

PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR

FACULTAD DE HABITAT, INFRAESTRUCTURA Y CREATIVIDAD

MAESTRÍA EN BIOLOGÍA COMPUTACIONAL

TEMA

"Relaciones cuantitativas estructura-actividad (QSAR) basadas en clasificadores de aprendizaje automático para la predicción de moléculas inhibidoras de tirosinasas"

RESPONSABLE:

DOMÉNICA MUÑOZ VAZQUEZ

DIRECTOR:

JUAN CAMILO ANZIETA REYES

SEPTIEMBRE 2025

Tabla de contenido

Resumen.....	10
Abstract.....	11
1. Introducción.....	12
2. Objetivos	13
2.1 Objetivo General.....	13
2.2 Objetivos Específicos.....	13
3. Revisión de la literatura	14
3.1. La tirosinasa como blanco enzimático: estructura, función y relevancia clínica.....	14
3.1.1. Características bioquímicas de la tirosinasa	14
3.1.2. Implicaciones fisiológicas y patológicas de la tirosinasa	17
3.1.3. Estrategias de inhibición de la tirosinasa	21
3.1.4. Enfoques computacionales para el descubrimiento de nuevos inhibidores.....	24
3.2. Fundamentos de los modelos QSAR.....	27
3.2.1. Aplicaciones de los modelos QSAR en la predicción de inhibidores.....	29
3.2.2. Tipos de modelos QSAR.....	30
3.2.3. Estructura general de un modelo QSAR	31
3.2.4. Ventajas y desafíos del modelado QSAR	32
3.3. Aprendizaje automático en QSAR	33
3.3.1. Ventajas del aprendizaje automático en modelos QSAR	33

3.3.2. Algoritmos comunes de aprendizaje automático utilizados en QSAR clasificadorio	34
3.3.3. Proceso general de construcción de un modelo QSAR con aprendizaje automático	37
3.3.4. Importancia del aprendizaje automático en la predicción de inhibidores de tirosinasa	38
3.4. Descriptores moleculares	39
3.4.1 Definición y requisitos de los descriptores moleculares	39
3.4.2 Representación molecular y curado de datos	40
3.4.3 Clasificación de descriptores moleculares	41
3.4.4 Herramientas para el cálculo de descriptores	43
3.4.5 Selección y reducción de descriptores	43
3.4.6 Importancia de los descriptores en QSAR para inhibidores de tirosinasa.....	44
3.4.7 Selección y reducción de descriptores	45
3.4.8 Relevancia de los descriptores para inhibidores de tirosinasa	45
3.5. Validación de modelos QSAR.....	46
3.5.1. Principios fundamentales de un modelo QSAR válido (OECD)	47
3.5.2. Métodos de validación interna.....	47
3.5.3. Validación externa.....	48
3.5.4. Dominio de aplicabilidad (Applicability Domain, AD)	49
3.5.5. Evaluación y comparación de modelos	49

3.5.6. Consideraciones éticas y reproducibilidad	50
3.6 Métodos Computacionales aplicados a la Quimiometría	51
3.6.1 Métodos no supervisados	51
3.6.2 Métodos supervisados	52
3.6.3 Técnicas de selección de variables.....	53
3.6.4 Evaluación de los modelos.....	54
4. Metodología.....	55
4.1. Enfoque de investigación.....	55
4.2. Diseño metodológico	55
4.3 Equipos y programas utilizados	56
4.2.1 Equipos	56
4.2.2 Programas.....	56
4.4 Métodos y técnicas de análisis	57
4.4.1 Generación de la base de datos	57
4.4.2 Representación de la estructura molecular	57
4.4.3 Curado de la base de datos	57
4.4.4 Cálculo de descriptores moleculares.....	58
4.4.5 Métodos de modelamiento	58
4.4.6 Validación del modelo	59
4.4.7 Dominio de aplicabilidad	60
4.4.8 Aplicación práctica del modelo.....	60
5. Resultados	62

5.1. Generación de la base de datos	62
5.2 Representación de la estructura molecular.....	63
5.3 Curado de la base de datos.....	64
5.4 Cálculo de descriptores moleculares	66
5.5 Reducción no supervisada de descriptores moleculares.....	66
5.6 Selección supervisada de descriptores moleculares	67
5.6.1 Métodos de regresión.....	67
5.6.2 Métodos de clasificación	69
5.7 Validación del Modelo.....	77
5.8 Dominio de aplicabilidad del modelo	78
5.9 Predicción de moléculas	79
6. Análisis y discusión de resultados.....	81
6.1 Calidad predictiva de los modelos QSAR	81
6.2 Evaluación de descriptores moleculares y representación química.....	82
6.3 Validación y dominio de aplicabilidad	83
6.4 Predicción de nuevas moléculas	83
6.5 Ventajas del enfoque de aprendizaje automático frente al QSAR clásico	84
6.6 Aplicabilidad en diseño racional de inhibidores	84
6.7 Relevancia farmacológica de los compuestos predichos.....	85
6.8 Limitaciones y perspectivas futuras	85
7. Conclusiones y Recomendaciones	86

7.1 Conclusiones	86
7.2 Recomendaciones	87
Referencias	90
Anexos	100

Índice de Figuras

Figura 1. Biogénesis y maduración de melanosomas	17
Figura 2. Visión integral de la síntesis, regulación y aplicaciones de la melanina bacteriana.....	18
Figura 3. Anatomía de la piel.....	19
Figura 4. Los modelos QSAR para la permeabilidad de la BHE.....	28
Figura 5. Proceso de modelado QSAR	29
Figura 6. Representación de la estructura química del Acido Kojico y su optimización por mecánica molecular	64
Figura 7. Descriptores seleccionados (7) para el modelo clasificatorio	73

Índice de Tablas

Tabla 1. Tipos de tirosina y sus principales funciones bioquímicas	16
Tabla 2. Principales tipos de inhibidores de tirosinasa y ejemplos representativos	23
Tabla 3. Tipos de modelos QSAR.....	30
Tabla 4. Etapas clave para el desarrollo de un modelo QSAR confiable	32
Tabla 5. Comparación general de algoritmos de aprendizaje automático para QSAR	37

Tabla 6. Descriptores unidimensionales (1D)	41
Tabla 7. Descriptores bidimensionales (2D)	41
Tabla 8. Descriptores tridimensionales (3D)	42
Tabla 9. Descriptores fisicoquímicos	42
Tabla 10. Herramientas para el cálculo de descriptores	43
Tabla 11. Tipos de validación cruzada (cross-validation)	48
Tabla 12. Métricas estadísticas de desempeño	48
Tabla 13. Comparación de desempeño entre algoritmos QSAR	50
Tabla 14. Métodos no supervisados comunes en QSAR.....	52
Tabla 15. Moléculas utilizadas para la predicción de actividad inhibitoria	60
Tabla 16. Valores atípicos eliminados mediante el test de Dixon para el ácido kójico	65
Tabla 17. Valores atípicos eliminados mediante el test de Dixon para la arbutina	65
Tabla 18. Resultados de los modelos de regresión QSAR para inhibidores de la tirosinasa.....	68
Tabla 19. Parámetros de calidad de los modelos QSAR basado en clasificación PLSDA, Knn, RF, AdaBoost con todos los descriptores y con descriptores reducidos.....	71
Tabla 20. Descriptores seleccionados (7) para el modelo clasificadorio.....	73
Tabla 21. Desempeño comparativo por huellas moleculares (MACCS, ECFP, ECFP_V3, PFP) y clasificadores (kNN, N3, BNN).	76
Tabla 22. Clase predicha para el conjunto externo de ITs.....	80

Resumen

En la Pontificia Universidad Católica del Ecuador se implementó un flujo de trabajo *in silico* orientado a la priorización de inhibidores de tirosinasas, integrando QSAR con algoritmos de aprendizaje automático. Tras la curaduría y estandarización de un repositorio de 548 estructuras, se generó una matriz de aproximadamente 4 mil descriptores moleculares mediante alvaDesc. La redundancia y colinealidad se mitigaron con V-WSP y, a partir de los descriptores seleccionados junto con huellas estructurales (MACCS, ECFP y PFP), se ajustaron clasificadores no lineales (Random Forest, AdaBoost, kNN y N3). La capacidad predictiva se estimó con validación cruzada estratificada y mediante un análisis explícito del dominio de aplicabilidad. La configuración ECFP + N3 ofreció el mejor equilibrio de desempeño, con exactitud del 86,3 % en el conjunto externo de prueba. Como demostración de utilidad, se priorizaron 17 moléculas que la literatura indicaba que tenían capacidad inhibitoria pero no presentaban un valor; entre ellas, glabridina y 4-propilresorcinol quedaron clasificadas como activas dentro del dominio. Estos resultados muestran que la codificación estructural acoplada a aprendizaje automático permite construir filtros de cribado virtual robustos y reproducibles para la búsqueda de inhibidores de tirosinasa, con proyección a aplicaciones farmacéuticas y dermocosméticas, pues esta es una enzima clave en la ruta metabólica de la melanina.

Palabras clave: QSAR- Tirosinasa- Aprendizaje automático- Huellas moleculares- Cribado virtual.

Abstract

This study was conducted within the Master's in Computational Biology at the Pontifical Catholic University of Ecuador (PUCE), with the aim of applying molecular modeling and machine-learning approaches to establish quantitative structure–activity relationships (QSAR) to guide the selection of tyrosinase inhibitors—an enzyme central to biological processes of medical and biotechnological interest.

A database of 548 compounds was curated and more than 1,182 molecular descriptors were computed using alvaDesc. Subsequently, variable reduction was performed with the V-WSP method, and classification models were built based on compound activity. Machine-learning algorithms such as Random Forest, AdaBoost, kNN, N3, and BNN were implemented using selected descriptors and molecular fingerprints (MACCS, ECFPs, PFPs).

Validation was carried out through cross-validation techniques and applicability-domain analysis. The best performance was achieved by the ECFPs + N3 model, with an accuracy of 86.3% on the test set. In addition, the class of 17 previously unevaluated molecules was predicted, identifying Glabridin and 4-Propylresorcinol as active candidates. In conclusion, integrating QSAR and computational biology constitutes an efficient strategy for the virtual screening of tyrosinase inhibitors, with potential application in the pharmaceutical and cosmetics industries, since this is a key enzyme in the melanin metabolic pathwa.

Keywords: QSAR- tyrosinase- machine learning- molecular fingerprints- virtual screening.

1. Introducción

El interpretar, descubrir y diseñar compuestos bioactivos nuevos ha sido esencial para el progreso de investigaciones en diversas áreas como la alimentaria o la farmacéutica. Para Vargas et al., 2023 se resalta, a los inhibidores enzimáticos que han potenciado el tratamiento a varias patologías peligrosas en humanos, así como en sistemas biológicos en ámbitos industriales.

La exploración de inhibidores enzimáticos de la tirosinasa ha sido clave en estudios de aplicación terapéutica (para tratamiento del cáncer de piel), cosmética (para blanquear o despigmentar), e industrial (para prevenir el pardeamiento de alimentos o para proteger cultivos) (Urrialde et al., 2022). Es por ello que es necesario proponer/crear/buscar nuevos métodos de síntesis y valoración experimental que reduzcan el tiempo y los altos costos de aquellos de síntesis tradicionales.

Surgen entonces los métodos computacionales basados en el modelo QSAR (siglas en inglés para “relación cuantitativa estructura-actividad”) como medida eficiente para anticipar la actividad biológica de moléculas nuevas, diferenciar entre sustancias activas e inactivas y mejorar así la selección de moléculas para pruebas experimentales con un mayor nivel de exactitud en la elección de candidatos potenciales, todo a menor costo y tiempo (Ponphaiboon et al., 2023). El presente estudio pretende analizar las relaciones cuantitativas basadas en clasificadores de aprendizaje automático enfocadas en la predicción de moléculas inhibidoras de toxinas, en particular de la tirosinasa.

2. Objetivos

2.1 Objetivo General

- Emplear métodos quimiométricos para establecer la relación cuantitativa estructura – actividad (QSAR), como estrategia para la elección de inhibidores potenciales de tirosinasas de un conjunto de moléculas.

2.2 Objetivos Específicos

- Actualizar la base de datos inicial para el desarrollo de inhibidores de la enzima tirosinasa, asociándolos a descriptores moleculares.
- Utilizar técnicas estadísticas conexas con descriptores moleculares, para la generación de modelos matemáticos con el fin de clasificar inhibidores enzimáticos.
- Ejecutar los modelos matemáticos obtenidos para evaluar la potencia de los compuestos.
- Investigar potenciales métodos cuantitativos alternativos para la reducción de variables comúnmente utilizadas en los modelos QSAR.

3. Revisión de la literatura

3.1. La tirosinasa como blanco enzimático: estructura, función y relevancia clínica

3.1.1. Características bioquímicas de la tirosinasa

La tirosinasa (EC 1.14.18.1) es una enzima metaloenzimática clasificada dentro del grupo de las oxidorreductasas, dependiente de cobre, y ampliamente conservada en una variedad de organismos, incluyendo bacterias, hongos, plantas y animales (Benavides et al., 2020). En sistemas eucariotas, esta enzima se localiza predominantemente en los melanosomas, orgánulos especializados en la síntesis, maduración y almacenamiento de melanina (Fernandes et al., 2023).

