *Spodoptera frugiperda* y el otro compuesto fue la combinación EM2 con un valor de $\text{Log}_{10}CL_{50}$ de 0,281. Cabe destacar que el desarrollo de nuevos compuestos larvicidas constituye un desafío; por consiguiente, la información interpretada del análisis ANOVA (Anexo 27) de los mejores descriptores seleccionados (ATSC6e, AATSC8, ALogP, RDF140v, TDB2u y ATSC3s), presentaron valores de Fisher-Snedecor (F) y significancia estadística (p) altas en comparación con los demás descriptores moleculares de los modelos, independientemente si los valores calculados y la probabilidad del Fisher individual (*F values* y Pr (>F)) fueran altos o bajos. Así mismo, estos descriptores moleculares pueden ser empleados para estudiar el mecanismo de acción de los compuestos.

Tabla 11. Resultados consolidados de la actividad larvicida predicha para las distintas combinaciones moleculares basadas en los modelos QSAR.

Combinación molecular	Modelo estadístico	Log ₁₀ (DL ₅₀) Predicha	Combinación molecular	Modelo estadístico	Log ₁₀ (CL ₅₀) Predicha
CM2	Eugenoles	-8,768	EM2	Amidas	0,281
EM2	Amidas	-2,099	AM2	Eugenoles	0,722
EM1	Amidas	-1,502	BM2	Eugenoles	1,064
FM2	Terpenos	0,686	DM1	Amidas	1,117
FM1	Terpenos	0,929	CM1	Eugenoles	1,129
CM1	Eugenoles	2,501	BM1	Eugenoles	1,267
DM2	Amidas	4,075	DM2	Amidas	1,423
DM1	Amidas	4,665	FM1	Terpenos	1,476
BM2	Eugenoles	5,039	EM1	Amidas	1,677
AM1	Eugenoles	13,852	CM2	Eugenoles	1,739
BM1	Eugenoles	16,227	AM1	Eugenoles	2,299
AM2	Eugenoles	19,168	FM2	Terpenos	2,357

Con base en estas características estructurales, se diseñaron 12 moléculas consideradas prometedoras para la actividad larvicida, como se muestra en la Figura 21. Aunque estos diseños son teóricos, se propone su evaluación experimental en futuros estudios para confirmar las predicciones realizadas y ampliar su aplicabilidad. En este marco, el análisis de la actividad larvicida predicha subraya que la combinación CM2 destaca por su notable eficacia, al presentar un valor de Log₁₀(DL₅₀) sumamente negativo (-8,768); lo que indica que se requiere una cantidad extremadamente baja del compuesto para producir el efecto letal. Por su parte, EM2 presentó un perfil igualmente relevante, con un Log₁₀(DL₅₀) de -2,099 y un CL₅₀ sobresaliente de 0,281. Esto sugiere no solo una alta efectividad en dosis pequeñas, consolidando su posición como compuesto efectivo y consistente.

Al analizar las demás combinaciones incluidas en la tabla, sus valores de Log₁₀(DL₅₀) y Log₁₀(CL₅₀) se ubican en rangos moderados. Aunque no alcanzan la eficacia de CM2 y EM2, varias combinaciones muestran resultados significativos, lo que refleja una tendencia positiva en la actividad larvicida de las amidas y eugenoles. Estas variaciones parecen depender de la naturaleza de los grupos funcionales y la posición de los sustituyentes en las moléculas, lo que refuerza la importancia de optimizar el diseño estructural. Estos resultados consolidan a las combinaciones moleculares CM2 y EM2 como las alternativas más prometedoras dentro del análisis QSAR y justifican la propuesta de explorar experimentalmente las moléculas diseñadas.

5. CONCLUSIONES

-La investigación demostró que la metodología QSAR es una herramienta eficaz para predecir la actividad larvívica de compuestos naturales y sintéticos frente al *Spodoptera frugiperda*. Se construyeron cuatro modelos estadísticos con derivados de eugenol, amidas sintéticas, terpenos y flavonoides, logrando correlaciones significativas entre las propiedades estructurales y la actividad biológica. La validación interna y externa de los modelos confirmó su solidez estadística y su capacidad predictiva, con coeficientes de determinación y de predicción superiores a 0,70; lo que respalda su aplicabilidad en la selección de moléculas con potencial insecticida.

-De manera particular, los modelos basados en amidas sintéticas y flavonoides mostraron predicciones más confiables, con errores menores al 30%, mientras que los derivados de eugenol evidenciaron limitaciones en ciertos casos. Estos hallazgos subrayan la importancia de los descriptores moleculares seleccionados y la influencia de los grupos funcionales en la actividad larvívica. Además, el diseño teórico de nuevas moléculas permitió identificar combinaciones como CM2 y EM2, que se destacan por su alta eficacia, lo que justifica su evaluación experimental futura.

-El uso de cálculos mecanocuánticos y descriptores moleculares demostró ser fundamental para optimizar la calidad estadística de los modelos, ya que permitió seleccionar parámetros con una interpretación química coherente y vinculada al mecanismo de acción de los compuestos. Este enfoque evidencia cómo la integración de métodos computacionales puede reducir la necesidad de ensayos experimentales extensos, agilizando el diseño racional de insecticidas alternativos.

-El estudio resalta la relevancia de continuar explorando metabolitos secundarios de origen vegetal, especialmente terpenos y flavonoides, por su bajo impacto en la salud humana y el ambiente. En este sentido, la aplicación de modelos QSAR no solo abre la puerta al diseño de compuestos más efectivos contra *Spodoptera frugiperda*, sino que también contribuye a la construcción de estrategias agrícolas más sostenibles y alineadas con las demandas de seguridad alimentaria global.

-En conjunto, este trabajo aporta una base metodológica coherente para el desarrollo de alternativas más seguras y sostenibles en el control de plagas agrícolas, reduciendo la dependencia de insecticidas sintéticos altamente tóxicos. Los resultados obtenidos no solo fortalecen la comprensión de la relación entre estructura y actividad, sino que también ofrecen una herramienta predictiva que puede acelerar la búsqueda de compuestos con alto potencial larvívica y menor impacto ambiental.

6. RECOMENDACIONES

De este trabajo de investigación, se presentan las siguientes recomendaciones:

-Es pertinente ampliar la base de datos de moléculas con actividad larvícida y que tengan estructuras químicas similares entre sí. Éstas podrían ser incluidas en los modelos QSAR de actividad larvícida propuestos, y así sería posible observar el comportamiento de cada uno e incrementar sus capacidades predictivas.

-Para comprobar la validez estadística de los modelos matemáticos, se sugiere emplear compuestos externos con actividad larvícida y que sean de las mismas familias del conjunto de datos empleados en este estudio.

-Con el fin de garantizar la fiabilidad y robustez de los modelos QSAR, se recomienda sintetizar los nuevos compuestos hipotéticos, evaluar su actividad larvícida y comparar los resultados con las predicciones de los modelos estadísticos.

7. DIVULGACIÓN DE LOS RESULTADOS

Los resultados del trabajo de grado se presentarán para el año 2026 a manera de artículo científico y con miras a publicarse en la revista colombiana de Química.

8. BIBLIOGRAFÍA

- Abdulfatai, U., Uzairu, A., Shallangwa, G. A., & Uba, S. (2020). Computational modelling of some anti-friction lubricant additives. *South African Journal of Chemical Engineering*, 32(December 2019), 27–31. <https://doi.org/10.1016/j.sajce.2020.01.003>
- Ahissou, B. R., Sawadogo, W. M., Bokonon-Ganta, A. H., Somda, I., Kestemont, M. P., & Verheggen, F. J. (2021). Baseline Toxicity Data of Different Insecticides against the Fall Armyworm *Spodoptera frugiperda* (J.E. Smith) (Lepidoptera: Noctuidae) and Control Failure Likelihood Estimation in Burkina Faso. *African Entomology*, 29(2), 435–444. <https://doi.org/10.4001/003.029.0435>
- Altürk, S., Avcı, D., Tamer, Ö., & Atalay, Y. (2017). Comparison of different hybrid DFT methods on structural, spectroscopic, electronic and NLO parameters for a potential NLO material. *Computational and Theoretical Chemistry*, 1100, 34–45. <https://doi.org/10.1016/j.comptc.2016.12.007>
- Ambure, P., Halder, A. K., González Díaz, H., & Cordeiro, M. N. D. S. (2019). QSAR-Co: An Open Source Software for Developing Robust Multitasking or Multitarget Classification-Based QSAR Models. *Journal of Chemical Information and Modeling*, 59(6), 2538–2544. <https://doi.org/10.1021/acs.jcim.9b00295>
- Andrade, C. (2019). The P Value and Statistical Significance: Misunderstandings, Explanations, Challenges, and Alternatives. *Indian Journal of Psychological Medicine*, 41(3), 210–215. <https://doi.org/10.4103/IJPSYM.IJPSYM>
- Atkins, P., & Friedman, R. (2011). *Molecular Quantum Mechanics*. [http://www.nanoer.net/d/img/Molecular Quantum Mechanics, 5th Edition.pdf](http://www.nanoer.net/d/img/Molecular%20Quantum%20Mechanics,%205th%20Edition.pdf)
- Arthur, D. E., Soliman, M. E. S., Adeniji, S. E., Adedirin, O., & Peter, F. (2022). Qsar and Molecular Docking Study of Gonadotropin-Releasing Hormone Receptor Inhibitors. *Scientific African*, 17, e01291. <https://doi.org/10.1016/j.sciaf.2022.e01291>
- Arthur, D. E., Uzairu, A., Mamza, P., & Abechi, S. (2016). Quantitative structure–activity relationship study on potent anticancer compounds against MOLT-4 and P388 leukemia cell lines. *Journal of Advanced Research*, 7(5), 823–837. <https://doi.org/10.1016/j.jare.2016.03.010>
- Ávila-Martínez, D., Cervantes-Ortiz, F., Rodríguez-Pérez, G., Gámez-Vázquez, A. J., García-Rodríguez, J. G., & Mendoza-Elos, M. (2023). Damage and population dynamics of the fall armyworm (*Spodoptera frugiperda*) in elite lines of corn. *Agronomía Mesoamericana*, 34(3). <https://doi.org/10.15517/am.2023.53809>
- Ávila-Murillo, M. C., Cuca-Suárez, L. E., & Salamanca-Cerón, J. A. (2014). Actividad insecticida sobre *Spodoptera frugiperda* (Lepidoptera: Noctuidae) de los compuestos aislados de la parte aérea de *Piper septuplinervium* (Miq.) C. DC. y las inflorescencias de *Piper subtomentosum* Trel. & Yunck. (Piperaceae). *Química Nova*, 37(3), 442–446. <https://doi.org/10.5935/0100-4042.20140067>
- Avogadro. (2018). *Avogadro Chemistry*. <https://avogadro.cc/>
- Ayil-Gutiérrez, B. A., Sánchez-Teyer, L. F., Vázquez-Flota, F., Monforte-González, M., Tamayo-Ordóñez, Y., Tamayo-Ordóñez, M. C., & Rivera, G. (2018). Biological effects of natural products against *Spodoptera* spp. *Crop Protection*, 114, 195–207. <https://doi.org/10.1016/j.cropro.2018.08.032>
- Baloch, M. N., Fan, J., Haseeb, M., & Zhang, R. (2020). Mapping potential distribution of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) in Central Asia. *Insects*, 11(3), 1–10. <https://doi.org/10.3390/insects11030172>
- Barrera, G., Simón, O., Villamizar, L., Williams, T., & Caballero, P. (2011). *Spodoptera*