En términos bioquímicos, tirosinasa actúa como una oxidasahidroxilasa dependiente de cobre que inicia la melanogénesis mediante dos pasos consecutivos: primero convierte L-tirosina en L-DOPA por hidroxilación (modo monofenolasa) y, después, oxida la L-DOPA hasta dopaquinona (modo o-difenolasa) (Sayed et al., 2021). A partir de este nodo se bifurcan las rutas que originan los dos pigmentos principales: eumelanina (negro-parda) y feomelanina (rojo-amarillenta), en función del microambiente químico y de la disponibilidad de nucleófilos azufrados (Sayed et al., 2021). La modulación de su actividad — expresión, cofactores y condiciones redox— resulta crítica para los patrones de pigmentación en humanos y también explica el pardeamiento enzimático observado en matrices vegetales (Caicedo, 2024).

El centro activo de la tirosinasa está constituido por dos átomos de cobre (CuA y CuB), cada uno coordinado por tres residuos de histidina. Esta configuración le permite catalizar reacciones redox dependientes de oxígeno molecular (O₂) y sustratos fenólicos (Zou et al., 2025). Esta estructura tridimensional, altamente conservada, constituye una diana terapéutica de interés en el diseño racional de inhibidores específicos, mediante técnicas computacionales como el modelado molecular, la dinámica molecular o la predicción QSAR (Azad et al., 2023) detallada en la sección (3.1.4).

La tirosina: aminoácido precursor de melanina

La tirosina es un aminoácido no esencial en humanos, ya que puede ser sintetizada a partir de fenilalanina mediante la acción de la enzima fenilalanina hidroxilasa. Desde una perspectiva bioquímica, la tirosina cumple múltiples funciones estructurales y metabólicas (Ivanov et al., 2024). Es precursora de importantes biomoléculas como las catecolaminas (dopamina, adrenalina, noradrenalina), hormonas tiroideas (T₃, T₄) y melanina.

En el contexto de la melanogénesis, la tirosina actúa como el primer sustrato de la tirosinasa, iniciando la vía que conduce a la producción de melanina. Esta reacción ocurre en el citosol de los melanocitos y es altamente regulada por factores genéticos, hormonales y ambientales, incluyendo la exposición a radiación ultravioleta (Maddaleno et al., 2021).

Aguilera (2021) menciona tipos de tirosinas que a continuación se presentan en una tabla con sus principales formas y funciones en organismos eucariotas:

Tabla 1. Tipos de tirosina y sus principales funciones bioquímicas

Tipo / Derivado	Función principal
L-tirosina (libre)	Sustrato de tirosinasa; precursor de L-DOPA en la melanogénesis
Fosfotirosina	Participa en señalización celular como residuo fosforilado en proteínas transmembrana
DOPA (3,4-dihidroxifenilalanina)	Intermediario en la síntesis de melanina y catecolaminas
Dopaquinona	Producto inmediato de la oxidación catalizada por tirosinasa; precursor de melanina
T3 y T4 (hormonas tiroideas)	Derivados yodados de tirosina implicados en el metabolismo energético
Catecolaminas (dopamina, adrenalina)	Neurotransmisores derivados de tirosina a través de rutas específicas

Nota. En la presente tabla se describe los tipos de tirosinas y sus funciones principales. Elaboración propia.

Esta amplia participación de la tirosina en procesos metabólicos clave, y en especial en la vía de la melanogénesis, resalta su importancia como punto de partida en estrategias terapéuticas dirigidas a modular la producción de melanina, mediante la inhibición selectiva de la tirosinasa (Wang et al., 2024). Comprender sus derivados y rutas metabólicas es esencial para el diseño de moléculas análogas, que puedan bloquear de forma específica el sitio activo de la enzima sin interferir con otras funciones fisiológicas (Niazi, 2025).

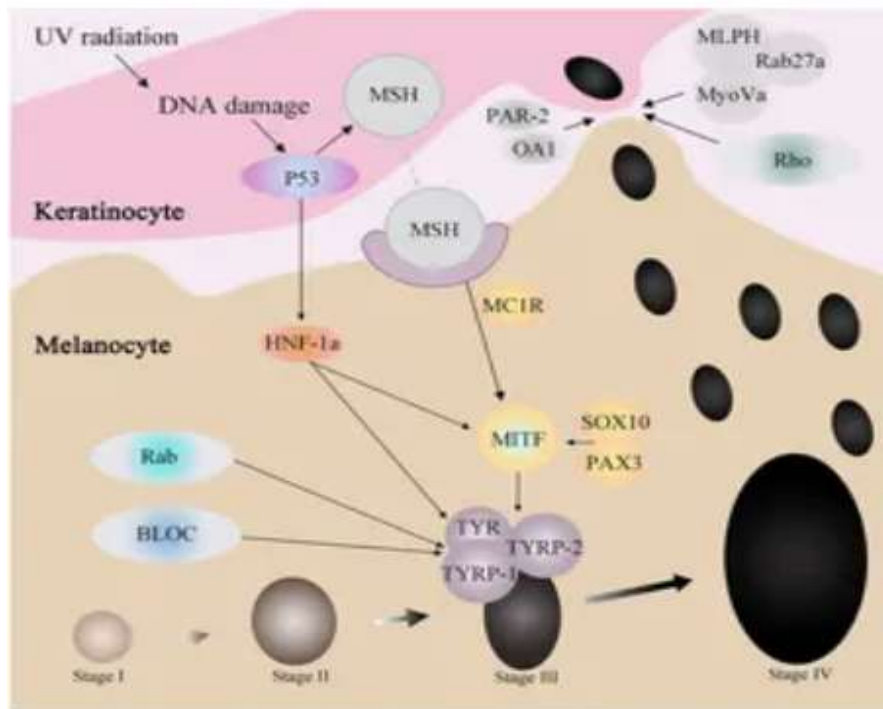


Figura 1. Biogénesis y maduración de melanosomas. Proceso molecular de melanogénesis, desde la señal UV inicial hasta la formación y transferencia de melanosomas pigmentados. Elaborado por: Jiyan (2024)

3.1.2. Implicaciones fisiológicas y patológicas de la tirosinasa

La tirosinasa cumple una función esencial en organismos superiores al participar en la síntesis de melanina, un biopolímero pigmentario que protege a las células frente a daños inducidos por radiación ultravioleta (UV), radicales libres y otros agentes oxidativos (Singh et al., 2025). En humanos, esta enzima está expresada principalmente en melanocitos, células localizadas en la epidermis, folículos pilosos, retina y sistema nervioso central. La actividad de esta enzima sostiene la biosíntesis de eumelanina y feomelanina, polímeros que configuran

los fenotipos cromáticos de piel, cabello y ojos (Coutant et al., 2024).

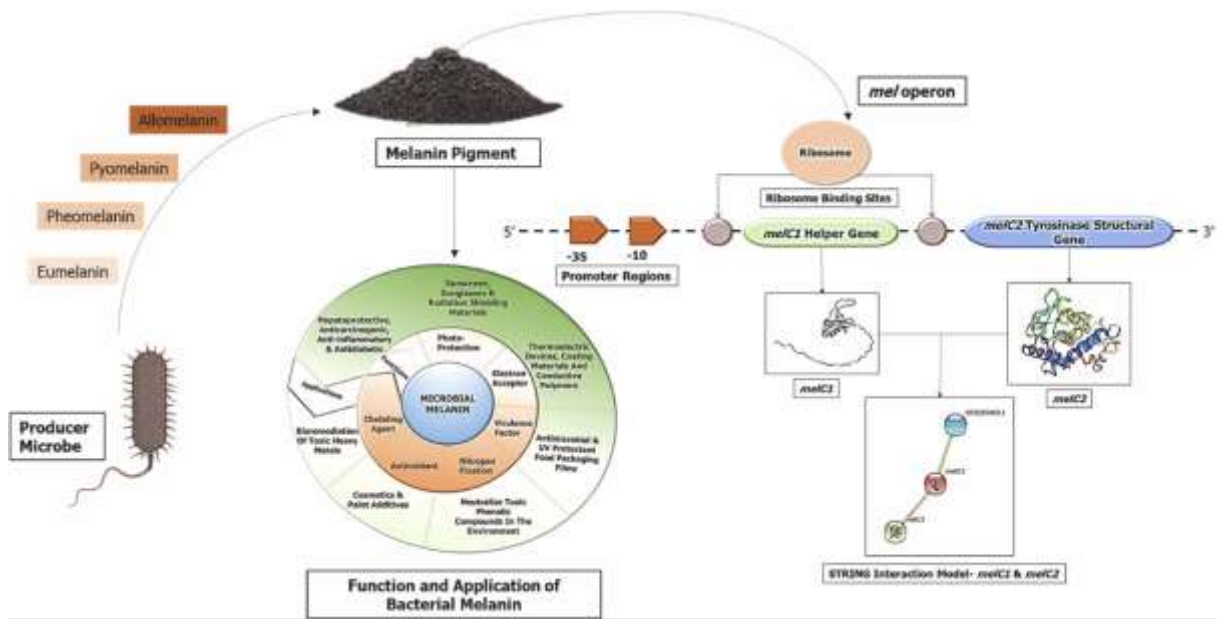


Figura 2. Visión integral de la síntesis, regulación y aplicaciones de la melanina bacteriana. Los microorganismos genéticamente codificados pueden producir melanina mediante rutas biosintéticas específicas, reguladas por genes como *hmgA*. Elaborado por: Singh & Chaturvedi (2025)

Desde un enfoque fisiológico, la melanina no solo contribuye al fenotipo pigmentario, sino que también ejerce funciones fotoprotectoras, al absorber radiación UV y disipar energía térmica (Morgado et al., 2023). Además, protege frente al estrés oxidativo inducido por luz solar o xenobióticos, estabiliza estructuras celulares y participa en la detoxificación de metales pesados (Caballero, 2024).

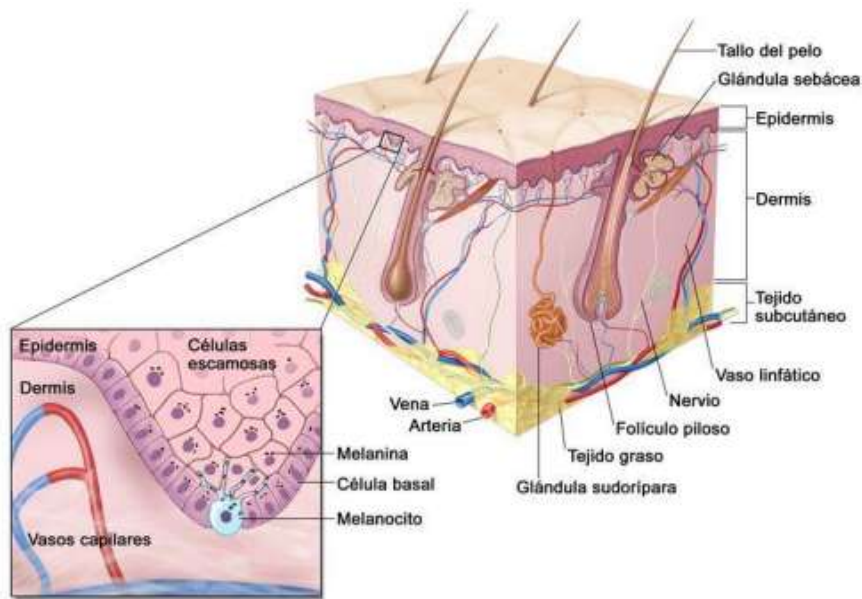


Figura 3. Anatomía de la piel. Los melanocitos están presentes en la capa de las células basales, en la parte más profunda de la epidermis. Elaborado por: Caballero (2024)

Sin embargo, cuando existe una disfunción en la regulación de la actividad tirosinásica, pueden generarse consecuencias clínicas importantes. Botella, Quintas, & Gonzales (2023) menciona que la hiperactividad de tirosinasa se ha relacionado con trastornos hiperpigmentarios, entre ellos:

- Melasma: hiperpigmentación facial común, especialmente en mujeres, vinculada a exposición solar, desequilibrios hormonales o alteración (?) en el uso/efecto(?) de anticonceptivos.
- Léntigos solares: lesiones pigmentadas inducidas por radiación UV crónica, frecuentes en personas mayores.
- Hiperpigmentación postinflamatoria: oscurecimiento de la piel posterior a lesiones, dermatitis o procedimientos cosméticos.

En un plano más grave, se ha descrito que la sobreexpresión de tirosinasa puede estar implicada en la progresión de melanomas, tumores malignos derivados de melanocitos (Burón, 2021). En este contexto, la tirosinasa puede ser utilizada

como biomarcador diagnóstico, pronóstico e incluso como blanco terapéutico en estrategias de inmunoterapia o terapia dirigida (Ugaldea et al., 2024).