- frugiperda multiple nucleopolyhedrovirus as a potential biological insecticide: Genetic and phenotypic comparison of field isolates from Colombia. *Biological Control*, 58(2), 113–120. <https://doi.org/10.1016/j.biocontrol.2011.04.009>
- Bateman, M. L., Day, R. K., Luke, B., Edgington, S., Kuhlmann, U., & Cock, M. J. W. (2018). Assessment of potential biopesticide options for managing fall armyworm (*Spodoptera frugiperda*) in Africa. *Journal of Applied Entomology*, 142(9), 805–819. <https://doi.org/10.1111/jen.12565>
- Batista-Pereira, L. G., Castral, T. C., Da Silva, M. T. M., Amaral, B. R., Fernandes, J. B., Vieira, P. C., Da Silva, M. F. G. F., & Corrêa, A. G. (2006). Insecticidal activity of synthetic amides on *Spodoptera frugiperda*. *Zeitschrift Fur Naturforschung - Section C Journal of Biosciences*, 61(3–4), 196–202. <https://doi.org/10.1515/znc-2006-3-408>
- Becke, A. D. (1993). A new mixing of Hartree–Fock and local density-functional theories. *The Journal of Chemical Physics*, 98, 1372. <https://doi.org/10.1063/1.464304>
- Bijlsma, L., Bade, R., Celma, A., Mullin, L., Cleland, G., Stead, S., Hernandez, F., & Sancho, J. V. (2017). Prediction of Collision Cross-Section Values for Small Molecules: Application to Pesticide Residue Analysis. *Analytical Chemistry*, 89(12), 6583–6589. <https://doi.org/10.1021/acs.analchem.7b00741>
- Bochevarov, A. D., Harder, E., Hughes, T. F., Greenwood, J. R., Braden, D. a., Philipp, D. M., Rinaldo, D., Halls, M. D., Zhang, J., & Friesner, R. A. (2013). Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences. *International Journal of Quantum Chemistry*, 113, 2110–2142. <https://doi.org/10.1002/qua.24481>
- Bugeac, C. A., Ancuceanu, R., & Dinu, M. (2021). QSAR Models for Active Substances against *Pseudomonas aeruginosa* Using Disk-Diffusion Test Data. *Molecules*, 26(6), 1–18. <https://doi.org/10.3390/molecules26061734>
- Caldeweyher, E., & Brandenburg, J. G. (2018). Simplified DFT methods for consistent structures and energies of large systems. *Journal of Physics Condensed Matter*, 30(21). <https://doi.org/10.1088/1361-648X/aabcfb>
- Carvalho, R. A., Omoto, C., Field, L. M., Williamson, M. S., & Bass, C. (2013). Investigating the Molecular Mechanisms of Organophosphate and Pyrethroid Resistance in the Fall Armyworm *Spodoptera frugiperda*. *PLoS ONE*, 8(4). <https://doi.org/10.1371/journal.pone.0062268>
- Castral, T. C., Matos, A. P., Monteiro, J. L., Araujo, F. M., Bondancia, T. M., Batista-Pereira, L. G., Fernandes, J. B., Vieira, P. C., Da Silva, M. F. G. F., & Corrêa, A. G. (2011). Synthesis of a combinatorial library of amides and its evaluation against the fall armyworm, *Spodoptera frugiperda*. *Journal of Agricultural and Food Chemistry*, 59(9), 4822–4827. <https://doi.org/10.1021/jf104903t>
- Chaudhry, Q., Chrétien, J., Craciun, M., Guo, G., Lemke, F., Müller, J.-A., Neagu, D., Piclin, N., Pintore, M., & Trundle, P. (2007). Algorithms for (Q)SAR model building. In E. Benfenati (Ed.), *Quantitative Structure-Activity Relationships (QSAR) for Pesticide Regulatory Purposes* (pp. 111–144). <https://doi.org/10.1016/B978-044452710-3/50006-9>
- Cheng, C. L., Shalabh, & Garg, G. (2014). Coefficient of determination for multiple measurement error models. *Journal of Multivariate Analysis*, 126(1), 137–152. <https://doi.org/10.1016/j.jmva.2014.01.006>
- Cherkasov, A., Muratov, E. N., Fourches, D., Varnek, A., Baskin, I. I., Cronin, M., Dearden, J., Gramatica, P., Martin, Y. C., Todeschini, R., Consonni, V., Kuz'Min, V. E., Cramer, R., Benigni, R., Yang, C., Rathman, J., Terfloth, L., Gasteiger, J., Richard, A., & Tropsha, A. (2014). QSAR modeling: Where have you been? Where are you going to?

- Journal of Medicinal Chemistry*, 57(12), 4977–5010.
<https://doi.org/10.1021/jm4004285>
- Chirico, N., & Gramatica, P. (2011). Real external predictivity of QSAR models: How to evaluate It? Comparison of different validation criteria and proposal of using the concordance correlation coefficient. *Journal of Chemical Information and Modeling*, 51(9), 2320–2335. <https://doi.org/10.1021/ci200211n>
- Cramer, R. D., Bunce, J. D., Patterson, D. E., & Frank, I. E. (1988). Crossvalidation, Bootstrapping, and Partial Least Squares Compared with Multiple Regression in Conventional QSAR Studies. *Quant. Struct.-Act. Relat.*, 7, 18–25. <https://doi.org/10.1002/qsar.19880070105>
- Cronin, M. T. D. (2010). Quantitative Structure–Activity Relationships (QSARs) – Applications and Methodology. In T. Puzyn, J. Leszczynski, & M. Cronin (Eds.), *Recent Advances in QSAR Studies* (pp. 3–11). https://doi.org/10.1007/978-1-4020-9783-6_1
- Cruz, G. S., Wanderley-Teixeira, V., Oliveira, J. V., Correia, A. A., Breda, M. O., Alves, T. J. S., Cunha, F. M., Teixeira, A. A. C., Dutra, K. A., & Navarro, D. M. A. F. (2014). Bioactivity of Piper hispidinervum (Piperales: Piperaceae) and Syzygium aromaticum (Myrtales: Myrtaceae) oils, with or without formulated bta on the biology and immunology of Spodoptera frugiperda (Lepidoptera: Noctuidae). *Journal of Economic Entomology*, 107(1), 144–153. <https://doi.org/10.1603/EC13351>
- Da Silva, J. B. P., Navarro, D. M. D. A. F., Da Silva, A. G., Santos, G. K. N., Dutra, K. A., Moreira, D. R., Ramos, M. N., Espíndola, J. W. P., De Oliveira, A. D. T., Brondani, D. J., Leite, A. C. L., Hernandez, M. Z., Pereira, V. R. A., Da Rocha, L. F., De Castro, M. C. a B., De Oliveira, B. C., Lan, Q., & Merz, K. M. (2015). Thiosemicarbazones as Aedes aegypti larvicidal. *European Journal of Medicinal Chemistry*, 100, 162–175. <https://doi.org/10.1016/j.ejmech.2015.04.061>
- Dambolena, J. S., Zunino, M. P., Herrera, J. M., Pizzolitto, R. P., Areco, V. A., & Zygadlo, J. A. (2016). Terpenes : Natural Products for Controlling Insects of Importance to Human Health - A Structure-Activity Relationship Study. *Hindawi Publishing Corporation Psyche*, 2016, 1–17. <https://doi.org/http://dx.doi.org/10.1155/2016/4595823>
- Darlington, R. B., & Hayes, A. F. (2016). The Simple Regression Model. In T. D. Little (Ed.), *Regression Analysis and Linear Models: Concepts, Applications, and Implementation* (pp. 17–42). <https://www.afhayes.com/index.html>
- De Lile, J. R., Kang, S. G., Son, Y. A., & Lee, S. G. (2020). Do HOMO-LUMO energy levels and band gaps provide sufficient understanding of Dye-sensitizer activity trends for water purification? *ACS Omega*, 5(25), 15052–15062. <https://doi.org/10.1021/acsomega.0c00870>
- Dearden, J. C. (2016). The History and Development of Quantitative Structure-Activity Relationships (QSARs). *International Journal of Quantitative Structure-Property Relationships*, 1(1), 1–44. <https://doi.org/10.4018/IJQSPR.2016010101>
- Dehmer, M., Varmuza, K., Borgert, S., & Emmert-Streib, F. (2009). On entropy-based molecular descriptors: Statistical analysis of real and synthetic chemical structures. *Journal of Chemical Information and Modeling*, 49(7), 1655–1663. <https://doi.org/10.1021/ci900060x>
- Delong, W., Yongling, W., Lanying, W., Juntao, F., & Xing, Z. (2017). Design, synthesis and evaluation of 3-arylidene azetidin-2-ones as potential antifungal agents against Alternaria solani Sorauer. *Bioorganic and Medicinal Chemistry*, 25(24), 6661–6673. <https://doi.org/10.1016/j.bmc.2017.11.003>
- Devore, J. L. (2007). *Probabilidad y Estadística para Ingeniería y Ciencias* (CENGAGE Learning (ed.); 7th ed.).

- Diaz Napal, G. N., & Palacios, S. M. (2015). Bioinsecticidal effect of the flavonoids pinocembrin and quercetin against *Spodoptera frugiperda*. *Journal of Pest Science*, *88*(3), 629–635. <https://doi.org/10.1007/s10340-014-0641-z>
- Djouaka, R., Soglo, M. F., Kusimo, M. O., Adéoti, R., Talom, A., Zeukeng, F., Paraiso, A., Afari-Sefa, V., Saethre, M. G., Manyong, V., Tamò, M., Waage, J., Lines, J., & Mahuku, G. (2018). The rapid degradation of lambda-cyhalothrin makes treated vegetables relatively safe for consumption. *International Journal of Environmental Research and Public Health*, *15*(7). <https://doi.org/10.3390/ijerph15071536>
- dos Santos, I. M., Agra, J. P. G., de Carvalho, T. G. C., de Azevedo Maia, G. L., & de Alencar Filho, E. B. (2018). Classical and 3D QSAR studies of larvicidal monoterpenes against *Aedes aegypti*: new molecular insights for the rational design of more active compounds. *Structural Chemistry*, *29*(5), 1287–1297. <https://doi.org/10.1007/s11224-018-1110-8>
- Draper, N. R., & Smith, H. (1998). *Applied Regression Analysis* (John Wiley & Sons (ed.); 3rd ed.).
- Duchowicz, P. R., Goodarzi, M., Ocsachoque, M. A., Romanelli, G. P., Ortiz, E. del V., Autino, J. C., Bennardi, D. O., Ruiz, D. M., & Castro, E. A. (2009). QSAR analysis on *Spodoptera litura* antifeedant activities for flavone derivatives. *Science of the Total Environment*, *408*(2), 277–285. <https://doi.org/10.1016/j.scitotenv.2009.09.041>
- Early, R., González-Moreno, P., Murphy, S. T., & Day, R. (2018). Forecasting the global extent of invasion of the cereal pest *Spodoptera frugiperda*, the fall armyworm. *NeoBiota*, *40*, 25–50. <https://doi.org/10.3897/neobiota.40.28165>
- Eriksson, L., Jaworska, J., Worth, A. P., Cronin, M. T. D., McDowell, R. M., & Gramatica, P. (2003). Methods for Reliability and Uncertainty Assessment and for Applicability Evaluations of Classification- and Regression-Based QSARs. *Environmental Health Perspectives*, *111*(10), 1361–1375. <https://doi.org/10.1289/ehp.5758>
- Essin, A. M., & Griffiths, D. J. (2006). Quantum mechanics of the $1/x^2$ potential. *American Journal of Physics*, *74*(2), 109–117. <https://doi.org/10.1119/1.2165248>
- Estrela, J. L. V., Guedes, R. N. C., Maltha, C. R. A., & Fazolin, M. (2003). Toxicidade de Amidas análogas à piperina a larvas de *Ascia monuste orseis* Godart (Lepidoptera: Pieridae) e *Spodoptera frugiperda* (J.E. Smith) (Lepidoptera: Noctuidae). *Neotropical Entomology*, *32*(2), 343–346. <https://doi.org/10.1590/S1519-566X2003000200022>
- FAO, & WHO. (2016). *International Code of Conduct on Pesticide Management: Guidelines on highly hazardous pesticides*. <http://www.fao.org/publications/card/en/c/a5347a39-c961-41bf-86a4-975cdf2fd063/>
- Figueroa-Brito, R., Huerta-De La Peña, A., Moreno, I. P., Mancebón, V. S. M., & López-Olguín, J. F. (2011). Insecticidal activity of seed extracts of *Carica papaya* (L.) against the fall armyworm *Spodoptera frugiperda* (J.E. Smith) (Lepidoptera: Noctuidae). *Interciencia*, *36*(10), 752–756. <https://www.redalyc.org/articulo.oa?id=33921449008>
- Friesner, R. A. (2005). Ab initio quantum chemistry: Methodology and applications. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(19), 6648–6653. <https://doi.org/10.1073/pnas.0408036102>
- Fujita, T. (1997). Recent success stories leading to commercializable bioactive compounds with the aid of traditional QSAR procedures. *Quantitative Structure-Activity Relationships*, *16*(2), 107–112. <https://doi.org/10.1002/qsar.19970160202>
- Gadaleta, D., Mangiatordi, G. F., Catto, M., Carotti, A., & Nicolotti, O. (2016). Applicability Domain for QSAR Models. *International Journal of Quantitative Structure-Property Relationships*, *1*(1), 45–63. <https://doi.org/10.4018/ijqspr.2016010102>
- García-Domenech, R., García-Mujica, P., Gil, Ú., Beltrán, J. M., & Gálvez, J. (2010).

- Búsqueda de modelos QSAR para la actividad repelente de sesquiterpenos naturales frente al mosquito de la fiebre amarilla, *Aedes aegypti*. *Afinidad Revista de Química Teórica y Aplicada*, 67(547), 187–192. <https://raco.cat/index.php/afinidad/article/view/269141>
- Geldof, A. A., Engel, C., & Rao, B. R. (1992). Estrogenic action of commonly used fragrant agent citral induces prostatic hyperplasia. *Urological Research*, 20, 139–144. <https://doi.org/10.1007/BF00296526>
- Giaginis, C., Tsantili-Kakoulidou, A., & Theocharis, S. (2014). Applying quantitative structure-activity relationship (QSAR) methodology for modeling postmortem redistribution of benzodiazepines and tricyclic antidepressants. *Journal of Analytical Toxicology*, 38(5), 242–248. <https://doi.org/10.1093/jat/bku025>
- Golbraikh, A., Shen, M., Xiao, Z., Xiao, Y. De, Lee, K. H., & Tropsha, A. (2003). Rational selection of training and test sets for the development of validated QSAR models. *Journal of Computer-Aided Molecular Design*, 17(2–4), 241–253. <https://doi.org/10.1023/A:1025386326946>
- Golbraikh, A., & Tropsha, A. (2002). Beware of q²! *Journal of Molecular Graphics and Modelling*, 20(4), 269–276. [https://doi.org/10.1016/S1093-3263\(01\)00123-1](https://doi.org/10.1016/S1093-3263(01)00123-1)
- Golbraikh, A., Wang, X. S., Zhu, H., & Tropsha, A. (2017). Predictive QSAR modeling: Methods and applications in drug discovery and chemical risk assessment. In J. Leszczynski (Ed.), *Handbook of Computational Chemistry* (pp. 1–38). https://doi.org/10.1007/978-3-319-27282-5_37
- Govindarajan, M., Sivakumar, R., Rajeswari, M., & Yogalakshmi, K. (2012). Chemical composition and larvicidal activity of essential oil from *Mentha spicata* (Linn.) against three mosquito species. *Parasitology Research*, 110(5), 2023–2032. <https://doi.org/10.1007/s00436-011-2731-7>
- Gramatica, P. (2007). Principles of QSAR models validation: internal and external. *QSAR & Combinatorial Science*, 26(5), 694–701. <https://doi.org/10.1002/qsar.200610151>
- Gramatica, P. (2012). On the Development and Validation of QSAR Models. In B. Reisfeld & A. N. Mayeno (Eds.), *Computacional Toxicology* (pp. 499–526). <https://doi.org/10.1007/978-1-62703-059-5>
- Gramatica, P. (2020). Principles of QSAR Modeling: Comments and Suggestions From Personal Experience. *International Journal of Quantitative Structure-Property Relationships*, 5(3), 61–97. <https://doi.org/10.4018/ijqspr.20200701.0a1>
- Gramatica, P., & Sangion, A. (2016). A Historical Excursus on the Statistical Validation Parameters for QSAR Models: A Clarification Concerning Metrics and Terminology. *Journal of Chemical Information and Modeling*, 56(6), 1127–1131. <https://doi.org/10.1021/acs.jcim.6b00088>
- Grisoni, F., Consonni, V., & Todeschini, R. (2018). Impact of Molecular Descriptors on Computational Models. In O. Nicolotti (Ed.), *Methods in Molecular Biology* (pp. 171–209). https://doi.org/10.1007/978-1-4939-8639-2_5
- Grodniczky, J. A., & Coats, J. R. (2002). QSAR Evaluation of Monoterpenoids' Insecticidal Activity. *Journal of Agricultural and Food Chemistry*, 50, 4576–4580.
- Guha, R., & Willighagen, E. (2012). A Survey of Quantitative Descriptions of Molecular Structure. *Current Topics in Medicinal Chemistry*, 12(18), 1946–1956. <https://doi.org/10.2174/1568026611212180002>
- Gutiérrez-Moreno, R., Mota-Sanchez, D., Blanco, C. A., Whalon, M. E., Terán-Santofimio, H., Rodríguez-Maciel, J. C., & Difonzo, C. (2019). Field-Evolved Resistance of the Fall Armyworm (Lepidoptera: Noctuidae) to Synthetic Insecticides in Puerto Rico and Mexico. *Journal of Economic Entomology*, 112(2), 792–802.

- <https://doi.org/10.1093/jee/toy372>
- Hall, G. G. (1950). The molecular orbital theory of chemical valency. VI. Properties of equivalent orbitals. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 202(1070), 336–344. <https://doi.org/10.1098/rspa.1950.0104>
- Hall, G. G. (1991). The Lennard-Jones paper of 1929 and the foundations of Molecular Orbital Theory. *Advances in Quantum Chemistry*, 22, 1–6. [https://doi.org/10.1016/S0065-3276\(08\)60361-5](https://doi.org/10.1016/S0065-3276(08)60361-5)
- Hansch, C. (1976). On the Structure of Medicinal Chemistry. *Journal of Medical Chemistry*, 19(1), 1–6. <https://doi.org/https://doi.org/10.1021/jm00223a001>
- Hansch, C., Kurup, A., Garg, R., & Gao, H. (2001). Chem-bioinformatics and QSAR: A review of QSAR lacking positive hydrophobic terms. *Chemical Reviews*, 101(3), 619–672. <https://doi.org/10.1021/cr0000067>
- Hansch, C., Leo, A., & Elkins, D. (1974). Computerized Management of Structure-Activity Data. I. Multivariate Analysis of Biological Data. *Computerized Management of Structure-Activity Data*, 1(2), 57–61. <https://doi.org/10.1021/c160053a002>
- Hansch, C., Muir, R. M., Fujita, T., Veldstra, H., Maloney, P. P., & Muir, R. (1969). A Quantitative Approach to Biochemical Structure-Activity Relationships. *Acc. Chem. Res.*, 2(4), 232–239. <https://doi.org/10.1021/ar50020a002>
- Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E., & Hutchison, G. R. (2012). Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics*, 4(8). <https://doi.org/10.1186/1758-2946-4-17>
- Hardke, J. T., Lorenz, G. M., & Leonard, B. R. (2015). Fall armyworm (Lepidoptera: Noctuidae) ecology in Southeastern cotton. *Journal of Integrated Pest Management*, 6(1), 1–8. <https://doi.org/10.1093/jipm/pmv009>
- Harrell, F. E. (2001). Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. In *Statistical Methods in Medical Research*. <https://doi.org/10.1177/096228020401300512>
- Harve, G., & Kamath, V. (2004). Larvicidal activity of plant extracts used alone and in combination with known synthetic larvicidal agents against *Aedes aegypti*. *Indian Journal of Experimental Biology*, 42, 1216–1219. [http://nopr.niscair.res.in/bitstream/123456789/23822/1/IJEB_42\(12\)_1216-1219.pdf](http://nopr.niscair.res.in/bitstream/123456789/23822/1/IJEB_42(12)_1216-1219.pdf)
- He, H., Qin, X., Dong, F., Ye, J., Xu, C., Zhang, H., Liu, Z., Lv, X., Wu, Y., Jiang, X., & Cheng, X. (2020). Synthesis, characterization of two matrine derivatives and their cytotoxic effect on Sf9 cell of *Spodoptera frugiperda*. *Scientific Reports*, 10(1), 1–10. <https://doi.org/10.1038/s41598-020-75053-1>
- HE, L. mei, WU, Q. lin, GAO, X. wu, & WU, K. ming. (2021). Population life tables for the invasive fall armyworm, *Spodoptera frugiperda* fed on major oil crops planted in China. *Journal of Integrative Agriculture*, 20(3), 745–754. [https://doi.org/10.1016/S2095-3119\(20\)63274-9](https://doi.org/10.1016/S2095-3119(20)63274-9)
- Hernández-Carlos, B., & Gamboa-Angulo, M. (2019). Insecticidal and nematicidal contributions of Mexican flora in the search for safer biopesticides. *Molecules*, 24(5). <https://doi.org/10.3390/molecules24050897>
- Hernández-Trejo, A., Osorio-Hernández, E., López-Santillán, J. A., Ríos-Velasco, C., Varela-Fuentes, S., Rodríguez-Herrera, & R. (2018). Insectos benéficos asociados al control del gusano cogollero (*Spodoptera frugiperda*) en el cultivo de maíz (*Zea mays* L.). *AgroProductividad*, 11, 9–14. <https://revista-agroproductividad.org/index.php/agroproductividad/article/view/142>

- Høst, S., Olsen, J., Jansík, B., Thøgersen, L., Jørgensen, P., & Helgaker, T. (2008). The augmented Roothaan-Hall method for optimizing Hartree-Fock and Kohn-Sham density matrices. *Journal of Chemical Physics*, 129(12). <https://doi.org/10.1063/1.2974099>
- Huang, Y., Rong, C., Zhang, R., & Liu, S. (2017). Evaluating frontier orbital energy and HOMO/LUMO gap with descriptors from density functional reactivity theory. *Journal of Molecular Modeling*, 23(1). <https://doi.org/10.1007/s00894-016-3175-x>
- Ileer, V., Peralta, J., Palacios, C., & Burgos, A. (2022). Bioinsecticidas elaborados con extractos botánicos utilizados contra *Spodoptera* spp. en el cultivo de sandía (*Citrullus lanatus* T.) en Los Ríos-Ecuador. *Uniciencia*, 36(1), 1–11. <https://doi.org/10.15359/ru.36-1.42>
- JANPA. (2020). *The JANPA project: Program features in brief*. <http://janpa.sourceforge.net/>
- Jensen, F. (1999). *Introduction to computational chemistry*. [http://ktrunghuy.free.fr/SACH-BOOKS/Computational Chemistry/Introduction to computational chemistry.pdf](http://ktrunghuy.free.fr/SACH-BOOKS/Computational%20Chemistry/Introduction%20to%20computational%20chemistry.pdf)
- Jepson, P. C., Murray, K., Bach, O., Bonilla, M. A., & Neumeister, L. (2020). Selection of pesticides to reduce human and environmental health risks: a global guideline and minimum pesticides list. *The Lancet Planetary Health*, 4(2), e56–e63. [https://doi.org/10.1016/S2542-5196\(19\)30266-9](https://doi.org/10.1016/S2542-5196(19)30266-9)
- Kar, S., Harding, A. P., Roy, K., & Popelier, P. L. A. (2010). QSAR with quantum topological molecular similarity indices: Toxicity of aromatic aldehydes to *Tetrahymena pyriformis*. *SAR and QSAR in Environmental Research*, 21(1–2), 149–168. <https://doi.org/10.1080/10629360903568697>
- Kar, S., Roy, K., & Leszczynski, J. (2017). On Applications of QSARs in Food and Agricultural Sciences: History and Critical Review of Recent Developments. In K. Roy (Ed.), *Advances in QSAR Modeling, Challenges and Advances in Computational Chemistry and Physics* (pp. 203–302). <https://doi.org/10.1007/978-3-319-56850-8>
- Katritzky, A. R., Lobanov, V. S., & Karelson, M. (1994). *QSPR: The Correlation and Quantitative Prediction of Chemical and Physical Properties from Structure*. <https://doi.org/10.1039/CS9952400279>
- Kausar, S., & Falcao, A. O. (2018). An automated framework for QSAR model building. *Journal of Cheminformatics*, 10(1), 1–23. <https://doi.org/10.1186/s13321-017-0256-5>
- Kim, M.-C., Park, H., Son, S., Sim, E., & Burke, K. (2015). Improved DFT Potential Energy Surfaces via Improved Densities. *Journal of Physical Chemistry Letters*, 6(19), 3802–3807. <https://doi.org/10.1021/acs.jpcclett.5b01724>
- Kobayashi, Y., Uchida, T., & Yoshida, K. (2020). Prediction of Soil Adsorption Coefficient in Pesticides Using Physicochemical Properties and Molecular Descriptors by Machine Learning Models. *Environmental Toxicology and Chemistry*, 39(7), 1451–1459. <https://doi.org/10.1002/etc.4724>
- Kolossov, E., & Stanforth, R. (2010). The quality of QSAR models: problems and solutions. *SAR and QSAR in Environmental Research*, 18(1–2), 89–100. <https://doi.org/10.1080/10629360601053984>
- Kononov, D. A., Llewellyn, L. E., Heyden, Y. Vander, & Coomans, D. (2008). Robust cross-validation of linear regression QSAR models. *Journal of Chemical Information and Modeling*, 48(10), 2081–2094. <https://doi.org/10.1021/ci800209k>
- Krishnan, R., Binkley, J. S., Seeger, R., Pople, J. A., Krishnan, R., Binkley, J. S., Seeger, R., & Pople, J. A. (1980). Selfconsistent molecular orbital methods. XX. A basis set for correlated wave functions Self-consistent molecular orbital methods .XX. A basis set for correlated wave functions. *Journal of Chemical Physics*, 72, 650–654. <https://doi.org/10.1063/1.438955>