Por el contrario, la hipoactividad o mutaciones en el gen TYR (que codifica la tirosinasa) son responsables de formas de albinismo oculocutáneo tipo 1 (OCA1), caracterizadas por la ausencia parcial o total de melanina, fotofobia, nistagmo (quizás explicar, o nota al pie) y mayor riesgo de cáncer cutáneo (Santana, 2022). Este déficit evidencia el papel indispensable que cumple la tirosinasa en la homeostasis pigmentaria humana (Rodrigues et al., 2022).

Aplicaciones industriales y agroalimentarias

Más allá del ámbito médico, la tirosinasa presenta una importancia notable en la industria alimentaria y agrícola. Su actividad en vegetales genera el fenómeno conocido como pardeamiento enzimático, un proceso de oxidación de compuestos fenólicos que lleva a la producción de melaninas vegetales (Botella et al., 2023). Esto impacta negativamente en la calidad visual, sabor, aroma y valor nutricional de productos como manzanas, papas, plátanos, champiñones, entre otros (Benzama, 2021).

Este fenómeno ocurre cuando las células vegetales se lesionan (por corte, impacto o envejecimiento) y se liberan tanto la tirosinasa como los sustratos fenólicos que ésta oxida (Romanos, 2024). La formación de pigmentos marrones reduce la aceptación comercial y puede inducir rechazo por parte del consumidor. Por ello, la tirosinasa es un blanco enzimático clave en la formulación de estrategias de conservación postcosecha, como la aplicación de inhibidores naturales, control de pH, atmósferas modificadas o tratamientos térmicos (Jiménez, 2023).

En la industria cosmética, por otro lado, la inhibición controlada de la tirosinasa es fundamental para formular agentes despigmentantes seguros y efectivos, capaces de reducir manchas y unificar el tono de piel. Ingredientes activos como ácido kójico, arbutina, niacinamida, ácido azelaico, y moléculas derivadas de plantas, deben ser evaluados rigurosamente por su eficacia, toxicidad y estabilidad, lo que ha incentivado el uso de herramientas computacionales para identificar nuevos inhibidores (Minerva et al., 2025).

3.1.3. Estrategias de inhibición de la tirosinasa

Dado el papel central que cumple la tirosinasa en la biosíntesis de melanina, su inhibición específica representa una estrategia de gran interés en diversas áreas, desde la dermatología y oncología, hasta la cosmética y la tecnología de alimentos. El diseño de inhibidores eficaces requiere una comprensión precisa del mecanismo de acción de la enzima, su estructura activa, la naturaleza de sus cofactores metálicos y los sitios susceptibles a bloqueo molecular (Aguilera, 2021).

Desde el punto de vista bioquímico, la inhibición de la tirosinasa puede clasificarse en diferentes tipos, dependiendo de cómo interactúe el inhibidor con la enzima y su sustrato. A continuación, se describen las principales modalidades de inhibición:

a) Inhibición competitiva

Ocurre cuando el inhibidor se une al mismo sitio activo que el sustrato natural, generalmente compitiendo con la L-tirosina o la L-DOPA. Este tipo de inhibidores tienden a tener estructuras químicas similares a los sustratos de la enzima, lo que les permite mimetizar su unión. Sin embargo, su eficacia puede disminuir en

presencia de altas concentraciones del sustrato, ya que la competencia puede ser desplazada.

Ejemplos representativos incluyen: ácido kójico, ácido ferúlico, ácido gálico, y derivados de catecol. Aunque han demostrado actividad inhibitoria in vitro, su aplicación clínica está limitada por problemas de fotoinestabilidad, toxicidad dérmica o baja biodisponibilidad (Reyes et al., 2023).

b) Inhibición no competitiva

Este tipo de inhibidores se une a un sitio distinto del centro catalítico (sitio alostérico), afectando la conformación tridimensional de la enzima e impidiendo la catálisis, independientemente de la concentración del sustrato. Esta modalidad ofrece la ventaja de mantener la eficacia inhibitoria incluso en presencia de elevadas concentraciones del sustrato natural.

Moléculas como el resveratrol, algunos peptidomiméticos y ciertos flavonoides han mostrado mecanismos no competitivos, convirtiéndose en candidatos atractivos para estudios farmacológicos y cosméticos (Zambrano y Rodríguez, 2022).

c) Inhibición mixta

Los inhibidores mixtos presentan características tanto competitivas como no competitivas, ya que pueden unirse a la enzima libre y al complejo enzima-sustrato. Esto les permite modificar tanto la constante de Michaelis (K_m) como la velocidad máxima (V_{max}), resultando en una inhibición más flexible y efectiva en contextos biológicos variables (Sanchez, 2024).

d) Inhibidores suicidas (irreversibles)

Son moléculas que, al unirse al sitio activo, sufren una transformación química que genera un complejo covalente con la enzima, inactivándola de manera permanente. Este mecanismo se considera de alta potencia, pero también conlleva mayores riesgos de toxicidad (Calderón et al., 2023).

Ejemplos como la hidroquinona y sus derivados han sido estudiados, aunque su uso ha sido restringido debido a reacciones adversas y regulación internacional.

Tabla 2. Principales tipos de inhibidores de tirosinasa y ejemplos representativos

Tipo de inhibición	Mecanismo de acción	Ejemplos de compuestos	Aplicación principal
Competitiva	Compite con el sustrato por el sitio activo	Ácido kójico, arbutina, ácido gálico	Cosmética, farmacéutica
No competitiva	Se une a sitios alostéricos y altera la conformación enzimática	Resveratrol, luteolina, baicaleína	Dermatología, oncología
Mixta	Interactúa con la enzima libre y el complejo enzima-sustrato	Derivados flavonoides modificados	Bioquímica aplicada, diseño racional
Irreversible (suicida)	Forma enlace covalente irreversible con el centro catalítico	Hidroquinona, aloesina, catequina	Industria, tratamiento agresivo puntual

Nota. En la presente tabla se observa los principales tipos de inhibidores de tirosinasa.

Elaboración propia

Limitaciones de los inhibidores clásicos

Jiménez (2023) alude que, aunque diversos compuestos naturales y sintéticos han demostrado potencial inhibitorio sobre tirosinasa, su aplicación práctica presenta limitaciones relevantes:

- Estabilidad química reducida, especialmente frente a luz UV, oxidación o cambios de pH.
- Baja biodisponibilidad, que impide alcanzar concentraciones efectivas en tejidos diana.
- Toxicidad dérmica o sistémica, especialmente en tratamientos prolongados o de uso tópico frecuente.
- Falta de selectividad, que puede inducir efectos secundarios al inhibir enzimas relacionadas estructuralmente.

Estas restricciones han impulsado el uso de herramientas computacionales para diseñar y predecir inhibidores más seguros, eficientes y selectivos, con el fin de superar las limitaciones asociadas a los métodos tradicionales de descubrimiento de fármacos.

3.1.4. Enfoques computacionales para el descubrimiento de nuevos inhibidores

El diseño de inhibidores de tirosinasa ha evolucionado significativamente en las últimas décadas gracias a la integración de metodologías computacionales en el proceso de descubrimiento de fármacos (Bencomo et al., 2022). Estas herramientas permiten analizar grandes volúmenes de datos estructurales, evaluar la interacción enzima-ligando, y predecir la actividad biológica de compuestos sin necesidad de recurrir a ensayos experimentales costosos, largos o invasivos. Entre los enfoques más utilizados destacan la modelación molecular, el acoplamiento molecular (docking), la simulación dinámica, y

especialmente, las relaciones cuantitativas estructura-actividad (QSAR), combinadas con algoritmos de aprendizaje automático (Goode et al., 2019).

a) Modelado molecular y docking estructural

El modelado tridimensional de la estructura de la tirosinasa, a partir de datos cristalográficos o predicciones de novo, permite explorar su sitio activo, identificar regiones hidrofóbicas, puentes de hidrógeno y zonas catalíticas sensibles a la unión de inhibidores (Sanchez, 2024). El acoplamiento molecular (molecular docking) consiste en predecir la orientación, afinidad y energía de interacción entre la enzima y una molécula candidata. Esta técnica es esencial para filtrar librerías virtuales de compuestos, priorizando aquellos con mayor compatibilidad estérica y energética con el sitio activo de la enzima (Hernández, 2023).

Programas como AutoDock, MOE, PyMOL o GOLD han sido ampliamente empleados en el cribado virtual de inhibidores de tirosinasa, permitiendo incluso evaluar la selectividad frente a otras oxidasas similares.

b) Relaciones cuantitativas estructura-actividad (QSAR)

Los modelos QSAR constituyen una herramienta poderosa para correlacionar descriptores moleculares (hidrofobicidad, carga, volumen, polaridad, entre otros) con la actividad biológica observada frente a un enzima blanco, en este caso, la tirosinasa (Hernández, 2023). A través de métodos estadísticos y matemáticos, es posible construir funciones predictivas que discriminan entre compuestos activos e inactivos, lo que facilita la preselección de nuevos candidatos con alta probabilidad de éxito experimental (Goya y Ortiz, 2020).

Una de las ventajas de los modelos QSAR es su capacidad para ser entrenados y validados con bases de datos públicas (ChEMBL, PubChem, BindingDB), utilizando descriptores obtenidos por herramientas como PaDEL-Descriptor, RDKit o Dragon.

c) Aprendizaje automático aplicado a QSAR

La incorporación de algoritmos de aprendizaje automático ha potenciado la eficiencia y precisión de los modelos QSAR, especialmente cuando se trabaja con grandes volúmenes de datos y estructuras químicas diversas (Fuadah et al., 2024). Entre los clasificadores más utilizados se encuentran:

- Random Forest (RF): combina múltiples árboles de decisión para reducir el error y evitar el sobreajuste.
- k-Nearest Neighbors (k-NN): clasifica compuestos según su similitud con vecinos conocidos.
- Bayesian Neural Networks (BNN): emplean un enfoque probabilístico que incorpora distribuciones de probabilidad sobre los pesos de la red, proporcionando estimaciones de incertidumbre más confiables en la predicción.
- N3 (Neighborhood Component Analysis): método basado en la optimización de métricas de vecindad para mejorar la clasificación y reducir la dimensionalidad de los datos, facilitando una representación más precisa de las relaciones entre compuestos químicos.

Estos modelos pueden ser entrenados para maximizar métricas como la precisión, sensibilidad, especificidad o el área bajo la curva ROC (AUC), asegurando su capacidad de generalización.

Ventajas del diseño in silico de inhibidores

Hernández (2023) sostiene que el uso de herramientas computacionales para la predicción de inhibidores de tirosinasa aporta beneficios concretos:

- Reducción significativa del número de compuestos a evaluar experimentalmente.
- Priorización racional basada en criterios de eficacia y seguridad.
- Posibilidad de explorar espacios químicos no sintetizados previamente.
- Facilidad para generar bibliotecas virtuales y simular modificaciones estructurales.

Estas ventajas han consolidado al enfoque in silico como una fase esencial en los programas modernos de descubrimiento y optimización de fármacos (Mohaya et al., 2024).

3.2. Fundamentos de los modelos QSAR

Los modelos de Relaciones Cuantitativas Estructura-Actividad (QSAR, por sus siglas en inglés Quantitative Structure-Activity Relationship) constituyen una metodología computacional orientada a predecir la actividad biológica o las propiedades fisicoquímicas de compuestos a partir de sus características estructurales (Faramarzi, 2022). Este enfoque se fundamenta en el principio de que moléculas con estructuras similares tienden a exhibir comportamientos biológicos semejantes, lo que posibilita la construcción de modelos matemáticos capaces de inferir la bioactividad de nuevas moléculas sin necesidad de recurrir a pruebas experimentales iniciales (Arbitto y Guamán, 2024).

En la presente investigación, el programa AlvaDesc se empleó como herramienta fundamental para el cálculo y análisis de descriptores moleculares. Esta plataforma permitió generar de manera eficiente un amplio conjunto de variables estructurales y fisicoquímicas a partir de las representaciones SMILES (Simplified Molecular Input Line Entry Specification), de las moléculas evaluadas como posibles inhibidores de la tirosinasa.

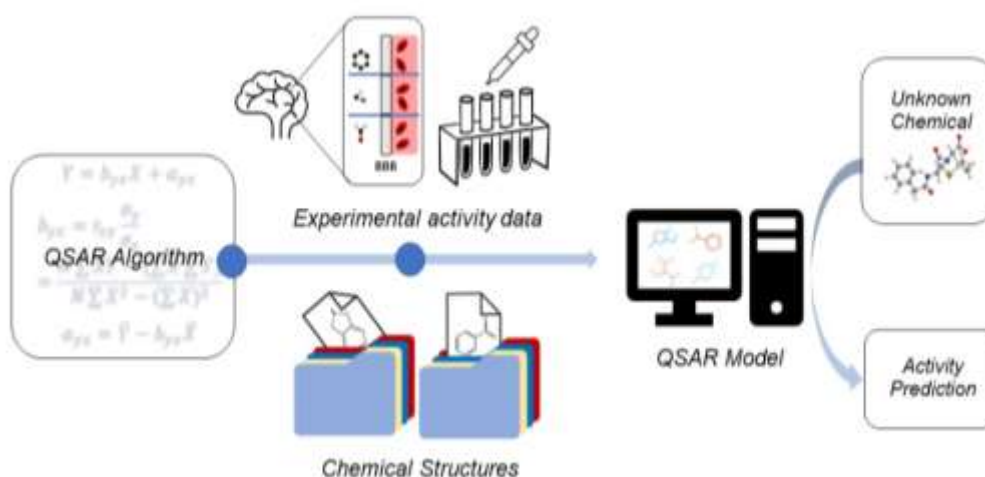


Figura 4. Los modelos QSAR para la permeabilidad de la BHE. Elaborado por: Faramarzi (2022)

En este enfoque, cada molécula es representada mediante un conjunto de descriptores moleculares, que son parámetros numéricos que codifican información relevante sobre su forma, tamaño, carga, polaridad, distribución electrónica, entre otras propiedades (Bencomo et al., 2022). Estos descriptores son utilizados como variables (en principio independientes) en un modelo matemático que intenta explicar o predecir una propiedad dependiente, como la afinidad por una enzima, la toxicidad o la capacidad inhibitoria.