- Kwon, S., Bae, H., Jo, J., & Yoon, S. (2019). Comprehensive ensemble in QSAR prediction for drug discovery. *BMC Bioinformatics*, *20*(1), 1–12. <https://doi.org/10.1186/s12859-019-3135-4>
- Leach, A. R., & Gillet, V. J. (2007). *A Introduction to Chemoinformatics*. <https://doi.org/10.1017/CBO9781107415324.004>
- Lewars, E. G. (2016). *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics*. <https://doi.org/10.1002/anie.200485057>
- Leyva, M., Marquetti, C., Tacoronte, J. E., Scull, R., Tiomno, O., Mesa, A., & Montada, D. (2009). Actividad larvica de aceites esenciales de plantas contra *Aedes aegypti* (L.) (Diptera: Culicidae). *Biomédica*, *20*(1), 5–13. <https://www.medigraphic.com/pdfs/revbio/bio-2009/bio091b.pdf>
- Lima, A. P. S., Santana, E. D. R., Santos, A. C. C., Silva, J. E., Ribeiro, G. T., Pinheiro, A. M., Santos, Í. T. B. F., Blank, A. F., Araújo, A. P. A., & Bacci, L. (2020). Insecticide activity of botanical compounds against *Spodoptera frugiperda* and selectivity to the predatory bug *Podisus nigrispinus*. *Crop Protection*, *136*, 105230. <https://doi.org/10.1016/j.cropro.2020.105230>
- Lima, M. S., Silva, P. S. L., Oliveira, O. F., Silva, K. M. B., & Freitas, F. C. L. (2010). Corn yield response to weed and fall armyworm controls. *Planta Daninha*, *28*(1), 103–111. <https://doi.org/10.1590/s0100-83582010000100013>
- Lin, P., Qin, G., Zhang, N., Fan, J., Hao, X., Zou, L., & Ren, A. (2020). The roles of heteroatoms and substituents on the molecular packing motif from herringbone to π -stacking: A theoretical study on electronic structures and intermolecular interaction of pentacene derivatives. *Organic Electronics*, *78*, 1–9. <https://doi.org/10.1016/j.orgel.2019.105606>
- Lin, Z., Zhang, W., Pang, S., Huang, Y., Mishra, S., Bhatt, P., & Chen, S. (2020). Current approaches to and future perspectives on methomyl degradation in contaminated soil/water environments. *Molecules*, *25*(3), 1–16. <https://doi.org/10.3390/molecules25030738>
- Lira, E. C., Bolzan, A., Nascimento, A. R. B., Amaral, F. S. A., Kanno, R. H., Kaiser, I. S., & Omoto, C. (2020). Resistance of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) to spinetoram: inheritance and cross-resistance to spinosad. *Pest Management Science*, *76*(8), 2674–2680. <https://doi.org/10.1002/ps.5812>
- Liu, P., & Long, W. (2009). Current mathematical methods used in QSAR/QSPR studies. *International Journal of Molecular Sciences*, *10*(5), 1978–1998. <https://doi.org/10.3390/ijms10051978>
- Lizarazo, K., Mendoza, C., & Carrero, R. (2008). Efecto de extractos vegetales de polygonum hydropiperoides, Solanum nigrum y Calliandra pittieri sobre el gusano cogollero (*Spodoptera frugiperda*). *Agronomía Colombiana*, *26*(3), 427–434. <https://revistas.unal.edu.co/index.php/agrocol/article/view/11474/12120>
- Lokendra, O., Rachana, S., & Rani, B. M. (2013). Modern drug design with advancement in QSAR: A review. *International Journal of Research in BioSciences*, *2*(1), 1–12. https://www.idjsr.com/uploads/23/1224_pdf.pdf
- Manzhos, S. (2020). Machine learning for the solution of the Schrödinger equation. *Machine Learning: Science and Technology*, *1*(1), 013002. <https://doi.org/10.1088/2632-2153/ab7d30>
- Martin, Y. C. (2012). Hansch analysis 50 years on. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, *2*(3), 435–442. <https://doi.org/10.1002/wcms.1096>
- Masan, S. E. P. P., Febriana, F. N., Zaidan, A. H., Puspitasari, I., & Rusydi, F. (2021). Evaluation of the Electronic Structure Resulting from ab-initio Calculations on Simple

- Molecules Using the Molecular Orbital Theory. *Jurnal Penelitian Pendidikan IPA*, 7(1), 107. <https://doi.org/10.29303/jppipa.v7i1.545>
- Mohi El-Din, M. M., Mostafa, A. M., & Abd-Elkader, A. (2014). Experimental studies on the effect of (Lambda-Cyhalothrin) insecticide on lungs and the ameliorating effect of plant extracts (Ginseng (Panax Ginseng) and garlic (Allium sativum L.) on asthma development in albino rats. *BMC Research Notes*, 7(1), 1–10. <https://doi.org/10.1186/1756-0500-7-243>
- Morillo, F., & Notz, A. (2001). Resistencia de Spodoptera frugiperda (Smith) (Lepidoptera: Noctuidae) a lambdacihalotrina y metomil. *Entomotrópica: Revista Internacional Para El Estudio de La Entomología Tropical*, 16(2), 79–87. http://saber.ucv.ve/ojs/index.php/rev_ento/article/view/7648/7572
- Moriwaki, H., Tian, Y. S., Kawashita, N., & Takagi, T. (2018). Mordred: A molecular descriptor calculator. *Journal of Cheminformatics*, 10(1), 1–14. <https://doi.org/10.1186/s13321-018-0258-y>
- Mousavi, M., Ilkhani, a R., Sharifi, S., Mehrzad, J., Eghdami, A., & Monajjemi, M. (2013). DFT studies of nano anticancer on vinblastine and vincristine molecules. *International Journal of Microbiology Research and Reviews*, 1(2), 32–38. https://www.researchgate.net/profile/Majid-Monajjemi/publication/257935928_DFT_studies_of_nano_ant anticancer_on_vinblastine_and_vincristine_molecules/links/5767734208ae421c448c50c7/DFT-studies-of-nano-anticancer-on-vinblastine-and-vincristine-molecules.pdf
- Muhammad, U., Uzairu, A., & Arthur, E. (2018). Review on : quantitative structure activity relationship (QSAR) modeling. *Journal of Analytical & Pharmaceutical Research*, 7(2), 240–242. <https://doi.org/10.15406/japlr.2018.07.00232>
- Muratov, E. N., Bajorath, J., Sheridan, R. P., Tetko, I. V., Filimonov, D., Poroikov, V., Oprea, T. I., Baskin, I. I., Varnek, A., Roitberg, A., Isayev, O., Curtalolo, S., Fourches, D., Cohen, Y., Aspuru-Guzik, A., Winkler, D. A., Agrafiotis, D., Cherkasov, A., & Tropsha, A. (2020). QSAR without borders. *Chemical Society Reviews*, 49(11), 3525–3564. <https://doi.org/10.1039/d0cs00098a>
- Nagaraja, C. H. (2014). Introduction to R. In M. B. Rao & C. R. Rao (Eds.), *Handbook of Statistics: Computational Statistics with R* (Vol. 32, pp. 1–48). <https://doi.org/10.1016/B978-0-444-63431-3.00001-2>
- Nakagawa, Y., Smaghe, G., Paemel, M. Van, Tirry, L., & Fujita, T. (2001). Quantitative structure – activity studies of insect growth regulators: XVIII. Effects of substituents on the aromatic moiety of dibenzoylhydrazines on larvicidal activity against the Colorado potato beetle *Leptinotarsa decemlineata*†. *Pest Management Science*, 865(May), 117–123. <https://doi.org/10.1002/ps.376>
- Neese, F., Wennmohs, F., Becker, U., & Riplinger, C. (2020). The ORCA quantum chemistry program package. *The Journal of Chemical Physics*, 152(22), 224108. <https://doi.org/10.1063/5.0004608>
- Neves, B. J., Braga, R. C., Melo-Filho, C. C., Moreira-Filho, J. T., Muratov, E. N., & Andrade, C. H. (2018). QSAR-based virtual screening: Advances and applications in drug discovery. *Frontiers in Pharmacology*, 9, 1–7. <https://doi.org/10.3389/fphar.2018.01275>
- Nikolaienko, T. Y., Bulavin, L. A., & Hovorun, D. M. (2014). JANPA: an open source cross-platform implementation of the Natural Population Analysis on the Java platform. *Computational & Theoretical Chemistry*, 1050, 15–22. <https://doi.org/10.1016/j.comptc.2014.10.002>
- Nikolic, K., & Agababa, D. (2009). Design and QSAR study of analogs of γ -tocotrienol with

- enhanced antiproliferative activity against human breast cancer cells. *Journal of Molecular Graphics and Modelling*, 27(7), 777–783. <https://doi.org/10.1016/j.jmngm.2008.11.007>
- Njuguna, E., Nethononda, P., Maredia, K., Mbabazi, R., Kachapulula, P., Rowe, A., & Ndolo, D. (2021). Experiences and Perspectives on Spodoptera frugiperda (Lepidoptera: Noctuidae) Management in Sub-Saharan Africa. *Journal of Integrated Pest Management*, 12(1), 1–9. <https://doi.org/10.1093/jipm/pmab002>
- ORCA. (2017). *phpBB® Forum Software* ©. <https://orcaforum.kofo.mpg.de/app.php/portal>
- Ordóñez-García, M., Rios-Velasco, C., Ornelas-Paz, J. de J., Bustillos-Rodríguez, J. C., Acosta-Muñiz, C. H., Berlanga-Reyes, D. I., Salas-Marina, M. Á., Cambero-Campos, O. J., & Gallegos-Morales, G. (2020). Molecular and Morphological characterization of multiple nucleopolyhedrovirus from Mexico and their insecticidal activity against Spodoptera frugiperda (Lepidoptera: Noctuidae). *Journal of Applied Entomology*, 144(1–2), 123–132. <https://doi.org/10.1111/jen.12715>
- Paier, J., Marsman, M., & Kresse, G. (2007). Why does the B3LYP hybrid functional fail for metals? *Journal of Chemical Physics*, 127. <https://doi.org/https://doi.org/10.1063/1.2747249>
- Paluch, G., Grodnitzky, J., Bartholomay, L., & Coats, J. (2009). Quantitative structure-activity relationship of botanical sesquiterpenes: spatial and contact repellency to the yellow fever mosquito, *Aedes aegypti*. *Journal of Agricultural and Food Chemistry*, 57(16), 7618–7625. <https://doi.org/10.1021/jf900964e>
- Pérez-Gutiérrez, S., Zavala-Sánchez, M. A., González-Chávez, M. M., Cárdenas-Ortega, N. C., & Ramos-López, M. A. (2011). Bioactivity of *Carica papaya* (Caricaceae) against Spodoptera frugiperda (Lepidoptera: Noctuidae). *Molecules*, 16(9), 7502–7509. <https://doi.org/10.3390/molecules16097502>
- Perkins, R., Fang, H., Tong, W., & Welsh, W. J. (2003). Quantitative Structure-Activity Relationship Methods: Perspectives on Drug Discovery and Toxicology. *Environmental Toxicology and Chemistry*, 22(8), 1666–1679. <https://doi.org/10.1145/1459352.1459353>
- Phambala, K., Tembo, Y., Kasambala, T., Kabambe, V. H., Stevenson, P. C., & Belmain, S. R. (2020). Bioactivity of common pesticidal plants on fall Armyworm Larvae (*Spodoptera frugiperda*). *Plants*, 9(1), 1–10. <https://doi.org/10.3390/plants9010112>
- Pingaew, R., Mandi, P., Prachayasittikul, V., Thongnum, A., Prachayasittikul, S., Ruchirawat, S., & Prachayasittikul, V. (2021). Investigations on Anticancer and Antimalarial Activities of Indole-Sulfonamide Derivatives and in Silico Studies. *ACS Omega*, 6(47), 31854–31868. <https://doi.org/10.1021/acsomega.1c04552>
- Pirhadi, S., Shiri, F., & Ghasemi, J. B. (2015). Multivariate statistical analysis methods in QSAR. *The Royal Society of Chemistry*, 5(127), 104635–104665. <https://doi.org/10.1039/c5ra10729f>
- Pople, J. A., Adamson, R. D., & Gill, P. M. W. (1996). Density Functional Partitions. *Journal Physics Chemistry*, 100, 6348–6353. <https://doi.org/https://doi.org/10.1021/jp963467y>
- Put, R., Perrin, C., Questier, F., Coomans, D., Massart, D. L., & Vander Heyden, Y. (2003). Classification and regression tree analysis for molecular descriptor selection and retention prediction in chromatographic quantitative structure-retention relationship studies. *Journal of Chromatography A*, 988(2), 261–276. [https://doi.org/10.1016/S0021-9673\(03\)00004-9](https://doi.org/10.1016/S0021-9673(03)00004-9)
- Puzyn, T., Leszczynski, J., & Cronin, M. T. D. (2010). *Recent Advances in QSAR Studies: Methods and Applications*. <https://doi.org/10.1007/978-1-4020-9783-6>
- RácRác, A., Bajusz, D., & Héberger, K. (2019). Intercorrelation Limits in Molecular

- Descriptor Preselection for QSAR/QSPR. *Molecular Informatics*, 38(8–9), 1–7. <https://doi.org/10.1002/minf.201800154z>, A., Bajusz, D., & Héberger, K. (2019). Intercorrelation Limits in Molecular Descriptor Preselection for QSAR/QSPR. *Molecular Informatics*, 38(8–9), 1–7. <https://doi.org/10.1002/minf.201800154>
- Rácz, A., Bajusz, D., & Héberger, K. (2018). Modelling methods and cross-validation variants in QSAR: a multi-level analysis. *SAR and QSAR in Environmental Research*, 29(9), 661–674. <https://doi.org/10.1080/1062936X.2018.1505778>
- Randić, M. (1995). Molecular Shape Profiles. *Journal of Chemical Information and Computer Sciences*, 35(3), 373–382. <https://doi.org/10.1021/ci00025a005>
- Randić, M., & Basak, S. C. (1999). Optimal molecular descriptors based on weighted path numbers. *Journal of Chemical Information and Computer Sciences*, 39(2), 261–266. <https://doi.org/10.1021/ci9800763>
- Rasulev Saidkhodzhaev A, Nazrullaev SS, Akhmedkhodzhaeva K, Khushbaktova Z, L. J. (2007). Molecular modelling and QSAR analysis of the estrogenic activity of terpenoids isolated from *Ferula* plants. *Computational Center for Molecular Structure and Interactions*, 18, 663–673.
- Reséndiz-Ramírez, Z., López-Santillán, J. A., Estrada-Drouaillet, B., Osorio-Hernández, E., Pecina-Martínez, J. A., & Mendoza-Castillo, M. D. C. (2017). Efectos genéticos de la resistencia a *Spodoptera frugiperda* en líneas de maíz derivadas de germoplasma nativo de Tamaulipas. *Revista Mexicana de Ciencias Agrícolas*, 8(6), 1329. <https://doi.org/10.29312/remexca.v8i6.298>
- Rezić, I., Kracher, D., Oros, D., Mujadžić, S., Anđelini, M., Kurtanjek, Ž., Ludwig, R., & Rezić, T. (2022). Application of Causality Modelling for Prediction of Molecular Properties for Textile Dyes Degradation by LPMO. *Molecules*, 27(19). <https://doi.org/10.3390/molecules27196390>
- Rinnan, Å., Christensen, N. J., & Engelsen, S. B. (2010). How the energy evaluation method used in the geometry optimization step affect the quality of the subsequent QSAR/QSPR models. *Journal of Computer-Aided Molecular Design*, 24(1), 17–22. <https://doi.org/10.1007/s10822-009-9308-x>
- Ríos-Díez, J. D., & Saldamando-Benjumea, C. I. (2011). Susceptibility of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) strains from central Colombia to two insecticides, methomyl and lambda-cyhalothrin: A study of the genetic basis of resistance. *Journal of Economic Entomology*, 104(5), 1698–1705. <https://doi.org/10.1603/EC11079>
- Ríos-Díez, J. D., & Saldamando-Benjumea, C. I. (2011). Susceptibility of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) strains from central Colombia to two insecticides, methomyl and lambda-cyhalothrin: A study of the genetic basis of resistance. *Journal of Economic Entomology*, 104(5), 1698–1705. <https://doi.org/10.1603/EC11079>
- Romanelli, G. P., Virla, E. G., Duchowicz, P. R., Gaddi, A. L., Ruiz, D. M., Bennardi, D. O., Ortiz, E. D. V., & Autino, J. C. (2010). Sustainable Synthesis of Flavonoid Derivatives, QSAR Study and Insecticidal Activity against the Fall Armyworm, *Spodoptera frugiperda* (Lep. : Noctuidae). *Journal of Agricultural and Food Chemistry*, 58, 6290–6295. <https://doi.org/10.1021/jf100073j>
- Romo-Asunción, D., Martínez-González, D. E., Vásquez-Martínez, B. E., Ramos-López, M. A., Figueroa-Brito, R., Flores-Macías, A., Rea-López, M. Á., & Rodríguez-Muñoz, M. E. (2015). Compuestos botánicos como alternativa para el manejo del gusano cogollero del maíz. *Nthe*, December. <https://www.researchgate.net/publication/308087672>
- Roy, K. (2007). On some aspects of validation of predictive quantitative structure–activity relationship models. *Expert Opinion on Drug Discovery*, 2, 1567–1577.

- <https://doi.org/10.1517/17460441.2.12.1567>
- Roy, K., Kar, S., & Das, R. N. (2015a). Statistical Methods in QSAR/QSPR. In K. Roy, S. Kar, & R. N. Das (Eds.), *A Primer on QSAR/QSPR Modeling Fundamentals Concepts* (pp. 37–59). <https://doi.org/10.1007/978-3-319-17281-1>
- Roy, K., Kar, S., & Das, R. N. (2015b). *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*. <https://doi.org/10.1016/C2014-0-00286-9>
- Roy, P. P., Leonard, J. T., & Roy, K. (2008). Exploring the impact of size of training sets for the development of predictive QSAR models. *Chemometrics and Intelligent Laboratory Systems*, *90*(1), 31–42. <https://doi.org/10.1016/j.chemolab.2007.07.004>
- RStudio. (2021). *RStudio: Building Effective Data Science Teams*. <https://www.rstudio.com/>
- Rutte, R. R., Aguilar, R., Maldonado, E. A., & Ruiz, M. (2020). Native strains of *Bacillus thuringiensis* against *Spodoptera frugiperda* and *Alabama argillacea* in cotton crop (*Gossypium barbadense*) in Piura, Peru. *Chilean Journal of Agricultural and Animal Sciences*, *36*(1), 52–62. <https://doi.org/10.29393/CHJAAS36-2S30002>
- Sakura, J. J., & Napolitano, J. (1994). *Modern quantum mechanics*. <https://doi.org/10.1017/9781108587280>
- Saldamando, C. I., & Vélez-Arango, A. M. (2010). Host plant association and genetic differentiation of corn and rice strains of *Spodoptera frugiperda* Smith (Lepidoptera: Noctuidae) in Colombia. *Neotropical Entomology*, *39*(6), 921–929. <https://doi.org/10.1590/S1519-566X2010000600012>
- Salinas-Sánchez, D. O., Aldana-Llanos, L., Valdés-Estrada, M. E., Gutiérrez-Ochoa, M., Valladares-Cisneros, G., & Rodríguez-Flores, E. (2012). Insecticidal activity of *Tagetes erecta* extracts on *spodoptera frugiperda* (Lepidoptera: Noctuidae). *Florida Entomologist*, *95*(2), 428–432. <https://doi.org/10.1653/024.095.0225>
- Santiago, G. P., Pádua, L. E. de M., Silva, P. R. R., Carvalho, E. M. S., & Maia, C. B. (2008). Effects of plant extracts on the biology of *Spodoptera frugiperda* (J. E. Smith, 1797) (Lepidoptera: Noctuidae) maintained under artificial diet. *Ciência e Agrotecnologia*, *32*(3), 792–796. <https://www.scielo.br/j/cagro/a/ZH9sRr67hYXRNN6gXqNypqn/?lang=pt>
- Santos Amaya, O., Delgado Restrepo, O., Argüelles, J., & Aguilera Garramuño, E. (2009). Evaluación del comportamiento del complejo *Spodoptera* con la introducción de algodón transgénico al Tolima, Colombia. *Corpoica Ciencia y Tecnología Agropecuaria*, *10*(1), 24. https://doi.org/10.21930/rcta.vol10_num1_art:125
- Santos, S. R. L., Silva, V. B., & Barbosa, J. D. F. (2010). Toxic Effects on and Structure-Toxicity Relationships of Phenylpropanoids, Terpenes, and Related Compounds. *Vector-Borne and Zoonotic Diseases*, *10*(10), 1049–1054. <https://doi.org/10.1089/vbz.2009.0158>
- Saxena, A. K., & Prathipati, P. (2003). Comparison of MLR, PLS and GA-MLR in QSAR analysis. *SAR and QSAR in Environmental Research*, *14*(5–6), 433–445. <https://doi.org/10.1080/10629360310001624015>
- Schüürmann, G., Ebert, R. U., Chen, J., Wang, B., & Kühne, R. (2008). External validation and prediction employing the predictive squared correlation coefficient -Test set activity mean vs training set activity mean. *Journal of Chemical Information and Modeling*, *48*(11), 2140–2145. <https://doi.org/10.1021/ci800253u>
- Shahlaei, M. (2013). Descriptor selection methods in quantitative structure-activity relationship studies: A review study. *Chemical Reviews*, *113*(10), 8093–8103. <https://doi.org/10.1021/cr3004339>
- Siddhartha, K., Chinniah, C., & Shanthi, M. (2019). In vitro bioassay of certain botanical oils

- for their efficacy against maize fall army worm (J. E. Smith) *Spodoptera frugiperda* (Noctuidae: Lepidoptera). *Journal of Entomology and Zoology Studies*, 7(5), 606–609. <http://www.fao.org/3/a-bt415e.pdf>.
- Sisay, B., Simiyu, J., Malusi, P., Likhayo, P., Mendesil, E., Elibariki, N., Wakgari, M., Ayalew, G., & Tefera, T. (2018). First report of the fall armyworm, *Spodoptera frugiperda* (Lepidoptera: Noctuidae), natural enemies from Africa. *Journal of Applied Entomology*, 142(8), 800–804. <https://doi.org/10.1111/jen.12534>
- Soares Rodrigues, G. C., Dos Santos M., M., Muratov, E. N., Scotti, L., & Scotti, M. T. (2020). Quantitative Structure-Activity Relationship Modeling and Docking of Monoterpenes with Insecticidal Activity against *Reticulitermes chinensis* Snyder and *Drosophila melanogaster*. *Journal of Agricultural and Food Chemistry*, 68(16), 4687–4698. <https://doi.org/10.1021/acs.jafc.0c00272>
- Soberón Risco, G. V., Idrogo, C. R., Kato, M. J., Díaz, J. S., Armando-Jr, J., & Delgado Paredes, G. E. (2012). Larvicidal activity of *Piper tuberculatum* on *Spodoptera frugiperda* (Lepidoptera: Noctuidae) under laboratory conditions. *Revista Colombiana De Entomología*, 38(1), 35–41. <http://www.scielo.org.co/pdf/rcen/v38n1/v38n1a06.pdf>
- Sombra, K. E. S., de Aguiar, C. V. S., de Oliveira, S. J., Barbosa, M. G., Zocolo, G. J., & Pastori, P. L. (2020). Potential pesticide of three essential oils against *Spodoptera frugiperda* (J.E. smith) (Lepidoptera: Noctuidae). *Chilean Journal of Agricultural Research*, 80(4), 617–628. <https://doi.org/10.4067/S0718-58392020000400617>
- Stone, M., & Jonathan, P. (1993). Statistical thinking and technique for QSAR and related studies. Part I: General theory. *Journal of Chemometrics*, 7(1), 455–475. <https://doi.org/10.1002/cem.1180070603>
- Tandon, H., Chakraborty, T., & Suhag, V. (2019). A Concise Review on the Significance of QSAR in Drug Design. *Chemical and Biomolecular Engineering*, 4(4), 45. <https://doi.org/10.11648/j.cbe.20190404.11>
- Todeschini, R., & Consonni, V. (2000). *Handbook of Molecular Descriptors*. <https://doi.org/10.1002/9783527610907.scard>
- Todeschini, R., & Consonni, V. (2009). Molecular Descriptors for Chemoinformatics. In *Recent Advances in QSAR Studies*. <https://doi.org/10.1002/9783527628766>
- Todeschini, R., Consonni, V., & P. Gramatica. (2009). Chemometrics in QSAR. In Elsevier (Ed.), *Comprehensive Chemometrics* (1st ed., Vol. 4, pp. 129–172, 144).
- Todeschini, R., Gramatica, P., Provenzani, R., & Marengo, E. (1995). Weighted holistic invariant molecular descriptors. Part 2. Theory development and applications on modeling physicochemical properties of polyaromatic hydrocarbons. *Chemometrics and Intelligent Laboratory Systems*, 25, 221–229. [https://doi.org/10.1016/0169-7439\(95\)80026-6](https://doi.org/10.1016/0169-7439(95)80026-6)
- Toppur, B., & Jaims, K. J. (2021). Determining the best set of molecular descriptors for a Toxicity classification problem. *RAIRO - Operations Research*, 55(5), 2769–2783. <https://doi.org/10.1051/ro/2021134>
- Triola, M. F. (2004). *Estadística* (Pearson Educación (ed.); 9th ed.).
- Tropsha, A., & Golbraikh, A. (2007). Predictive QSAR Modeling Workflow, Model Applicability Domains, and Virtual Screening. *Current Pharmaceutical Design*, 3494–3504. <https://doi.org/10.2174/138161207782794257>
- Tropsha, A., Gramatica, P., & Gombar, V. K. (2003). The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR and Combinatorial Science*, 22(1), 69–77. <https://doi.org/10.1002/qsar.200390007>
- Trumper, E. V. (2014). Resistencia de insectos a cultivos transgénicos con propiedades