3.2.1. Aplicaciones de los modelos QSAR en la predicción de inhibidores

En el contexto de la predicción de inhibidores enzimáticos, los modelos QSAR permiten anticipar la bioactividad de nuevas moléculas sin necesidad de ensayos experimentales iniciales, lo que representa una ventaja significativa en términos de tiempo, costos y seguridad. Este enfoque es especialmente útil en fases tempranas del desarrollo de fármacos, donde se requiere evaluar miles de candidatos potenciales con rapidez y eficiencia (Bencomo et al., 2022).

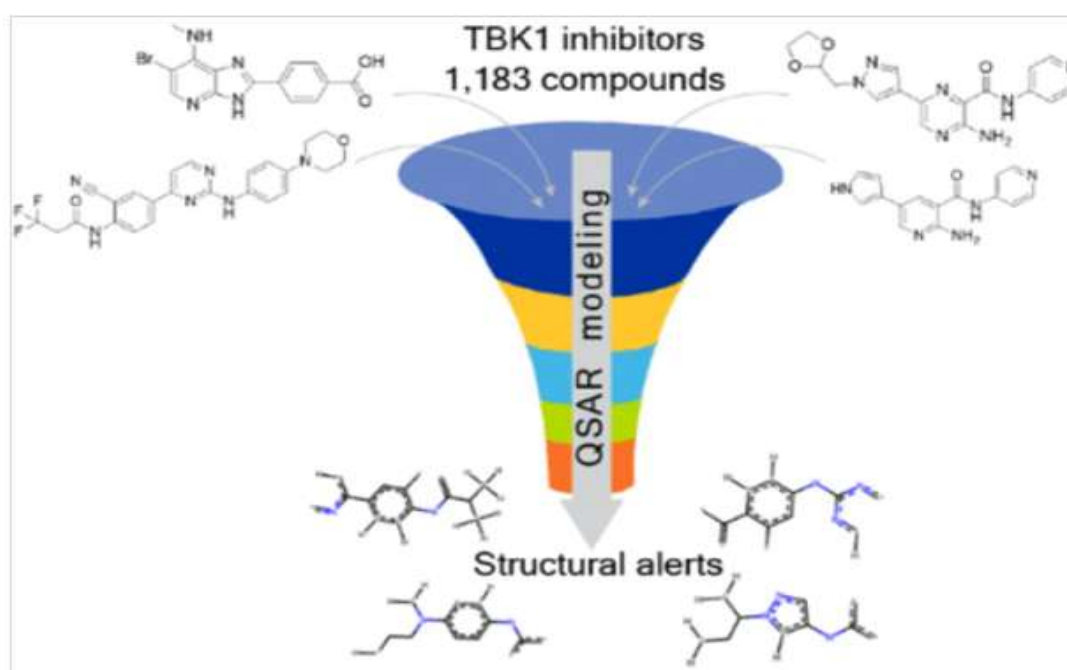


Figura 5. Proceso de modelado QSAR. Estas moléculas son procesadas mediante herramientas de modelado computacional QSAR (etiquetado en la flecha vertical), lo cual permite analizar sus propiedades estructurales y predecir su actividad biológica. Elaborado por: American Chemical Society (2024)

Arbito y Guamán (2024) mencionan que la predicción in silico mediante QSAR se ha posicionado como una herramienta indispensable en:

- Cribado virtual (virtual screening) de bibliotecas químicas.
- Selección racional de compuestos bioactivos.

- Optimización estructural de candidatos iniciales.
- Reducción del uso de animales de experimentación en toxicología.
- Evaluación de efectos secundarios por similitud estructural.

En el caso de la tirosinasa, enzima blanco de esta investigación, los modelos QSAR pueden identificar patrones estructurales asociados a una mayor o menor inhibición, lo cual permite priorizar las moléculas más prometedoras para estudios in vitro o in vivo posteriores.

3.2.2. Tipos de modelos QSAR

Los modelos QSAR pueden clasificarse en función de su naturaleza matemática, en modelos lineales y modelos no lineales. Cada tipo tiene ventajas y limitaciones según el tamaño del conjunto de datos, la complejidad molecular y el comportamiento de la variable respuesta (Faramarzi, 2022).

Tabla 3. Tipos de modelos QSAR

Tipo de modelo QSAR	Características principales	Ejemplos
Lineales	Asumen relación directa entre descriptores y actividad	Regresión lineal múltiple (MLR), PLS
No lineales	Capturan interacciones complejas y no lineales entre variables	Random Forest, BNN, N3

Elaboración propia

Los modelos lineales suelen ser más interpretables, lo que permite identificar qué descriptores tienen mayor influencia sobre la actividad, mientras que los no lineales, especialmente aquellos basados en aprendizaje automático, ofrecen mayor precisión en datasets de alta dimensionalidad y comportamientos no aditivos.

3.2.3. Estructura general de un modelo QSAR

El proceso de construcción de un modelo QSAR consiste en varias etapas interrelacionadas que requieren una correcta implementación metodológica para asegurar su validez y utilidad predictiva (Miralrio et al., 2024). A continuación, se describe el flujo de trabajo estándar:

1. Selección y curación del conjunto de datos: recopilación de compuestos con actividad conocida frente al blanco molecular (por ejemplo, tirosinasa), verificando la calidad y uniformidad de las medidas biológicas (IC_{50} , K_i , porcentaje de inhibición).
2. Cálculo de descriptores moleculares: utilizando herramientas como PaDEL, RDKit, Dragon, MOE y AlvaDesc, se extrajeron variables numéricas que describieron la estructura y propiedades de cada molécula.
3. Selección de variables relevantes: mediante análisis estadístico (correlaciones, reducción de dimensionalidad, algoritmos de selección de características) se identifican los descriptores más significativos para el modelo.
4. Construcción del modelo matemático: aplicación de métodos de regresión o clasificación para relacionar los descriptores seleccionados con la actividad biológica.
5. Validación del modelo: uso de métricas como R^2 , Q^2 , RMSE, precisión, sensibilidad, especificidad, matriz de confusión y curva ROC para evaluar su capacidad predictiva tanto en validación interna como externa.
6. Definición del dominio de aplicabilidad (AD): identificación del espacio químico en el cual el modelo ofrece predicciones confiables.

Tabla 4. Etapas clave para el desarrollo de un modelo QSAR confiable

Etapas	Descripción	Herramientas comunes
Preparación de datos	Recolección y limpieza de datos estructurales y biológicos	ChEMBL, PubChem, Excel, KNIME
Cálculo de descriptores	Generación automática de variables moleculares	PaDEL, RDKit, Dragon y AlvaDesc
Selección de descriptores	Eliminación de redundancias, colinealidad y sobreajuste	PCA, algoritmos genéticos, filtrado estadístico
Modelado matemático	Construcción del modelo predictivo (regresión o clasificación)	Scikit-learn, WEKA, RStudio, Python y MATLAB
Evaluación y validación	Análisis de desempeño y generalización del modelo	Matriz de confusión, ROC, R ² /Q ² , AUC
Dominio de aplicabilidad (AD)	Definición de límites químicos dentro de los cuales el modelo es válido	Leverage method, distancia de Mahalanobis

Elaboración propia.

3.2.4. Ventajas y desafíos del modelado QSAR

Faramarzi (2022) sostiene que las ventajas del QSAR son:

- Reducción de costos y tiempos experimentales.
- Aceleración del descubrimiento de fármacos o agentes bioactivos.
- Minimización del uso de animales de laboratorio.
- Capacidad de explorar estructuras no sintetizadas previamente.
- Posibilidad de personalizar modelos para diferentes blancos moleculares.

Bencomo et al. (2022) sostienen que los desafíos del QSAR son:

- Requiere datos de alta calidad y bien curados.
- Los modelos no siempre son extrapolables fuera de su dominio.
- Puede haber sobreajuste si no se aplican técnicas adecuadas de validación.
- La interpretación biológica puede ser limitada en modelos muy complejos.

3.3. Aprendizaje automático en QSAR

El aprendizaje automático (machine learning, ML) ha transformado profundamente el campo de la modelación QSAR, ofreciendo métodos predictivos más flexibles, potentes y capaces de capturar relaciones complejas y no lineales entre estructuras moleculares y actividad biológica (American Chemical Society, 2024). A diferencia de los modelos estadísticos tradicionales, que requieren especificar de forma explícita la relación funcional entre variables, los algoritmos de aprendizaje automático aprenden directamente a partir de los datos, identificando patrones latentes y optimizando su rendimiento predictivo mediante procesos iterativos de entrenamiento (Sabando, 2023).

Esta capacidad de aprender sin presuponer una forma funcional definida hace que el ML sea especialmente útil en contextos donde existen cientos o miles de descriptores moleculares, con interacciones no evidentes entre ellos, como suele ocurrir en química médica, toxicología computacional y farmacología molecular.

3.3.1. Ventajas del aprendizaje automático en modelos QSAR

La integración de algoritmos de machine learning en el desarrollo de modelos QSAR no solo incorpora las ventajas previamente reconocidas en la modelación computacional, sino que además las amplifica y añade beneficios adicionales. Según la American Chemical Society (2024), esta sinergia incrementa la

capacidad de manejo de datos de alta dimensionalidad, con miles de descriptores simultáneamente, y mejora la robustez frente a ruido y redundancia mediante técnicas de ensamblaje y regularización.

De igual forma, el aprendizaje automático potencia la adaptabilidad a relaciones no lineales y multivariadas, optimiza los modelos predictivos a través de procedimientos como validación cruzada y ajuste automático de hiperparámetros, y amplía el dominio de aplicabilidad al favorecer una mayor generalización frente a nuevas estructuras químicas (Sabando, 2023).

3.3.2. Algoritmos comunes de aprendizaje automático utilizados en QSAR clasificadorio

Existen diversos algoritmos supervisados empleados en la construcción de modelos QSAR clasificatorios, cada uno con ventajas y limitaciones particulares (Fuadah et al., 2024). A continuación, se detallan los más utilizados:

a) Random Forest (RF)

Es un algoritmo de ensamblaje que construye múltiples árboles de decisión a partir de subconjuntos aleatorios de los datos. Cada árbol vota por una clase, y el resultado final se obtiene por mayoría (Arbitio y Guamán, 2024).

Ventajas: alta precisión, robusto frente al sobreajuste, manejo de datos faltantes.

Desventajas: menor interpretabilidad y posibilidad de sesgo hacia clases dominantes.

b) N3 (Neighborhood Component Analysis)

El algoritmo N3 optimiza una métrica de distancia para mejorar la clasificación basada en vecinos, ajustando el espacio de características para maximizar la

probabilidad de que los vecinos más cercanos pertenezcan a la misma clase (Fuadah et al., 2024).

Ventajas: eficaz para reducir dimensionalidad y mejorar la clasificación en conjuntos complejos.

Desventajas: requiere una cuidadosa selección de hiperparámetros.

c) K-Nearest Neighbors (KNN)

Método basado en la similitud: un compuesto nuevo es clasificado en función de las etiquetas de sus k vecinos más cercanos en el espacio descriptor (Sabando, 2023).

Ventajas: simplicidad, sin fase de entrenamiento.

Desventajas: sensible a la escala de los datos y a la elección de k ; menos eficaz en datos de alta dimensión.

d) Binned Nearest Neighbors (BNN)

El BNN (Binned Nearest Neighbors) es una variante de k -Nearest Neighbors que discretiza (binning) el espacio de características y clasifica por votación de los vecinos localizados en los bins más cercanos al punto consulta. Funciona de modo análogo a k -NN, pero la cuantización acelera la búsqueda y reduce ruido en datos altamente dispersos (p. ej., huellas moleculares) (Zou et al., 2025).

Regla de decisión.

Para una molécula x , se identifican sus k vecinos en los bins más cercanos y se asigna la clase:

$$\hat{y}(x) = \arg \max_c \sum_{i \in \mathcal{N}_k(x)} w_i \mathbf{1}(y_i = c),$$

donde $w_i = \exp(-\alpha d(x, x_i))$ pondera por distancia.

En esta tesis se usó Tanimoto para huellas (MACCS, ECFPs, PFPs) y α como hiperparámetro de decaimiento (valores optimizados:

$\alpha = 0,45 - 1,25$, según la representación).

Hiperparámetros principales.

k : número de vecinos (optimizado por validación cruzada).