- insecticidas. Teoría, estado del arte y desafíos para la República Argentina. *AgriScientia*, 31(2), 109–126. <https://doi.org/10.31047/1668.298x.v31.n2.16538>
- Vanderkar, G., Jain, D., & Jain, S. (2007). Statistical concepts utilized to derive a QSAR and their application to pharmaceutical research. *Asian Journal of Chemistry*, 19(6), 4151–4160. http://www.asianjournalofchemistry.co.in/User/ViewFreeArticle.aspx?ArticleID=19_6_3
- Vargas-Méndez, L. Y., Sanabria-Flórez, P. L., Saavedra-Reyes, L. M., Merchan-Arenas, D. R., & Kouznetsov, V. V. (2019). Bioactivity of semisynthetic eugenol derivatives against *Spodoptera frugiperda* (Lepidoptera: Noctuidae) larvae infesting maize in Colombia. *Saudi Journal of Biological Sciences*, 26(7), 1613–1620. <https://doi.org/10.1016/j.sjbs.2018.09.010>
- Vásquez, J., Zeddám, J. L., & Tresierra, A. A. (2006). Control Biológico del "cogollero del maíz" *Spodoptera frugiperda*, (Lepidoptera; Noctuidae) con el Baculovirus SfVPN, en Iquitos-Perú. *Folia Amazónica*, 13(1–2), 25. <https://doi.org/10.24841/fa.v13i1-2.132>
- Veerasingam, R., Rajak, H., Jain, A., Sivadasan, S., Varghese, C. P., & Agrawal, R. K. (2011). Validation of QSAR Models - Strategies and Importance. *International Journal of Drug Design and Discovery*, 2(3), 511–519. <https://dokumen.tips/documents/validation-of-qsar-models-strategies-and-veerasamy-et-al-validation-of.html>
- Villaverde, J. J., Sevilla-Morán, B., López-Goti, C., Alonso-Prados, J. L., & Sandín-España, P. (2020). QSAR/QSPR models based on quantum chemistry for risk assessment of pesticides according to current European legislation. *SAR and QSAR in Environmental Research*, 31(1), 49–72. <https://doi.org/10.1080/1062936X.2019.1692368>
- Vreven, T., Frisch, M. J., Kudin, K. N., Schlegel, H. B., & Morokuma, K. (2006). Geometry optimization with QM/MM methods II: Explicit quadratic coupling. *Molecular Physics*, 104(5–7), 701–714. <https://doi.org/10.1080/00268970500417846>
- Vukovic, K., Gadaleta, D., & Benfenati, E. (2019). Methodology of AiQSAR: A group-specific approach to QSAR modelling. *Journal of Cheminformatics*, 11(1), 1–9. <https://doi.org/10.1186/s13321-019-0350-y>
- Walker, J. D., Jaworska, J., Comber, M. H. I., Schultz, T. W., & Dearden, J. C. (2003). Guidelines for developing and using quantitative structure-activity relationships. *Environmental Toxicology and Chemistry*, 22(8), 1653–1665. <https://doi.org/10.1897/01-627>
- WAN, J., HUANG, C., LI, C. you, ZHOU, H. xu, REN, Y. lin, LI, Z. yuan, XING, L. sheng, ZHANG, B., QIAO, X., LIU, B., LIU, C. hui, XI, Y., LIU, W. xue, WANG, W. kai, QIAN, W. qiang, MCKIRDY, S., & WAN, F. hao. (2021). Biology, invasion and management of the agricultural invader: Fall armyworm, *Spodoptera frugiperda* (Lepidoptera: Noctuidae). *Journal of Integrative Agriculture*, 20(3), 646–663. [https://doi.org/10.1016/S2095-3119\(20\)63367-6](https://doi.org/10.1016/S2095-3119(20)63367-6)
- Wan, J., Zhang, L., Yang, G., & Zhan, C. G. (2004). Quantitative structure-activity relationship for cyclic imide derivatives of protoporphyrinogen oxidase inhibitors: A study of quantum chemical descriptors from density functional theory. *Journal of Chemical Information and Computer Sciences*, 44(6), 2099–2105. <https://doi.org/10.1021/ci049793p>
- Wang, Y., Shao, Y., Wang, Y., Fan, L., Yu, X., Zhi, X., Yang, C., Qu, H., Yao, X., & Xu, H. (2012). Synthesis and quantitative structure-activity relationship (QSAR) study of novel isoxazoline and oxime derivatives of podophyllotoxin as insecticidal agents. *Journal of Agricultural and Food Chemistry*, 60, 8435–8443. <https://doi.org/10.1021/jf303069v>
- Weigend, F., & Ahlrichs, R. (2005). Balanced basis sets of split valence, triple zeta valence

- and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.*, 7(1), 3297–3305. <https://doi.org/https://doi.org/10.1039/B508541A>
- Winkler, D. A. (2002). The role of quantitative structure-activity relationships (QSAR) in biomolecular discovery. In H. Stewart (Ed.), *Briefings in bioinformatics* (pp. 73–86). <https://doi.org/10.1093/bib/3.1.73>
- Wold, S. (1978). Cross-Validatory Estimation of the Number of Components in Factor and Principal Components Models. *Technometrics*, 20(4), 397–405. <https://doi.org/10.1080/00401706.1978.10489693>
- Wold, S., & Eriksson, L. (1995). Statistical Validation of QSAR Results. In H. van Waterbeemd (Ed.), *Chemometric Methods in Molecular Design* (pp. 309–338). <https://doi.org/10.1002/9783527615452.ch5>
- Wood, D. J., Carlsson, L., Eklund, M., Norinder, U., & Stålring, J. (2013). QSAR with experimental and predictive distributions: An information theoretic approach for assessing model quality. *Journal of Computer-Aided Molecular Design*, 27(3), 203–219. <https://doi.org/10.1007/s10822-013-9639-5>
- Xia, L. Y., Wang, Q. Y., Cao, Z., & Liang, Y. (2019). Descriptor Selection Improvements for Quantitative Structure-Activity Relationships. *International Journal of Neural Systems*, 29(9). <https://doi.org/10.1142/S0129065719500163>
- Xia, L. Y., Wang, Y. W., Meng, D. Y., Yao, X. J., Chai, H., & Liang, Y. (2018). Descriptor selection via log-sum regularization for the biological activities of chemical structure. *International Journal of Molecular Sciences*, 19(1), 1–15. <https://doi.org/10.3390/ijms19010030>
- Yap, C. W. (2010). Software News and Update PaDEL-Descriptor: An Open Source Software to Calculate Molecular Descriptors and Fingerprints. *Journal of Computational Chemistry*, 32(7), 1466–1474. <https://doi.org/10.1002/jcc>
- Yap, C. W. (2014). *PaDEL-Descriptor*. <http://www.yapcwsoft.com/dd/padeldescriptor/>
- Yap, C. W., Li, H., Ji, Z. L., & Chen, Y. Z. (2007). Regression Methods for Developing QSAR and QSPR Models to Predict Compounds of Specific Pharmacodynamic, Pharmacokinetic and Toxicological Properties. *Mini-Reviews in Medicinal Chemistry*, 7(11), 1097–1107. <https://doi.org/10.2174/138955707782331696>
- Yee, L. C., & Wei, Y. C. (2012). Current Modeling Methods Used in QSAR/QSPR. In M. Dehmer & K. Varmuza (Eds.), *Statistical Modelling of Molecular Descriptors in QSAR/QSPR* (pp. 1–31). https://beckassets.blob.core.windows.net/product/readingsample/9492480/9783527324347_excerpt_001.pdf
- Zenner de Polanía, I., & Borrero Fonseca, F. (1996). Resistencia del cogollero del maíz, *Spodoptera frugiperda* (J.E. Smith) a algunos insecticidas y su manejo. In *Manejo Integrado de Plagas y Enfermedades* (pp. 39–50). <http://hdl.handle.net/20.500.12324/1219>
- Zenner de Polanía, I., Maldonado, H. A. A., Cruz, R. M., & Sánchez, J. L. D. (2009). *Spodoptera frugiperda*: Response of different populations to the Cry1Ab toxin. *Revista Colombiana de Entomología*, 35(1), 34–41. <https://doi.org/http://www.scielo.org.co/pdf/rcen/v35n1/v35n1a07.pdf>
- Zhao, J., Zhang, J., Xu, B., Wang, Z., Cheng, J., & Zhu, G. (2012). Design, Synthesis, and Analysis of the Quantitative Structure–Activity Relationships of 4-Phenyl-acyl-substituted 3-(2,5-Dimethylphenyl)-4-hydroxy-1-azaspiro[4.5]dec-3-ene-2,8-dione Derivatives. *J. Agric. Food Chem.*, 60, 4779–4787. <https://doi.org/10.1021/jf3002069>
- Zhao, Y. X., Huang, J. M., Ni, H., Guo, D., Yang, F. X., Wang, X., Wu, S. F., & Gao, C. F.

(2020). Susceptibility of fall armyworm, *Spodoptera frugiperda* (J.E.Smith), to eight insecticides in China, with special reference to lambda-cyhalothrin. *Pesticide Biochemistry and Physiology*, 168, 1–8. <https://doi.org/10.1016/j.pestbp.2020.104623>

ANEXOS

Anexo 1. Matriz de las propiedades mecanocuánticas y termoquímicas de los compuestos del estudio.

Compuesto químico	Energía E. Total (Kcal/mol)	E _{HOMO} (eV)	E _{LUMO} (eV)	Gap (eV)	Momento dipolar (D)	Entropía	Energía Libre de Gibbs (Kcal/mol)
Eugenol 1	-682088,8996	-5,289	0,347	-5,636	0,61322	0,04818	-337729,8329
Eugenol 2	-682082,3296	-5,184	0,356	-5,540	0,15100	0,05324	-362360,3296
Eugenol 3	-651056,6544	-5,284	0,347	-5,631	0,53020	0,06127	-435514,7549
Eugenol 4	-507385,1442	-5,174	0,408	-5,582	0,73002	0,06453	-460151,9659
Eugenol 5	-507378,3865	-5,275	0,097	-5,372	0,67713	0,06374	-507208,0394
Eugenol 6	-481505,0298	-5,943	-0,146	-5,797	5,26048	0,05798	-433449,9563
Eugenol 7	-460326,8475	-6,139	-2,834	-3,305	2,49893	0,06833	-681922,1937
Eugenol 8	-435672,7717	-5,047	-0,298	-4,749	0,75800	0,05044	-337739,1199
Eugenol 9	-433583,7219	-4,985	-0,290	-4,695	0,31878	0,05359	-362367,4832
Eugenol 10	-433576,2733	-5,072	-0,382	-4,690	0,73712	-808.3076	-507215,4439
Eugenol 11	-362488,9472	-5,726	-0,809	-4,917	5,53761	0,05793	-433457,2980
Eugenol 12	-362482,0912	-5,842	-2,852	-2,990	2,92984	0,07114	-681929,9747
Eugenol 13	-338608,5212	-5,192	0,479	-5,671	1,00349	0,04921	-338488,6095
Eugenol 14	-337843,5980	-5,402	0,224	-5,626	1,295562	0,05650	-481356,8610
Eugenol 15	-337835,1634	-5,404	0,045	-5,449	3,440459	0,06999	-650844,4801
Amida 1	-681091,8597	-6,102	-1,408	-4,694	4,433	0,06562	-680929,8604
Amida 2	-632553,7361	-6,348	-0,858	-5,490	4,195	0,05836	-632410,6382
Amida 3	-615458,3861	-5,686	-0,607	-5,079	2,772	0,06160	-492811,3008
Amida 4	-589288,5814	-5,725	-0,581	-5,144	2,414	0,05443	-443540,1418
Amida 5	-541506,7870	-5,707	-0,610	-5,097	2,764	0,06264	-517445,9391
Amida 6	-540746,3889	-5,482	-1,234	-4,248	2,408	0,06467	-541330,0210
Amida 7	-517620,3095	-5,458	-1,261	-4,197	3,052	0,07546	-615230,6728
Amida 8	-516091,0591	-5,986	-1,224	-4,762	3,783	-672.9929	34,82641650
Amida 9	-492967,6351	-5,476	-1,128	-4,348	4,005	0,05887	-540579,5902
Amida 10	-443663,1519	-5,272	-1,487	-3,785	4,065	0,06536	-589102,2637
Amida 11	-422497,0160	-5,483	-1,190	-4,293	4,070	0,05783	-515942,6929
Terpeno 1	-943246,1056	-4,991	-1,033	-3,958	1,459	0,09724	-847184,6468
Terpeno 2	-877532,7102	-5,058	-0,912	-4,146	1,359	0,09803	-871811,3158
Terpeno 3	-876483,5844	-5,056	-1,504	-3,552	3,402	0,10206	-942893,1613
Terpeno 4	-872915,6338	-5,185	-0,955	-4,230	1,097	0,09712	-847921,3980
Terpeno 5	-872909,3160	-5,170	-0,855	-4,315	0,147	0,10078	-872549,1338
Terpeno 6	-872159,4201	-5,232	-0,943	-4,676	0,162	0,14551	-876885,5560
Terpeno 7	-848265,2335	-6,006	-3,166	-2,840	1,189	0,09900	-847169,6494
Terpeno 8	-847515,0765	-6,095	-0,253	-5,458	2,075	0,094122	-874465,5907
Terpeno 9	-847497,8815	-6,155	-0,373	-5,782	2,160	0,09609	-876055,0558
Terpeno 10	-679312,7570	-5,991	0,281	-6,272	2,694	0,07214	-654396,0843
Terpeno 11	-654671,4894	-5,868	0,263	-6,131	2,983	0,07404	-679020,3060
Terpeno 12	-291440,3543	-5,613	0,276	-5,889	2,058	0,04708	-291323,5388
Flavonoide 1	-2713889,948	-5,862	-1,145	-4,717	2,600	0,05679	-524793,3090
Flavonoide 2	-2713889,948	-5,243	-0,708	-4,535	6,180	0,05234	-153192,3074
Flavonoide 3	-2712367,610	-5,559	-1,359	-4,200	4,753	0,04789	-186431,1413
Flavonoide 4	-2710687,767	-5,634	-1,312	-4,246	4,563	0,04587	-148184,8335
Flavonoide 5	-2504284,170	-5,665	-1,250	-4,415	4,499	0,06779	-148398,1845

Flavonoide 6	-2416250,219	-5,788	-1,236	-4,552	2,290	0,06217	-130803,0004
--------------	--------------	--------	--------	--------	-------	---------	--------------

Anexo 2. ANOVA de los descriptores moleculares elegidos por el submétodo *Forward* en el intervalo $0,00 \leq \alpha \leq 0,10$ al ingresarlos en el modelo QSAR de derivados de eugenol.