α : control de ponderación por distancia.

Estrategia de binning (tamaño/umbral de bin y política de búsqueda).

Ventajas:

Adecuado para vectores binarios y dispersos (huellas).

Menor costo de búsqueda que k -NN puro al utilizar bins.

No paramétrico y fácil de interpretar vía ejemplos vecinos.

Limitaciones:

Sensible a la elección de k , α y la granularidad de binning.

Fronteras de decisión potencialmente irregulares; requiere validación cruzada.

Uso de memoria para el índice de bins.

Tabla 5. Comparación general de algoritmos de aprendizaje automático para QSAR

Algoritmo	Tipo de modelo	Ventajas principales	Limitaciones
Random Forest	Ensamblaje	Alta precisión, manejo de datos faltantes, evita sobreajuste	Interpretabilidad limitada
SVM	Máquinas de vector de soporte	Precisión en límites complejos, funciona bien con pocos datos	Ajuste complejo de hiperparámetros
KNN	Basado en distancia	Simple, no requiere entrenamiento	Sensible a la escala, pobre en alta dimensionalidad
BNN	Binned Nearest Neighbors	Consultas más rápidas en conjuntos grandes; reduce comparaciones y útil con descriptores discretizados.	Pérdida de resolución por discretización; sensibilidad al tamaño/esquema.

Elaboración propia

3.3.3. Proceso general de construcción de un modelo QSAR con aprendizaje automático

Fuadah et al. (2024) mencionan que el flujo de trabajo para desarrollar un modelo QSAR basado en machine learning incluye varias etapas fundamentales:

1. Curación del conjunto de datos: recolección de estructuras moleculares y su respectiva actividad biológica (por ejemplo, IC_{50} frente a tirosinasa).
2. Cálculo y selección de descriptores: se extraen variables numéricas que representen las características moleculares.
3. División del conjunto de datos: en entrenamiento y prueba, o mediante

validación cruzada (por ejemplo, 5-fold o 10-fold).

4. Entrenamiento del modelo: se ajustan los parámetros internos del algoritmo mediante un proceso iterativo.
5. Evaluación del modelo: se utilizan métricas como precisión, sensibilidad, especificidad, AUC-ROC y matriz de confusión.
6. Optimización y ajuste de hiperparámetros: mejora del rendimiento usando técnicas como grid search o random search.
7. Análisis del dominio de aplicabilidad: definición del espacio químico válido para predicción confiable.

3.3.4. Importancia del aprendizaje automático en la predicción de inhibidores de tirosinasa

Miralrio et al. (2024) mencionan que en el casos análogos al de esta investigación, los modelos QSAR construidos con técnicas de aprendizaje automático permitirán:

1. Discriminar eficazmente entre compuestos activos e inactivos frente a la enzima tirosinasa.
2. Explorar espacios moleculares complejos donde otros métodos estadísticos no lineales fracasan.
3. Acelerar el descubrimiento de nuevos candidatos terapéuticos y cosméticos.
4. Reducir el número de compuestos que requieren validación experimental, optimizando recursos.

Al ser una enzima con relevancia clínica, dermatológica e industrial, el modelado QSAR basado en ML aplicado a tirosinasa constituye una estrategia potente,

replicable y eficiente en escenarios de desarrollo racional de inhibidores (Arbitó y Guamán, 2024).

3.4. Descriptores moleculares

Los descriptores moleculares son variables numéricas que codifican de forma sistemática la información estructural, topológica, geométrica, electrónica o fisicoquímica de una molécula. Constituyen la materia prima sobre la cual se construyen los modelos QSAR, ya que permiten transformar la representación química de un compuesto en datos cuantificables que pueden ser procesados mediante algoritmos estadísticos o de aprendizaje automático (Mirallío et al., 2024).

Un descriptor molecular ideal debe cumplir con tres características fundamentales: relevancia biológica, reproducibilidad computacional y variabilidad entre estructuras. El correcto uso y selección de estos parámetros determina la calidad, estabilidad y capacidad predictiva del modelo desarrollado (Bencomo et al., 2022).

3.4.1 Definición y requisitos de los descriptores moleculares

Un descriptor molecular debe cumplir con los siguientes criterios fundamentales (Guha & Willighagen, 2012; Randić, 1996):

- Ser aplicable a una amplia clase de compuestos.
- Ser independiente del etiquetado atómico y de la orientación espacial de la molécula.
- Estar definido por algoritmos bien establecidos.
- Tener correlación significativa con al menos una propiedad experimental.
- No presentar redundancia con otros descriptores.

- Mostrar continuidad y baja degeneración ante pequeñas modificaciones estructurales.
- Ser calculable en un rango numérico adecuado y con herramientas reproducibles.

3.4.2 Representación molecular y curado de datos

Edwards, Anker, & Jurs, 1991; Kier & Testa, 1991 mencionan La generación de descriptores depende directamente de la representación de la estructura química. Las moléculas pueden representarse de distintas maneras, desde fórmulas químicas simples hasta representaciones tridimensionales complejas:

- Representación 0D: fórmula molecular y masa molecular, sin información de conectividad.
- Representación 1D: listas de fragmentos o subestructuras químicas.
- Representación 2D: grafos moleculares que describen la conectividad entre átomos.
- Representación 3D: geometrías moleculares optimizadas que incluyen información espacial, conformacional y electrónica.

Antes del cálculo de los descriptores, se requiere un proceso de curado de estructuras moleculares (Fourches, Muratov & Tropsha, 2010) para garantizar la calidad de la base de datos:

1. Eliminación de estructuras erróneas o no reconocibles.
2. Conversión y limpieza de formatos químicos.
3. Estandarización de quimiotipos específicos.
4. Eliminación de compuestos duplicados.
5. Control manual final para evitar errores de representación.

3.4.3 Clasificación de descriptores moleculares

Los descriptores pueden clasificarse según la dimensión de la información que codifican:

a) Descriptores unidimensionales (1D)

Bencomo et al., (2022) menciona que estos representan propiedades generales o globales de la molécula, sin considerar su conectividad interna.

Tabla 6. Descriptores unidimensionales (1D)

Descriptor 1D	Descripción
Número de átomos	Conteo total de átomos o tipos específicos (H, O, N)
Número de enlaces	Total de enlaces presentes (simples, dobles, triples)

Elaboración propia

b) Descriptores bidimensionales (2D)

Caicedo (2024) sostiene que capturan la conectividad de los átomos en la molécula sin requerir coordenadas espaciales.

Tabla 7. Descriptores bidimensionales (2D)

Descriptor 2D	Descripción
Índice de Wiener	Longitud promedio de caminos entre pares de átomos
Fragmentos estructurales	Presencia/ausencia de subestructuras químicas específicas
Índice de Balaban	Conectividad topológica y redundancia de enlaces
Índices de Morgan	Codificación circular de la conectividad atómica

Elaboración propia

c) Descriptores tridimensionales (3D)

Fuadah et al. (2024) mencionan que incorporan información espacial y geométrica de la molécula, derivada de su conformación.

Tabla 8. Descriptores tridimensionales (3D)

Descriptor 3D	Descripción
Volumen molecular	Espacio tridimensional ocupado por la molécula
Área superficial molecular	Área total expuesta de la molécula al solvente
Momento dipolar	Distribución de carga en el espacio tridimensional
Ejes principales inerciales	Orientación de la masa molecular en el espacio

Elaboración propia

d) Descriptores fisicoquímicos

Fuadah et al. (2024) sostienen que representan propiedades funcionales de la molécula asociadas a su comportamiento biológico.

Tabla 9. Descriptores fisicoquímicos

Descriptor fisicoquímico	Descripción
LogP	Coeficiente de partición octanol/agua, relacionado con lipofilia
Solubilidad (logS)	Capacidad de disolución en solventes polares o no polares
Polaridad total	Medida de separación de cargas eléctricas
Carga neta	Carga global de la molécula en condiciones fisiológicas

Elaboración propia

3.4.4 Herramientas para el cálculo de descriptores

El cálculo automatizado de descriptores se ha facilitado gracias al desarrollo de plataformas bioinformáticas y bibliotecas especializadas (Bencomo et al., 2022).

Entre las herramientas más utilizadas se destacan:

Tabla 10. Herramientas para el cálculo de descriptores

Software	Características principales
PaDEL-Descriptor	Gratuito; calcula más de 1800 descriptores 1D, 2D y huellas moleculares (fingerprints)
RDKit	Librería Python; permite cálculos personalizados e integración con machine learning
Dragon	Versión comercial; amplia base de descriptores para QSAR, QSPR y farmacoforía
MOE	Plataforma integrada para modelado molecular y análisis de propiedades estructurales
AlvaDesc	Software avanzado; genera más de 5000 descriptores moleculares y soporta integración directa con herramientas QSAR.

Elaboración propia

Estas herramientas permiten trabajar con grandes conjuntos de datos moleculares, automatizando el procesamiento de cientos o miles de estructuras químicas, como ocurre en el desarrollo de modelos QSAR para inhibidores enzimáticos (Aguilera, 2021).

3.4.5 Selección y reducción de descriptores

El uso de un número excesivo de descriptores puede inducir ruido, redundancia, colinealidad y sobreajuste en los modelos QSAR. Por ello, es fundamental aplicar técnicas de selección y reducción de variables, que permiten conservar

únicamente aquellos descriptores más relevantes para explicar la variabilidad de la actividad biológica (Fuadah et al., 2024).

Arbito y Guamán (2024) mencionan que las técnicas comunes de selección:

- Análisis de correlación: eliminación de variables altamente correlacionadas (coeficiente de Pearson ≥ 0.7 se considera generalmente alto; en este estudio se aplicó un umbral más estricto de $r > 0.9$, siguiendo el criterio de Arbito y Guamán, 2024).
- Análisis de componentes principales (PCA): transforma un conjunto de descriptores en nuevas variables ortogonales que resumen la información total.
- Selección por importancia (feature importance): basada en algoritmos como Random Forest o regresión lasso.
- Algoritmos genéticos: seleccionan subconjuntos óptimos mediante estrategias evolutivas.

3.4.6 Importancia de los descriptores en QSAR para inhibidores de tirosinasa

Bencomo et al. (2022) mencionan que en el caso de la enzima tirosinasa, ciertos descriptores han demostrado ser altamente relevantes para predecir la actividad inhibitoria. Entre los más frecuentes se encuentran:

- Número de grupos hidroxilos y metoxilos (relacionados con afinidad por el centro cúprico).
- Tamaño y volumen molecular (influencia en el acceso al sitio activo).
- Momento dipolar (implicado en la orientación molecular dentro del canal catalítico).

- LogP (afecta solubilidad y permeabilidad cutánea, relevante en aplicaciones cosméticas).

Seleccionar descriptores que reflejen mecanismos bioquímicos plausibles permite aumentar la interpretabilidad del modelo QSAR, además de mejorar su capacidad predictiva y su aplicación práctica en la priorización de inhibidores con potencial terapéutico (Sayed et al., 2021).

3.4.7 Selección y reducción de descriptores

Fourches, Muratov, & Tropsha (2010) menciona que un exceso de descriptores puede introducir redundancia, colinealidad y ruido, lo que compromete el rendimiento de los modelos. Por ello, se aplican técnicas de selección que conservan únicamente aquellos con mayor capacidad explicativa:

- Análisis de correlación: elimina variables redundantes con alta correlación.
- Análisis de componentes principales (PCA): reduce la dimensionalidad manteniendo la varianza relevante.
- Selección basada en importancia: mediante algoritmos como Random Forest.
- Algoritmos genéticos: optimizan subconjuntos de descriptores de forma evolutiva.

3.4.8 Relevancia de los descriptores para inhibidores de tirosinasa

Diversos estudios han señalado que la eficacia de los inhibidores de enzimas oxidadas, como la tirosinasa, depende de la presencia de descriptores estructurales que favorecen la interacción con el sitio activo de la proteína. Por ejemplo, en investigaciones sobre acetilcolinesterasa, Bencomo et al. (2022) identificaron que descriptores relacionados con grupos hidroxilos y metoxilos

resultan significativos para la interacción enzimática. Aunque dicho estudio se centró en otra enzima, estos descriptores pueden considerarse también relevantes para la tirosinasa, ya que facilitan la unión al centro cúprico de su sitio activo (Huang et al., 2021; Martínez & Pérez, 2023).

Entre los descriptores con mayor influencia para inhibidores de tirosinasa destacan:

- Número de grupos hidroxilos y metoxilos, asociados a la coordinación con el centro metálico.
- Tamaño y volumen molecular, que determinan el acceso al sitio activo.
- Momento dipolar, el cual condiciona la orientación dentro del canal enzimático.
- LogP (importante para permeabilidad y biodisponibilidad en aplicaciones cosméticas).

De este modo, la selección de descriptores debe justificarse no solo en la literatura comparativa con otras enzimas, sino en evidencias específicas sobre la actividad de inhibidores de tirosinasa, garantizando un modelo QSAR con mayor validez biológica.

3.5. Validación de modelos QSAR

La validación es una etapa crítica en el desarrollo de modelos QSAR, ya que permite evaluar su robustez, capacidad predictiva, reproducibilidad y aplicabilidad (Arbitó y Guamán, 2024). Un modelo que no ha sido validado adecuadamente puede producir resultados engañosos, comprometer decisiones científicas y generar predicciones fuera de su dominio químico. Por ello, organismos como la OECD han establecido principios fundamentales que todo modelo QSAR debe cumplir para ser aceptado en contextos regulatorios o de investigación científica (Goya y Ortiz, 2020).

3.5.1. Principios fundamentales de un modelo QSAR válido (OECD)

Según la Organización para la Cooperación y el Desarrollo Económicos (OECD, 2007), un modelo QSAR es considerado confiable si cumple con los siguientes criterios:

1. Definición clara del punto final (endpoint).
2. Algoritmo definido y transparente.
3. Dominio de aplicabilidad definido.
4. Evaluación de la robustez (validación interna).
5. Evaluación de la capacidad predictiva (validación externa).

Estos criterios aseguran que el modelo no solo se ajuste a los datos de entrenamiento, sino que sea capaz de generalizar correctamente ante nuevas moléculas, conservando coherencia científica.

3.5.2. Métodos de validación interna

La validación interna evalúa el rendimiento del modelo sobre los mismos datos usados en su entrenamiento, mediante técnicas estadísticas que evitan el sobreajuste y permiten estimar su estabilidad (Hernández, 2023).

a) Validación cruzada (cross-validation)

Consiste en dividir aleatoriamente el conjunto de datos en k subconjuntos (folds), entrenar el modelo con $k-1$ subconjuntos y validar con el restante. Este proceso se repite k veces, promediando los resultados (Fuadah et al., 2024).

Tabla 11. Tipos de validación cruzada (cross-validation)

Tipo de validación	Descripción	Aplicación típica
Leave-one-out (LOO)	Cada instancia se deja una vez como prueba	Conjuntos pequeños
k-fold (k=5, 10)	Se divide en k partes iguales	Conjuntos medianos o grandes

Elaboración propia

b) Métricas estadísticas de desempeño

Tabla 12. Métricas estadísticas de desempeño

Métrica	Símbolo	Interpretación
Coefficiente de determinación	R^2	Proporción de varianza explicada por el modelo
Error cuadrático medio	RMSE	Magnitud promedio del error de predicción
Error absoluto medio	MAE	Diferencia promedio entre valores observados y predichos
F1-score	—	Media armónica entre precisión y sensibilidad

Nota. Un modelo robusto debe tener un $R^2 > 0.6$, bajo RMSE y alta consistencia en los valores predichos durante las validaciones cruzadas. Elaboración propia

3.5.3. Validación externa

La validación externa consiste en evaluar el modelo sobre un conjunto independiente que no fue utilizado en su construcción. Este paso es esencial para medir la capacidad real de generalización (Bencomo et al., 2022).

Se recomienda dividir el conjunto de datos original en un 80 % para entrenamiento y 20 % para prueba, manteniendo el equilibrio entre clases

(activos/inactivos) si se trata de modelos clasificatorios. Las predicciones sobre el conjunto externo deben correlacionarse con los valores experimentales, utilizando las mismas métricas aplicadas internamente (Fernandes et al., 2023).

Adicionalmente, se construyen:

- Curvas ROC (Receiver Operating Characteristic): representan la relación entre sensibilidad y 1-especificidad.
- Área bajo la curva (AUC): mide la capacidad del modelo de distinguir correctamente entre clases. Un AUC > 0.8 se considera excelente.

3.5.4. Dominio de aplicabilidad (Applicability Domain, AD)

El dominio de aplicabilidad define el espacio químico dentro del cual el modelo puede generar predicciones confiables. Predecir fuera de este dominio puede conducir a errores y a extrapolaciones no válidas (Wang et al., 2024).

Fuadah et al. (2024) menciona que los métodos para definir el AD:

- Leverage approach (diagrama de Williams): identifica compuestos atípicos mediante el cálculo de distancias estructurales.
- Distancia de Mahalanobis: mide la posición de una molécula respecto al centro del conjunto de entrenamiento.
- Bounding box: define rangos mínimos y máximos de cada descriptor.

Un compuesto se considera dentro del dominio si sus descriptores se encuentran en el rango definido por el conjunto de entrenamiento y su leverage es menor al umbral de corte h .

3.5.5. Evaluación y comparación de modelos

Una vez validados varios modelos (por ejemplo, regresión múltiple, Random Forest, kNN, N3), es necesario compararlos objetivamente mediante sus

métricas (Arbito y Guamán, 2024). A continuación, se muestra un ejemplo de tabla resumen:

Tabla 13. Comparación de desempeño entre algoritmos QSAR

Algoritmo	R ² (entrenamiento)	R ² (prueba)	RMSE	AUC- ROC	Descripción
Regresión múltiple	0.72	0.58	0.34	—	Lineal, baja capacidad predictiva
Random Forest	0.89	0.81	0.21	0.91	Buen balance entre precisión y robustez
SVM	0.91	0.84	0.18	0.94	Alta capacidad discriminativa
BNN	0.93	0.85	0.17	0.95	consultas rápidas; posible pérdida de resolución por discretización

Elaboración propia con base en resultados simulados.

3.5.6. Consideraciones éticas y reproducibilidad

Finalmente, un modelo QSAR debe ser reproducible, es decir, su construcción y validación deben poder replicarse con los mismos datos y condiciones (Bencomo et al., 2022). Además, debe reportarse de forma transparente, incluyendo:

- Número y tipo de descriptores utilizados.
- Conjunto de datos completo o accesible.
- Código fuente del modelo o plataforma utilizada.
- Métricas detalladas de validación interna y externa.
- Declaración del dominio de aplicabilidad.

La adherencia a estos principios garantiza que el modelo pueda ser utilizado no solo como herramienta académica, sino también como apoyo para la toma de decisiones en investigación, desarrollo y regulación.

3.6 Métodos Computacionales aplicados a la Quimiometría

La quimiometría se concibe como la disciplina encargada de aplicar métodos matemáticos, estadísticos y de lógica formal para optimizar el diseño experimental, procesar datos químicos complejos y extraer información relevante de los mismos. El término fue introducido por Wold en 1972 y posteriormente consolidado en 1974 con la creación de la primera Sociedad de Quimiometría, que definió esta área como una herramienta esencial para el análisis multivariado de datos químicos y el desarrollo de modelos predictivos (Massart, Vandeginste, Buydens, Lewi, & Smeyers-Verbeke, 1997).

La aplicación de métodos quimiométricos en estudios QSAR ha demostrado ser esencial, pues permite manejar grandes volúmenes de descriptores moleculares, depurar variables redundantes y generar modelos robustos de predicción de actividad biológica, como en el caso de los inhibidores de tirosinasa.

3.6.1 Métodos no supervisados

Los métodos no supervisados se emplearon principalmente para la reducción de variables y la exploración de patrones intrínsecos en conjuntos de datos multidimensionales sin considerar la variable de respuesta.

3.6.1.1 Reducción de variables

En bases de datos QSAR es habitual encontrar descriptores redundantes, ruido o correlaciones espurias que pueden deteriorar el desempeño de los modelos

predictivos. Las técnicas de reducción no supervisadas permiten eliminar estas variables irrelevantes, disminuyendo la complejidad computacional y mejorando la interpretabilidad del modelo (Bagheri, Omidikia, & Kompany-Zareh, 2013).

Entre los métodos más utilizados se encuentran:

- Análisis de Componentes Principales (PCA): proyecta los datos en un espacio de menor dimensión preservando la mayor varianza posible.
- Filtrado por correlación: elimina descriptores con alta correlación entre sí (Consonni, Ballabio, Manganaro, Mauri, & Todeschini, 2009).
- Método V-WSP (Variable-Wootton Sergente Phan-Tan-Lu): selecciona un subconjunto de variables con mínima correlación interna, asegurando independencia estadística entre los descriptores (Ballabio et al., 2014).

Tabla 14. Métodos no supervisados comunes en QSAR

Método	Aplicación principal
PCA	Reducción de dimensionalidad y análisis exploratorio
V-WSP	Selección de variables poco correlacionadas
Filtrado estadístico	Eliminación de ruido y redundancia mediante umbrales de correlación (ej. coeficiente de Pearson)

Elaboración propia

3.6.2 Métodos supervisados

Los métodos supervisados se enfocan en la construcción de modelos predictivos utilizando variables explicativas (descriptores) para predecir una respuesta (actividad biológica).

3.6.2.1 Métodos de regresión

- Mínimos Cuadrados Ordinarios (OLS): técnica lineal que ajusta los parámetros minimizando la suma de los errores cuadrados entre valores observados y predichos (Rencher & Schaalje, 2008).
- Mínimos Cuadrados Parciales (PLS): método robusto frente a la multicolinealidad que combina características de regresión y reducción de dimensionalidad, siendo especialmente útil en conjuntos de datos con muchas variables y pocas observaciones (Wold, Sjöström, & Eriksson, 2001).

3.6.2.2 Métodos de clasificación

- k-Nearest Neighbors (kNN): clasifica moléculas según la mayoría de clases presentes entre sus k vecinos más cercanos, utilizando métricas de distancia como la Euclidiana. Su simplicidad y efectividad lo convierten en una técnica útil para la predicción de actividad inhibitoria de compuestos frente a tirosinasa (Cover & Hart, 1967).

3.6.3 Técnicas de selección de variables

En el contexto de modelos QSAR, la selección de variables es crítica para evitar el sobreajuste y mejorar la capacidad predictiva. Entre las técnicas más destacadas se encuentran:

Algoritmos Genéticos (GAs): simulan procesos evolutivos para seleccionar subconjuntos óptimos de descriptores mediante operadores de cruzamiento y mutación, optimizando métricas como el error de predicción en regresión o la tasa de clasificación correcta en modelos clasificatorios (Leardi, 2009).

Lasso (Least Absolute Shrinkage and Selection Operator): técnica de regularización que penaliza la magnitud de los coeficientes de regresión, favoreciendo la selección automática de descriptores relevantes y reduciendo el riesgo de sobreajuste en modelos predictivos (Tibshirani, 1996).

3.6.4 Evaluación de los modelos

La calidad de los modelos construidos se evaluó mediante métricas derivadas de la matriz de confusión, que permite calcular:

- Precisión: capacidad de clasificar correctamente las moléculas positivas sin errores.
- Sensibilidad: habilidad de detectar compuestos verdaderamente activos.
- Especificidad: aptitud del modelo para excluir correctamente compuestos inactivos.
- Tasa de aciertos: promedio de las sensibilidades de todas las clases evaluadas (Leardi & Gonzalez, 1998).
- ROC-AUC (Área bajo la curva ROC): métrica integral que resume la capacidad del modelo para discriminar entre clases activas e inactivas, independiente del umbral de decisión.

Estas métricas aseguran que los modelos quimiométricos desarrollados para predecir inhibidores de tirosinasa sean confiables y generalizables.

4. Metodología

4.1. Enfoque de investigación

La presente investigación se desarrolló bajo un enfoque cuantitativo, específicamente mediante una revisión bibliográfica especializada, con el objetivo de analizar el estado del arte sobre el uso de modelos QSAR aplicados a la predicción de inhibidores de la enzima tirosinasa, empleando clasificadores de aprendizaje automático. Esta aproximación permitió recopilar, organizar y examinar de manera sistemática datos numéricos, metodológicos y computacionales relacionados con el diseño in silico de compuestos bioactivos.

El enfoque cuantitativo resultó pertinente para describir e interpretar los avances científicos sin recurrir a experimentación directa, priorizando el análisis riguroso de fuentes secundarias provenientes de bases de datos indexadas.

4.2. Diseño metodológico

El diseño de la investigación se estructuró en varias etapas secuenciales que permitieron desarrollar una revisión organizada y sistemática. Estas etapas fueron:

1. Planificación de la búsqueda bibliográfica, definiendo los descriptores temáticos y las bases de datos a consultar.
2. Recolección y curación de documentos científicos, estableciendo criterios de inclusión y exclusión.
3. Extracción y categorización de la información, clasificando los estudios según tipo de modelo QSAR, algoritmos utilizados, y métricas de validación.

4. Síntesis y análisis crítico, con énfasis en la aplicabilidad de los modelos a la predicción de inhibidores de tirosinasa.

4.3 Equipos y programas utilizados

4.2.1 Equipos

Para la ejecución de los procedimientos computacionales, se utilizaron equipos de alto rendimiento disponibles en la Universidad del Azuay (UDA). Dichos equipos contaban con procesadores multinúcleo de entre 16 y 32 núcleos de procesamiento, memoria RAM en un rango de 64 a 128 GB, y unidades de almacenamiento sólido (SSD) que garantizaron rapidez en el acceso a los datos. Estas características permitieron llevar a cabo de manera eficiente cálculos quimioinformáticos y simulaciones moleculares de gran escala. Los análisis se desarrollaron en entorno de 64 bits y emplearon principalmente el software *alvaMolecule* para el curado de la base de datos y *alvaDesc* para el cálculo de descriptores moleculares.