Ecuación para calcular el parámetro de Fisher-Snedecor:

$$F = \frac{(R^2)(GL)}{(1 - R^2)(m)} \quad GL = (n - 1) - m$$

Donde: R^2 : Varianza explicada, $1 - R^2$: Varianza no explicada, GL : Grados de libertad, $n - 1$: Número de datos libres de variar (12 valores de actividad larvicida) y m : Número de parámetros o variables estimadas (la cantidad de descriptores).

Descriptores moleculares	R^2	EER	Distribución de Fisher-Snedecor (F)	GL	Significancia estadística (p)
nssO	0,4032	0,2407	9,108	11	$1,170 \times 10^{-2}$
MATS3m	0,5492	0,2092	8,310	10	$7,481 \times 10^{-5}$
Entropía	0,6605	0,1815	8,782	9	$4,881 \times 10^{-3}$
GATS3v	0,8383	0,1253	16,550	8	$6,172 \times 10^{-4}$
ATSC6e	0,9714	0,0527	82,400	7	$4,587 \times 10^{-6}$
AATSC8e	0,9957	0,0204	463,600	6	$9,936 \times 10^{-8}$
SsCH3	0,9979	0,01434	808,800	5	$2,509 \times 10^{-7}$
TDB9e	0,9993	0,00809	2,222	4	$5,057 \times 10^{-7}$
AATSC0c	1,0	0,00298	$1,454 \times 10^4$	3	$8,520 \times 10^{-7}$
AATSC5c	1,0	$1,067 \times 10^{-4}$	$1,022 \times 10^7$	2	$9,781 \times 10^{-8}$
TDB7p	1,0	$6,702 \times 10^{-8}$	$2,357 \times 10^{13}$	1	$1,606 \times 10^{-7}$

*Los valores del parámetro F calculados son promedios entre los F individuales de cada descriptor molecular a partir del segundo.

Anexo 3. ANOVA de los descriptores teóricos mecanocuánticos y termoquímicos incluidos al análisis QSAR de los derivados de eugenol.

Descriptores teóricos	R^2	EER	Fisher - Snedecor (F)	Significancia estadística (p)
Energía E. Total	0,1903	0,2924	0,4249	0,1361
Energía E_{HOMO}	0,0372	0,3189	0,4249	0,5278
Energía E_{LUMO}	0,0218	0,3214	0,2451	0,6303
Diferencia E. Gap	0,0158	0,3224	2,5858	0,6818
Momento dipolar	0,0972	0,3088	1,1843	0,2997

Anexo 4. Matriz de los descriptores moleculares del modelo QSAR de eugenoles para actividad larvicida.

Compuesto químico	nssO	MATS3m	Entropía	GATS3v	ATSC6e	AATSC8e
Eugenol 1	1	-0,0386	30,232	1,1101	$-1,61 \times 10^{-1}$	-0,0154
Eugenol 2	2	-0,0010	33,407	1,0842	$-1,30 \times 10^{-1}$	-0,0045
Eugenol 3	2	-0,1067	38,447	1,2437	-0,1783	-0,0011

Eugenol 4	2	-0,1046	40,493	1,2050	-0,2906	-0,0129
Eugenol 5	2	-0,0416	39,994	1,0997	-0,2488	-0,0057
Eugenol 6*	2	-0,12631	36,383	1,0979	-8,82x10 ⁻¹	-0,0078
Eugenol 7	2	0,03418	42,880	1,0421	0,1357	-0,0084
Eugenol 8	1	0,00040	31,649	1,0634	2,43x10 ⁻²	-0,0231
Eugenol 9	2	0,0376	33,629	1,0394	0,0387	-0,0144
Eugenol 10	2	-0,0177	-507.220,623	1,0733	-0,0996	-0,0084
Eugenol 11*	2	-0,0982	36,348	0,9332	-0,6636	-0,0078
Eugenol 12	2	0,0559	44,643	1,0074	0,44035	-0,0083
Eugenol 13	1	0,0624	30,879	1,0327	-0,2517	-0,0135
Eugenol 14	3	-0,0603	35,457	1,2074	1,0456	-0,0049
Eugenol 15	2	0,0240	43,920	1,0342	-0,9000	0,00139

*Compuestos de prueba.

Anexo 5. Resultados de actividad larvica predicha de la forma $\text{Log}_{10}(\text{DL}_{50})$ para los compuestos a partir del modelo QSAR de eugenoles.

Compuesto químico	DL ₅₀ (mg/g de insecto)	Log ₁₀ (DL ₅₀) experimental	Log ₁₀ (DL ₅₀) predicha	Residuales (RES)
Eugenol 1	0,49	-0,309	-0,3282	1,8263x10 ⁻²
Eugenol 2	2,02	0,305	0,3160	-1,1047x10 ⁻²
Eugenol 3	1,93	0,285	0,2945	-8,5925x10 ⁻³
Eugenol 4	0,88	-0,055	-0,0649	8,9498x10 ⁻³
Eugenol 5	0,90	-0,045	-0,0281	-1,7885x10 ⁻²
Eugenol 7	2,03	0,307	0,3344	-2,7455x10 ⁻²
Eugenol 8	0,46	-0,337	-0,3291	-7,8880x10 ⁻³
Eugenol 9	2,03	0,307	0,3056	1,3025x10 ⁻³
Eugenol 10	4,25	0,628	0,6279	4,0567x10 ⁻⁸
Eugenol 12	2,02	0,305	0,2860	1,8912x10 ⁻²
Eugenol 13	2,02	0,305	0,3063	-1,3978x10 ⁻³
Eugenol 14	4,24	0,627	0,6180	8,9777x10 ⁻³
Eugenol 15	2,83	0,451	0,4341	1,7861x10 ⁻²

Anexo 6. ANOVA de los descriptores moleculares elegidos por el submétodo *Forward* en el intervalo $0,00 \leq \alpha \leq 0,10$ al ingresarlos en el modelo QSAR de amidas sintéticas.

Ecuación para calcular el parámetro de Fisher-Snedecor:

$$F = \frac{(R^2)(GL)}{(1 - R^2)(m)} \quad GL = (n - 1) - m$$

Donde: R^2 : Varianza explicada, $1 - R^2$: Varianza no explicada, GL : Grados de libertad, $n - 1$: Número de datos libres de variar (12 valores de actividad larvica) y m : Número de parámetros o variables estimadas (la cantidad de descriptores).

Descriptores moleculares	R ²	EER	Distribución de Fisher-Snedecor (F)	GL	Significancia estadística (p)
AlogP	0,9452	0,2160	120,80	7	1,143x10 ⁻⁵
ZMIC1	0,9487	0,2258	55,53	6	1,347x10 ⁻⁴

Anexo 7. ANOVA de los descriptores teóricos mecanocuánticos y termoquímicos incluidos al análisis QSAR de los derivados de amidas sintéticas.

Descriptores teóricos	R ²	EER	Fisher - Snedecor (F)	Significancia estadística (p)
Energía E. Total	0,1148	0,8685	0,9079	3,72x10 ⁻¹
Energía E _{HOMO}	0,0400	0,9026	0,3222	5,88x10 ⁻¹
Energía E _{LUMO}	0,0246	0,9117	0,1765	6,87x10 ⁻¹
Diferencia E. Gap	0,0511	0,8992	0,3771	5,59 x10 ⁻¹
Momento dipolar	0,0000	0,9229	0,0043	9,49x10 ⁻¹
Entropía	0,1550	0,8486	1,2840	2,94x10 ⁻¹
Energía L. Gibbs	0,0082	0,9193	0,0584	8,16x10 ⁻¹

Anexo 8. Matriz de los descriptores moleculares del modelo QSAR de amidas sintéticas para actividad larvicida.

Compuesto químico	ALogP
Amida 1	-0,6352
Amida 2	0,0947
Amida 3	-0,3268
Amida 4*	-0,8711
Amida 5	-0,6611
Amida 6	-0,3539
Amida 7	1,7199
Amida 8*	0,3261
Amida 9	-1,103
Amida 10	0,4083
Amida 11	1,8283

*Compuestos de prueba.

Anexo 9. Resultados de actividad larvicida predicha de la forma Log₁₀(DL₅₀) para los compuestos a partir del modelo QSAR de amidas sintéticas.

Compuesto químico	DL ₅₀ (µg/mg de larva)	Log ₁₀ (DL ₅₀) experimental	Log ₁₀ (DL ₅₀) predicha	Residuales (RES)
Amida 1	0,793	-0,101	0,1338	-0,2348
Amida 2	8,15	0,911	0,8731	0,0378
Amida 3	1,24	0,093	0,4462	-0,3532
Amida 5	1,35	0,130	0,1075	0,0224
Amida 6	2,63	0,420	0,4187	0,0012
Amida 7	388,8	2,590	2,5190	0,0709
Amida 9	14,14	1,150	1,1074	0,0425
Amida 10	1,07	0,029	-0,3398	0,3688
Amida 11	17,07	1,235	1,1907	0,0442

*Compuestos de prueba.

Anexo 10. ANOVA de los descriptores moleculares elegidos por el submétodo *Forward* en el intervalo 0,00 ≤ α ≤ 0,10 al ingresarlos en el modelo QSAR de terpenos.

Ecuación para calcular el parámetro de Fisher-Snedecor:

$$F = \frac{(R^2)(GL)}{(1 - R^2)(m)} \quad GL = (n - 1) - m$$

Donde: R^2 : Varianza explicada, $1 - R^2$: Varianza no explicada, GL : Grados de libertad, $n - 1$: Número de datos libres de variar (12 valores de actividad larvicida) y m : Número de parámetros o variables estimadas (la cantidad de descriptores).

Descriptores moleculares	R^2	EER	Distribución de Fisher-Snedecor (F)	GL	Significancia estadística (p)
GATS2m	0,7374	0,1546	22,46	8	$1,465 \times 10^{-3}$
MLFER_A	0,8642	0,1189	22,27	7	$9,227 \times 10^{-4}$
E1e	0,9094	0,1049	20,08	6	$1,569 \times 10^{-3}$
RDF140v	0,9681	0,06815	37,96	5	$6,208 \times 10^{-4}$
TDB2u	0,9998	0,00576	4.390	4	$1,452 \times 10^{-7}$
nHdsCH	1,0	$8,622 \times 10^{-5}$	$1,633 \times 10^7$	3	$2,344 \times 10^{-11}$
TDB8u	1,0	$2,777 \times 10^{-6}$	$1,349 \times 10^{10}$	2	$7,410 \times 10^{-11}$
SC-4	1,0	$2,008 \times 10^{-9}$	$2,258 \times 10^{16}$	1	$5,146 \times 10^{-9}$

*Los valores del parámetro F calculados son promedios entre los F individuales de cada descriptor molecular a partir del segundo.

Anexo 11. ANOVA de los descriptores teóricos mecanocuánticos y termoquímicos incluidos al análisis QSAR de los terpenos.

Descriptores teóricos	R^2	EER	Fisher - Snedecor (F)	Significancia estadística (p)
Energía E. Total	0,0270	0,2985	0,1693	$6,91 \times 10^{-1}$
Energía E_{HOMO}	0,1334	0,2809	1,2316	$2,99 \times 10^{-1}$
Energía E_{LUMO}	0,0089	0,3004	0,0720	$7,95 \times 10^{-1}$
Diferencia E. Gap	0,0493	0,2942	0,4149	$5,37 \times 10^{-1}$
Momento dipolar	0,0803	0,2894	0,6985	$4,28 \times 10^{-1}$
Entropía	0,0360	0,2962	0,2991	$5,99 \times 10^{-1}$
Energía L. Gibbs	0,0494	0,2942	0,4159	$5,37 \times 10^{-1}$

Anexo 12. Matriz de los descriptores moleculares del modelo QSAR de terpenos para actividad larvicida.