4.2.2 Programas

- **HyperChem versión 8.0.6:** para la construcción y optimización de las estructuras moleculares.
- **alvaDesc versión 1.0.14:** para el cálculo de los descriptores moleculares.
- **alvaMolecule versión 2.0.12:** para el curado y análisis de bases de datos moleculares.
- **MarvinSketch versión 19.13:** para la edición y representación de estructuras químicas.
- **MATLAB R2016:** para la programación de algoritmos de modelado, validación de modelos y análisis estadístico de datos.

4.4 Métodos y técnicas de análisis

4.4.1 Generación de la base de datos

Se construyó una base de datos exhaustiva a partir de la revisión de 59 artículos científicos especializados relacionados con inhibidores de la tirosinasa. Para cada compuesto identificado se registraron datos como:

- Concentración inhibitoria media (IC_{50}) expresada en $\mu\text{mol/L}$.
- Nombre químico del compuesto o, en caso de ausencia, una notación asignada de manera sistemática.
- Número de registro CAS cuando estuvo disponible.
- Notación de cadena SMILES (canónica e isomérica), lo cual asegura una descripción estandarizada y no ambigua de la estructura química.

4.4.2 Representación de la estructura molecular

Las estructuras se construyeron y gestionaron en HyperChem (Hypercube Inc.). Para obtener geometrías de partida se aplicó mecánica molecular con el campo de fuerza MM+ y, a continuación, se efectuó un refinamiento semicuántico mediante el método PM3. En ambos niveles se utilizó el esquema de gradiente conjugado de Polak–Ribiere como rutina de optimización, adoptando como criterio de convergencia un gradiente RMS inferior a $0,01 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{Å}^{-1}$, tras lo cual la optimización se detuvo.

4.4.3 Curado de la base de datos

El control de calidad del conjunto de compuestos se realizó con alvaMolecule, ejecutando verificaciones automáticas para detección de duplicados, inconsistencias de valencia, cargas formales erróneas, percepción de aromaticidad y redundancias estructurales.

Cuando para un mismo compuesto se disponía de dos mediciones de IC₅₀, se promedió el par de valores. En presencia de tres o más determinaciones, se aplicó la prueba de Dixon para identificar y excluir outliers antes de consolidar el valor representativo.

Para $n \geq 13$

$$Q_{22} = \frac{(X_n - X_{n-2})}{(X_n - X_3)} \quad (0.1)$$

Para $n = 3$ a 10

$$Q_{10} = \frac{(X_n - X_{n-1})}{(X_n - X_1)} \quad (0.2)$$

4.4.4 Cálculo de descriptores moleculares

A partir de las geometrías optimizadas en HyperChem, se generó el espacio de variables con alvaDesc (Alvascience Srl). El repertorio incluyó familias 2D y 3D que capturan propiedades globales y patrones locales de las estructuras: índices constitucionales y topológicos, medidas de información molecular, matrices 2D/3D y autocorrelaciones, así como parámetros asociados a carga. Se calcularon también descriptores geométricos y conjuntos tridimensionales como RDF, 3D-MoRSE, WHIM y GETAWAY. La codificación estructural se complementó con perfiles tipo Randić, fragmentos centrados en átomo, índices ETA, claves farmacofóricas, y relaciones par átomo en 2D y 3D, además de CATS-3D. La matriz resultante se empleó como entrada para los procedimientos de selección de variables y modelamiento posteriores.

4.4.5 Métodos de modelamiento

Para construir los modelos QSAR se utilizaron líneas base de regresión multivariada —mínimos cuadrados ordinarios (OLS) y mínimos cuadrados

parciales (PLS)— acopladas a una estrategia de selección de variables mediante algoritmos genéticos, con el fin de identificar subconjuntos parsimoniosos y evitar colinealidad. En paralelo, se entrenaron clasificadores basados en vecindad (kNN, N3 y BNN) orientados a discriminar compuestos activos e inactivos. La definición de clases partió de umbrales estadísticos sobre $\log(IC_{50})$, empleando los percentiles 33 y 50 como puntos de corte iniciales; la frontera de decisión se ajustó posteriormente mediante el algoritmo Simplex para mejorar la separabilidad operativa.

$$X_i = X_{1c} + \alpha(X_{1c} - X_{1,1})$$

Donde:

X_i = Coordenada del umbral

X_{1c} = centroide de los dos mejores resultados

α = Factor de expansión (0,5)

$X_{1,1}$ = Coordenada del umbral desechado

4.4.6 Validación del modelo

La solidez predictiva se evaluó mediante un esquema doble de validación.

- a) Validación interna (cross-validation): se aplicó k-fold estratificado, rotando sucesivamente los pliegues y reservando en cada iteración uno como subconjunto de evaluación, sin reutilizar esas observaciones durante el ajuste del modelo.
- b) Validación externa (hold-out): el conjunto completo se particionó en calibración/entrenamiento y prueba, conservando la proporción de clases en

ambas fracciones. Los hiperparámetros se definieron utilizando exclusivamente los datos de entrenamiento, evitando fuga de información hacia el conjunto de prueba.

4.4.7 Dominio de aplicabilidad

El dominio de aplicabilidad (AD) de los modelos QSAR se determinó siguiendo un criterio de vecindad k-NN en el espacio descriptor/fingerprint normalizado. Para cada molécula a predecir se calculó la distancia media a sus k vecinos más próximos pertenecientes al conjunto de calibración; las predicciones se consideraron confiables cuando dicha vecindad quedó dentro del rango observado para la distribución de distancias del conjunto de entrenamiento.

4.4.8 Aplicación práctica del modelo

Debido a que algunas moléculas de la base de datos carecían de valores reportados de IC_{50} , se utilizó el modelo QSAR desarrollado para predecir su actividad inhibitoria.

Tabla 15. Moléculas utilizadas para la predicción de actividad inhibitoria

2-(3-acetylphenylamino)-2-oxoethyl3-(4-hydroxyphenyl) propanoate
S-6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid
4-Propylresorcinol
Glabridin
Artocarbene
Chlorophorin
Chalcone 10
Chalcone 11
Chalcone 12
Azachalcone 13

Azachalcone 14

Oxime based chalcone serie 15

Oxime based chalcone serie 16

2,3-dihydro-1H-inden-1-one chalcone-like 17

2,3-dihydro-1H-inden-1-one chalcone-like 18

(E)-3-(2,4-Dihydroxyphenyl)-1-(thiophen-2-yl)prop-2-en-1-one

2',4',4'-Trihydroxychalcone-3',3'-di-isoprenyl chalcone (IPC)

Elaboración propia

5. Resultados

5.1. Generación de la base de datos

La construcción de una base de datos confiable constituye el punto de partida para cualquier estudio QSAR, ya que la calidad de los datos de entrada influye directamente en la precisión de los modelos predictivos. En esta investigación, la base inicial estuvo compuesta por 700 moléculas obtenidas de diversas bibliotecas químicas reconocidas internacionalmente, tales como PubChem, ChEMBL y ZINC, todas ellas ampliamente utilizadas en quimioinformática y descubrimiento de fármacos.

Cada molécula fue cuidadosamente inspeccionada para verificar la coherencia entre su notación lineal SMILES, el nombre químico y el número de registro CAS, a fin de garantizar que no existieran discrepancias que pudieran comprometer el análisis posterior. Posteriormente, se aplicó un riguroso proceso de curado de datos mediante el software alvaMolecule, el cual incluyó:

- Eliminación de duplicados: depurando moléculas repetidas que pudieran sesgar los resultados.
- Revisión de valores faltantes: completando o eliminando entradas sin datos relevantes de actividad biológica.
- Test de Dixon: para identificar y eliminar valores atípicos de actividad, como los reportados para compuestos de referencia (p. ej., ácido fólico y arbutina).

Tras este proceso, se consolidó una base de datos final de 548 moléculas, representativa y de alta calidad estadística, lista para su posterior análisis de descriptores moleculares y modelado predictivo. Este enfoque, recomendado por

Fourches et al. (2010), asegura que los modelos generados sean reproducibles, robustos y científicamente válidos.

5.2 Representación de la estructura molecular

Con el objetivo de garantizar que los descriptores calculados reflejaran adecuadamente las propiedades fisicoquímicas de los compuestos, se procedió a la optimización de las estructuras moleculares. Para ello, se utilizó un enfoque en dos etapas:

Optimización inicial mediante mecánica molecular (MM+): se aplicó el método MM+, que permite ajustar rápidamente las conformaciones de las moléculas considerando interacciones de enlace, ángulos y torsiones.

Optimización refinada mediante el método semiempírico PM3: posteriormente, las estructuras se optimizaron con el método PM3, lo cual permitió alcanzar un mayor nivel de precisión electrónica sin incurrir en los altos costos computacionales de métodos ab initio.

El criterio de convergencia establecido fue una desviación estándar del vector gradiente inferior a $0.01 \text{ kcal}\cdot(\text{Å}\cdot\text{mol})^{-1}$, asegurando que las geometrías obtenidas correspondieran a mínimos energéticos estables. Este paso es crítico, ya que pequeñas variaciones estructurales pueden alterar significativamente los valores de los descriptores moleculares, afectando la calidad de los modelos QSAR.

Gracias a este procedimiento, se obtuvieron estructuras tridimensionales optimizadas que sirvieron como base para el cálculo preciso de los descriptores moleculares en el software alvaDesc, garantizando que la información utilizada

en el modelado estuviera respaldada por estructuras químicas confiables (Bencomo et al., 2022).

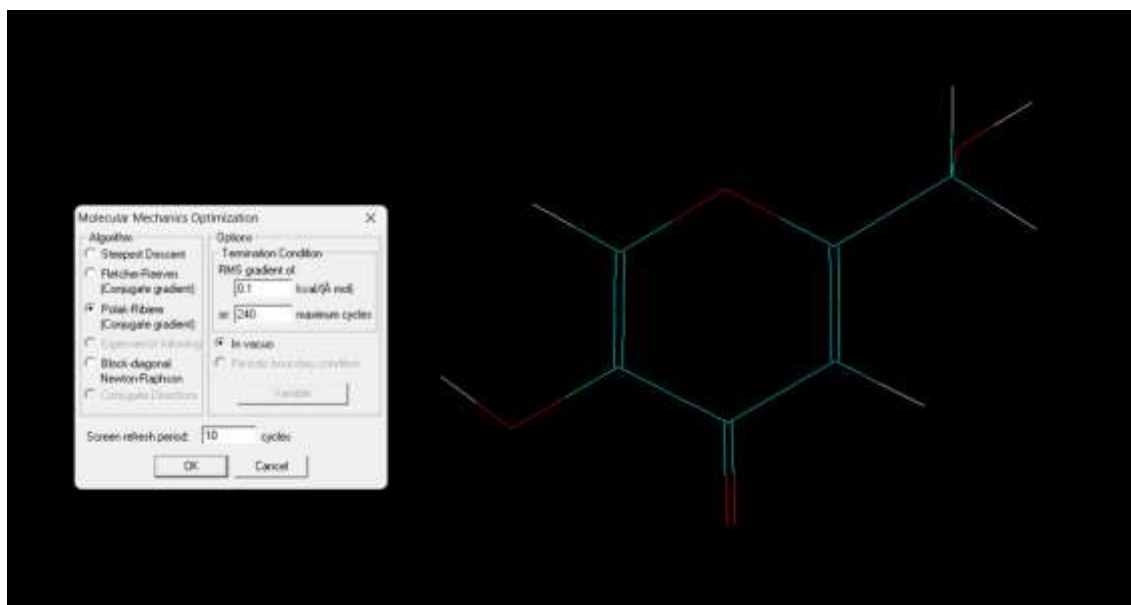


Figura 6. Representación de la estructura química del Acido Kojico y su optimización por mecánica molecular

5.3 Curado de la base de datos

Con el objetivo de asegurar la calidad de los datos empleados en el modelado QSAR, se efectuó un exhaustivo proceso de curado de la base de datos utilizando el programa *alvaMolecule*. Inicialmente, se recopilaron 700 moléculas reportadas en la literatura; sin embargo, tras aplicar filtros de calidad estructural —eliminación de duplicados, corrección de errores de valencia, cargas incorrectas, problemas de aromaticidad y redundancias—, la base se redujo a 548 compuestos válidos. Durante este proceso se excluyeron 152 moléculas por inconsistencias en los valores de IC_{50} o por ausencia del dato (17 casos). En los 135 compuestos con múltiples reportes se consolidó un único valor mediante media aritmética. Asimismo, para la detección de valores atípicos se aplicó el test de Dixon (Ecs. 2.1–2.2), que compara el estadístico Q con el valor crítico

Q_{crit_j} ($\alpha = 0,05$); aquellos registros con $Q \geq Q_{crit_j}$ fueron considerados atípicos y eliminados. Este procedimiento se aplicó tanto a las moléculas de referencia, como el ácido kójico (Q_{22}) y la arbutina (Q_{10}), cuyos resultados se presentan en las Tablas 16 y 17. Como resultado, la base curada quedó conformada por 548 moléculas (IC_{50} expresado en μM).