Compuesto químico	GATS2m	MLFER_A	E1e	RDF140v	TDB2u
Terpeno 1	0,8258	1,134	0,5386	7,2312	2,2031
Terpeno 2	0,8556	0,546	0,6182	2,2853	2,1948
Terpeno 3	0,7857	0,546	0,5508	5,1522	2,2008
Terpeno 4	0,8691	1,515	0,6405	0,6040	2,1900
Terpeno 5	0,8976	0,927	0,6383	4,2422	2,1818
Terpeno 6*	0,8245	0,927	0,6250	3,7245	2,1789
Terpeno 7	0,8488	0,591	0,5509	2,8852	2,1964
Terpeno 8*	0,8416	0,604	0,4980	7,0524	2,3528
Terpeno 9	0,8836	0,259	0,5077	2,8850	2,2245
Terpeno 10	0,83427	0,591	0,3939	$8,04 \times 10^{-57}$	2,2307

Terpeno 11	0,8689	0,003	0,4011	8,11x10 ⁻⁶⁹	2,2192
Terpeno 12	0,8032	0,546	0,5299	0,00	2,1593

*Compuestos de prueba.

Anexo 13. Resultados de actividad larvica predicha de la forma Log₁₀(DL₅₀) para los compuestos a partir del modelo QSAR de terpenos.

Compuesto químico	DL ₅₀ (mg/g de larva)	Log ₁₀ (DL ₅₀) experimental	Log ₁₀ (DL ₅₀) predicha	Residuales (RES)
Terpeno 1	2,94	0,468	0,4678	0,0001
Terpeno 2	15,52	1,191	1,1920	-0,0010
Terpeno 3	3,89	0,59	0,58901	0,0009
Terpeno 4	10,17	1,007	1,0066	0,0003
Terpeno 5	14,89	1,173	1,1718	0,0011
Terpeno 7	9,23	0,965	0,9739	-0,0089
Terpeno 9	19,8	1,297	1,2911	0,0058
Terpeno 10	6,59	0,819	0,8184	0,0005
Terpeno 11	15,05	1,178	1,1802	-0,0022
Terpeno 12	4,91	0,691	0,6878	0,0031

Anexo 14. ANOVA de los descriptores moleculares elegidos por el submétodo *Forward* en el intervalo 0,00 ≤ α ≤ 0,10 al ingresarlos en el modelo QSAR de flavonoides.

Ecuación para calcular el parámetro de Fisher-Snedecor:

$$F = \frac{(R^2)(GL)}{(1 - R^2)(m)} \quad GL = (n - 1) - m$$

Donde: R²: Varianza explicada, 1 - R²: Varianza no explicada, GL: Grados de libertad, n - 1: Número de datos libres de variar (12 valores de actividad larvica) y m: Número de parámetros o variables estimadas (la cantidad de descriptores).

Descriptores moleculares	R ²	EER	Distribución de Fisher-Snedecor (F)	GL	Significancia estadística (p)
ALogP	0,9982	0,03369	1.670	3	3,224x10 ⁻⁵
RotBFrac	0,9983	0,03984	597,2	2	1,672x10 ⁻³

*Los valores del parámetro F calculados son promedios entre los F individuales de cada descriptor molecular a partir del segundo.

Anexo 15. ANOVA de los descriptores teóricos mecanocuánticos y termoquímicos incluidos al análisis QSAR de los flavonoides.

Descriptores teóricos	R ²	EER	Fisher - Snedecor (F)	Significancia estadística (p)
Energía E. Total	0,3808	4,17x10 ⁻¹⁵	1,8449	0,2675
Energía E _{HOMO}	0,1991	0,00	0,7460	0,4517
Energía E _{LUMO}	0,1137	8,88x10 ⁻¹⁶	0,3849	0,5788
Diferencia E. Gap	0,0141	2,17x10 ⁻¹⁵	0,0429	0,8491

Momento dipolar	0,0977	$3,11 \times 10^{-16}$	0,3249	0,6085
Entropía	0,1568	$4,88 \times 10^{-16}$	0,5582	0,5092
Energía L. Gibbs	0,7484	$-1,77 \times 10^{-16}$	8,9251	0,0582

Anexo 16. Matriz de los descriptores moleculares del modelo QSAR de flavonoides para actividad larvícida.

Compuesto químico	ATSC3s
Flavonoide 1	13,3333
Flavonoide 2	-23,0902
Flavonoide 3	0,9899
Flavonoide 4*	-9,4040
Flavonoide 5	-20,1273
Flavonoide 6	-23,6648

*Compuesto de prueba.

Anexo 17. Resultados de actividad larvícida predicha de la forma $\text{Log}_{10}(\text{CL}_{50})$ para los compuestos a partir del modelo QSAR de flavonoides.

Compuesto químico	CL ₅₀ (ppm)	Log ₁₀ (CL ₅₀) experimental	Log ₁₀ (CL ₅₀) predicha	Residuales (RES)
Flavonoide 1	3,9	0,591	0,6032	-0,0121
Flavonoide 2	126,9	2,103	2,0987	0,0046
Flavonoide 3	13,63	1,134	1,1100	0,0244
Flavonoide 5	85,77	1,933	1,9771	-0,0437
Flavonoide 6	141	2,149	2,1223	0,0268

Anexo 18. Resultados de la actividad larvícida predicha de la forma $\text{Log}_{10}(\text{DL}_{50})$ para los compuestos en la CV-LOO a partir del modelo QSAR de eugenoles.

Compuesto químico	Log ₁₀ (DL ₅₀) experimental	Log ₁₀ (DL ₅₀) CV-LOO	Residuales CV-LOO
Eugenol 1	-0,309	-0,3429	0,0329
Eugenol 2	0,305	0,3141	-0,00913
Eugenol 3	0,285	0,3822	-0,0962
Eugenol 4	-0,055	-0,0813	0,0253
Eugenol 5	-0,045	-0,0269	-0,0191
Eugenol 7	0,307	0,3374	-0,0304
Eugenol 8	-0,337	-0,3705	0,0335
Eugenol 9	0,307	0,2456	0,0614
Eugenol 10	0,628	0,5812	0,0468
Eugenol 12	0,305	0,2426	0,0624
Eugenol 13	0,305	0,3113	-0,0063
Eugenol 14	0,627	0,5928	0,0342
Eugenol 15	0,451	0,4188	0,0332

Anexo 19. Resultados de la actividad larvícida predicha de la forma $\text{Log}_{10}(\text{DL}_{50})$ para los compuestos en la CV-LOO a partir del modelo QSAR de amidas sintéticas.

Compuesto químico	Log ₁₀ (DL ₅₀) experimental	Log ₁₀ (DL ₅₀) CV-LOO	Residuales CV-LOO
Amida 1	-0,101	0,0026	-0,1036
Amida 2	0,911	0,7430	0,1680
Amida 3	0,093	0,505	-0,4120
Amida 5	0,130	-0,0266	0,1566
Amida 6	0,420	0,4778	-0,0578
Amida 7	2,590	2,082	0,5080
Amida 9	1,150	1,150	0,0000
Amida 10	0,029	-0,523	0,552
Amida 11	1,235	1,001	0,234

*Los valores del parámetro F calculados son promedios entre los F individuales de cada descriptor molecular a partir del segundo.

Anexo 20. Resultados de la actividad larvicida predicha de la forma Log₁₀(DL₅₀) para los compuestos en la CV-LOO a partir del modelo QSAR de terpenos.

Compuesto químico	Log ₁₀ (DL ₅₀) experimental	Log ₁₀ (DL ₅₀) CV-LOO	Residuales CV-LOO
Terpeno 1	0,468	0,4369	0,0311
Terpeno 2	1,191	1,2105	-0,0195
Terpeno 3	0,59	0,6194	-0,0294
Terpeno 4	1,007	1,0055	0,0014
Terpeno 5	1,173	1,1351	0,0379
Terpeno 7	0,965	0,9727	-0,0076
Terpeno 9	1,297	1,2659	0,0311
Terpeno 10	0,819	0,8018	0,0172
Terpeno 11	1,178	1,1487	0,0293
Terpeno 12	0,691	0,6738	0,0172

Anexo 21. Resultados de la actividad larvicida predicha de la forma Log₁₀(CL₅₀) para los compuestos en la CV-LOO a partir del modelo QSAR de flavonoides.

Compuesto químico	Log ₁₀ (CL ₅₀) experimental	Log ₁₀ (CL ₅₀) CV-LOO	Residuales CV-LOO
Flavonoide 1	0,591	0,6316	-0,0405
Flavonoide 2	2,103	2,0924	0,0111
Flavonoide 3	1,134	1,0980	0,0370
Flavonoide 5	1,933	1,9941	-0,0608
Flavonoide 6	2,149	2,1077	0,0415

Anexo 22. Resultados de actividad larvicida predicha de la forma Log₁₀(DL₅₀)/Log₁₀(CL₅₀) para los compuestos de prueba a partir de los modelos QSAR para eugenoles, amidas, terpenos y flavonoides.

Compuesto de prueba	Log ₁₀ (DL ₅₀)/Log ₁₀ (CL ₅₀) experimental	Log ₁₀ (DL ₅₀)/Log ₁₀ (CL ₅₀) predicha	Residuales CV-Ext.
Eugenol 6	0,305	0,362	-0,057
Eugenol 11	-0,040	-0,672	0,632
Amida 4	-0,215	-0,105	-0,110

Amida 8	2,702	2,628	0,074
Terpeno 6	0,684	0,785	-0,101
Terpeno 8	1,093	1,271	0,178
Flavonoide 4	1,459	1,535	-0,076

Anexo 23. Matriz de los descriptores moleculares para los compuestos propuestos (modelo QSAR de eugenoles).

Compuesto químico	nssO	MATS3m	Entropía	GATS3v	ATSC6e	AATSC8e
Combinación A (1)	2	0,0038	0,0702	1,0854	-0,8436	-0,0112
Combinación A (2)	1	0,0872	0,0521	0,9709	-1,1772	-0,0010
Combinación B (1)	2	-0,1005	0,0563	1,2585	-0,9737	0,0110
Combinación B (2)	1	-0,0961	0,0563	1,2388	-0,3186	-0,0148
Combinación C (1)	3	0,0195	0,0704	1,0067	-0,1455	-0,0144
Combinación C (2)	2	0,0432	0,0693	1,1157	0,6128	0,0132

Anexo 24. Matriz de los descriptores moleculares para los compuestos propuestos (modelo QSAR de amidas).

Compuesto químico	ALogP
Combinación A (1)	1,5036
Combinación A (2)	-0,0547
Combinación D (1)	3,8388
Combinación D (2)	3,2565
Combinación E (1)	-2,2506
Combinación E (2)	-2,8405

Anexo 25. Matriz de los descriptores moleculares para los compuestos propuestos (modelo QSAR de terpenos).

Compuesto químico	GATS2m	MLFER_A	E1e	RDF140v	TDB2u
Combinación B (1)	0,887608	0,927	0,581	0,0000	2,2029
Combinación B (2)	0,908033	0,883	0,546	3,3858	2,1785
Combinación D (1)	0,920922	0,371	0,514	6,1401	2,1759
Combinación D (2)	0,995677	0,591	0,522	5,9317	2,1712
Combinación F (1)	0,888334	1,219	0,568	2,5489	2,1795
Combinación F (2)	0,771617	0,128	0,572	6,2731	2,2041

Anexo 26. Matriz de los descriptores moleculares para los compuestos propuestos (modelo QSAR de flavonoides).

Compuesto químico	ATSC3s
Combinación C (1)	0,5212
Combinación C (2)	-14,3280
Combinación E (1)	-12,8410
Combinación E (2)	21,1814
Combinación F (1)	-7,9232

Combinación F (2)	-29,4032
-------------------	----------

Anexo 27. ANOVA individual para la selección de los dos mejores descriptores moleculares del modelo QSAR para derivados del eugenol, amidas, terpenos y flavonoides.

Descriptores moleculares	Fisher - Snedecor (F)	Significancia estadística (p)	Fisher calculado (F value)	Probabilidad Fisher (Pr (>F))
ATSC6e (E)	82,4	4,587x10 ⁻⁶	253,99	3,87x10 ⁻⁶
AATSC8e (E)	463,6	9,936x10 ⁻⁸	40,58	7,03x10 ⁻⁴
ALogP (A)	120,8	1,143x10 ⁻⁵	120,8	1,14x10 ⁻⁵
RDF140v (T)	37,96	0,0006208	1.288,4	3,60x10 ⁻⁶
TDB2u (T)	4390	1,452x10 ⁻⁷	695,9	1,23x10 ⁻⁵
ATSC3s (F)	1670	3,224x10 ⁻⁵	1,670	3,22x10 ⁻⁵

E: Derivados del eugenol; A: Amidas sintéticas; T: Terpenos naturales; F: Flavonoides.