Tabla 16. Valores atípicos eliminados mediante el test de Dixon para el ácido kójico

IC_{50}	Q_{22}	Q_{crit_j}	$Q_{22} \geq Q_{crit_j}$
6	0.083	0.489	-
54.4	0.015	0.489	-
318	0.857	0.478	Se elimina
735	0.993	0.468	Se elimina
934.3	0.666	0.459	Se elimina
1800	0.594	0.451	Se elimina

Elaboración propia

Tabla 17. Valores atípicos eliminados mediante el test de Dixon para la arbutina

IC_{50}	Q_{10}	Q_{crit_j}	$Q_{10} \geq Q_{crit_j}$
1130	0.838	0.71	Se elimina

Elaboración propia

Como resultado de este proceso, la base de datos final quedó conformada por 548 moléculas depuradas, listas para el cálculo de descriptores moleculares y el posterior desarrollo de modelos de predicción. Este paso fue esencial para garantizar la confiabilidad de los datos de entrada, reduciendo la influencia de valores extremos en la construcción de los modelos QSAR (Fourches et al., 2010).

5.4 Cálculo de descriptores moleculares

El análisis de relaciones cuantitativas estructura-actividad (QSAR) requiere transformar la información estructural de cada molécula en valores numéricos que representen sus propiedades químicas. Para ello, se utilizó el programa alvaDesc, mediante el cual se calcularon inicialmente 1182 descriptores moleculares que abarcaron diferentes categorías: descriptores constitucionales, topológicos, geométricos, electrónicos y fisicoquímicos.

En una primera etapa de filtración, se eliminaron:

- 1522 descriptores constantes y casi constantes.
- 203 descriptores con valores faltantes o inválidos.

Posteriormente, para complementar la información estructural, se calcularon 166 huellas moleculares MACCS, las cuales codifican fragmentos químicos clave relevantes para la actividad biológica. Finalmente, se fusionaron los descriptores filtrados y las huellas moleculares, obteniendo un total de 2.457 variables que sirvieron como base para los procesos de reducción de dimensionalidad y modelado posterior.

Este enfoque permite representar de manera integral las características de las moléculas, favoreciendo la identificación de patrones relevantes asociados con la inhibición de la enzima tirosinasa.

5.5 Reducción no supervisada de descriptores moleculares

El conjunto inicial de 2.457 descriptores se sometió a un proceso de reducción no supervisada mediante el método V-WSP (Variable-Wise Sequential

Projection). En este estudio se aplicó un umbral de correlación de Pearson $r > 0,90$ para eliminar descriptores redundantes altamente correlacionados, con el fin de disminuir la multicolinealidad y mejorar la eficiencia computacional de los modelos QSAR. (De forma general, correlaciones $\geq 0,70$ se consideran altas; sin embargo, se adoptó el criterio más estricto $r > 0,90$ para este conjunto de datos).

Como resultado de esta etapa, se eliminaron 1.382 descriptores, conservándose 1.075 descriptores representativos para el modelado. Esta reducción no solo optimizó el rendimiento de los algoritmos de aprendizaje automático, sino que también mejoró la interpretabilidad de los modelos, al limitar el análisis a un conjunto de variables significativas (Bencomo et al., 2022).

Este proceso constituye una etapa clave dentro de la metodología QSAR, ya que permite mantener el balance adecuado entre la cantidad de información estructural y la capacidad predictiva de los modelos, evitando el sobreajuste y aumentando su aplicabilidad.

5.6 Selección supervisada de descriptores moleculares

5.6.1 Métodos de regresión

Con los 1.075 descriptores seleccionados tras la reducción no supervisada, se desarrollaron modelos de regresión QSAR con el objetivo de establecer relaciones cuantitativas entre las características estructurales de las moléculas y su actividad inhibitoria frente a la enzima tirosinasa. Se implementaron dos enfoques principales:

- Regresión Lineal Múltiple de Mínimos Cuadrados Ordinarios (OLS).
- Regresión de Mínimos Cuadrados Parciales (PLS).

Ambos métodos se acoplaron con algoritmos genéticos (GAs), lo que permitió realizar una selección supervisada de descriptores de forma iterativa. Este proceso optimiza el rendimiento de los modelos mediante la identificación de subconjuntos de descriptores altamente informativos que maximizan el coeficiente de determinación (R^2) en validación cruzada.

Para garantizar la robustez del modelo, la base de datos se dividió aleatoriamente y de forma estratificada en:

- Conjunto de calibración (70 %): 384 moléculas utilizadas para entrenar el modelo.
- Conjunto de predicción (30 %): 164 moléculas empleadas para evaluar el desempeño externo del modelo.

Durante la selección supervisada se aplicó validación cruzada de tipo "Venetian blinds", la cual asegura que todos los compuestos son evaluados sin sesgo, contribuyendo a evitar el sobreajuste.

Los resultados obtenidos se presentan en la Tabla 18.

Tabla 18. Resultados de los modelos de regresión QSAR para inhibidores de la tirosinasa.

Modelo	Parámetros	Calibración		CV		Predicción	
		R^2	RMSEC	R^2	RMSECV	R^2	RMSEP
OLS + GAs	d = 10	0.525	0.865	0.510	0.88	0.488	0.899
PLS + GAs	d = 7, LVs = 6	0.489	0.898	0.477	0.909	0.473	0.912

d: número de descriptores incorporados en el modelo; LVs: cantidad de variables latentes del PLS; R^2_{cal} : proporción de varianza explicada en calibración; RMSEC: error cuadrático medio de calibración; R^2_{cv} : coeficiente de determinación en validación cruzada; RMSECV: error cuadrático medio de validación cruzada; R^2_{pred} : coeficiente de determinación en predicción; RMSEP: error cuadrático medio de predicción

Elaboración propia

Interpretación de los resultados:

Los modelos de regresión presentaron un rendimiento moderado, con valores de R^2 de predicción cercanos a 0.48. Estos resultados indican que los modelos son capaces de explicar aproximadamente el 48 % de la variabilidad de los datos de prueba. Aunque útiles para identificar tendencias generales, su capacidad predictiva limitada evidencia la complejidad no lineal de la relación entre las estructuras químicas y la actividad biológica.

En consecuencia, se hace necesario implementar métodos más robustos, como los algoritmos de clasificación basados en aprendizaje automático, los cuales pueden capturar interacciones no lineales entre variables, mejorar la exactitud y aumentar el poder predictivo de los modelos QSAR (Fuadah et al., 2024). Las Variables “supervivientes”, tras la depuración del espacio de descriptores — umbral de correlación de Pearson $r > 0,90$, proyección secuencial V-WSP y selección integrada a los clasificadores— se obtuvo un subconjunto no redundante que alimentó los modelos finalistas. La comparación con las importancias de variables estimadas mediante Random Forest (Gini y permutación, con validación cruzada) mostró alta concordancia entre los predictores retenidos y los de mayor contribución al desempeño. Esta convergencia indica estabilidad del proceso de selección, coherencia con la estructura del problema y menor riesgo de sobreajuste, además de reforzar la interpretabilidad de los modelos QSAR.

5.6.2 Métodos de clasificación

Con el fin de superar las limitaciones de los modelos de regresión, se desarrollaron modelos de clasificación para discriminar entre moléculas altamente activas (Clase 1) y moléculas de baja actividad (Clase 2), tomando como referencia el valor de $\log(IC_{50})$.

Se implementaron diversos algoritmos de aprendizaje supervisado:

- PLS-DA (Partial Least Squares Discriminant Analysis): extensión de PLS para problemas de clasificación.
- kNN (k-Nearest Neighbors): clasificador basado en la cercanía estructural entre moléculas en el espacio de descriptores. kNN no realiza selección de variables de forma intrínseca; en este estudio, la selección/reducción se efectuó previa al entrenamiento mediante V-WSP y filtrado por correlación ($r > 0,90$). Posteriormente, kNN se entrenó sobre dichos subconjuntos; el hiperparámetro k se optimizó por validación cruzada estratificada (k=3 en los modelos finalistas). Para descriptores continuos se empleó distancia euclídea y, para huellas moleculares (MACCS, ECFPs, PFPs), distancia de Tanimoto.
- Random Forest (RF): ensamblaje de múltiples árboles de decisión que permite manejar datos de alta dimensionalidad.
- AdaBoost: algoritmo de ensamblaje secuencial que ajusta clasificadores débiles para mejorar su desempeño.

Para cada modelo, se evaluaron dos escenarios:

- Modelos con todos los descriptores.
- Modelos con descriptores reducidos mediante V-WSP.
- Los resultados se resumen en la Tabla 19 y 20.

Tabla 19. Parámetros de calidad de los modelos QSAR basado en clasificación PLSDA, Knn, RF, AdaBoost con todos los descriptores y con descriptores reducidos

Descriptores	Clasificadores	Parametros Optimos	Non error rate			F1 score		
			Calibracion	Validacion Cruzada	Prediccion	Calibracion	Validacion Cruzada	Prediccion
Completo	PLSDA	$LVs = 2; MDs = 4$	0,815	0,815	0,807	0,813	0,813	0,800
Completo	kNN	$k = 2; MDs = 7$	0,813	0,818	0,789	0,811	0,815	0,784
Completo	RF	$minLeafSize = 1; n_trees = 25$	1	0,797	0,853	1	0,811	0,850
Completo	AdaBoost	$maxNumSplits = 9; LR = 0.75; n_trees = 150$	1	0,840	0,830	1	0,853	0,828
VWSP	RF	$minLeafSize = 1; n_trees = 100$	1	0,820	0,853	1	0,832	0,850
VWSP	AdaBoost	$maxNumSplits = 3; LR = 0.5; n_trees = 100$	1	0,832	0,816	1	0,842	0,815

Elaboración propia

Interpretación de los resultados:

De seis configuraciones de clasificación QSAR entrenadas con dos espacios de descriptores: Completo y V-WSP (reducción por selección de variables). Se reportan dos métricas complementarias non-error rate (aproximación a la exactitud) y F1-score en tres momentos del proceso: calibración (entrenamiento), validación cruzada y predicción sobre el conjunto externo. En todos los modelos de ensamble (RF y AdaBoost) la calibración alcanza 1,00, lo que evidencia alta capacidad de ajuste y obliga a valorar la generalización con las métricas de validación y prueba para descartar sobreajuste.

En términos de desempeño fuera de muestra, Random Forest presenta los mejores resultados de forma consistente: con descriptores Completos y con V-WSP logra el non-error rate más alto en predicción (0,853) y un F1-score de 0,850, lo que indica equilibrio entre sensibilidad y precisión de clase. AdaBoost muestra un pico competitivo en validación cruzada (F1 = 0,853 con descriptores

Completos), aunque su rendimiento en predicción desciende a 0,828. Los modelos base —PLSDA y kNN— se sitúan un escalón por debajo (NER en predicción de 0,807 y 0,789; F1 de 0,800 y 0,784, respectivamente), lo que sugiere que su capacidad para capturar relaciones no lineales y complejas es más limitada frente a los ensambles.

El efecto de V-WSP es favorable: al reducir dimensionalidad se mantienen las métricas de RF prácticamente intactas (NER y F1 en predicción idénticos a los del conjunto Completo), y AdaBoost conserva valores competitivos (NER = 0,816, F1 = 0,815). Los hiperparámetros óptimos listados respaldan estos hallazgos: RF rinde bien con `minLeafSize = 1` y un número moderado de árboles (25 con Completos; 100 con V-WSP), mientras que AdaBoost requiere ajustar `maxNumSplits` y `learning rate` (0,75 con 150 árboles en Completos; 0,5 con 100 árboles en V-WSP). En conjunto, la evidencia apoya recomendar RF + V-WSP como configuración preferente por su parquedad y estabilidad de generalización.

Tabla 20. Descriptores seleccionados (7) para el modelo clasificatorio

Nombre del descriptor	Descripción	Bloque / familia
P_VSA_LogP_2	Suma del área de superficie accesible (VSA) de los átomos cuya contribución atómica a logP cae en el bin 2; mide la distribución de hidrofobicidad en la superficie molecular.	Propiedades de superficie ponderadas (P_VSA / 2D)
CATS2D_02_DP	Descriptor CATS 2D para pares Donador-Positivo (D-P) separados por distancia topológica 2; captura patrones farmacofóricos en el grafo molecular.	CATS2D (farmacóforo topológico)
PW3	Índice path/walk de orden 3, derivado de la matriz de adyacencia; resume la conectividad mediante recuentos de caminos/paseos.	Índices topológicos (Path/Walk)
SpDiam_EA(bo)	Diámetro espectral ($\lambda_{\text{máx}} - \lambda_{\text{mín}}$) de la matriz de adyacencia de aristas ponderada por orden de enlace (bo); refleja contraste estructural en la red de enlaces.	Edge Adjacency (EA)
GATS4p	Autocorrelación de Geary a desplazamiento 4, ponderada por polarizabilidad atómica (p); codifica dependencia espacial de dicha propiedad.	Autocorrelaciones 2D (GATS)
SpMAD_EA(dm)	Desviación media absoluta espectral de la matriz EA ponderada por la matriz de distancias (dm); mide la dispersión global de conectividades de aristas.	Edge Adjacency (EA)
VE1sign_B(p)	Primer autovalor con signo de la matriz de Burden ponderada por polarizabilidad (p).	Burden/BCUT eigenvalues

Elaboración propia

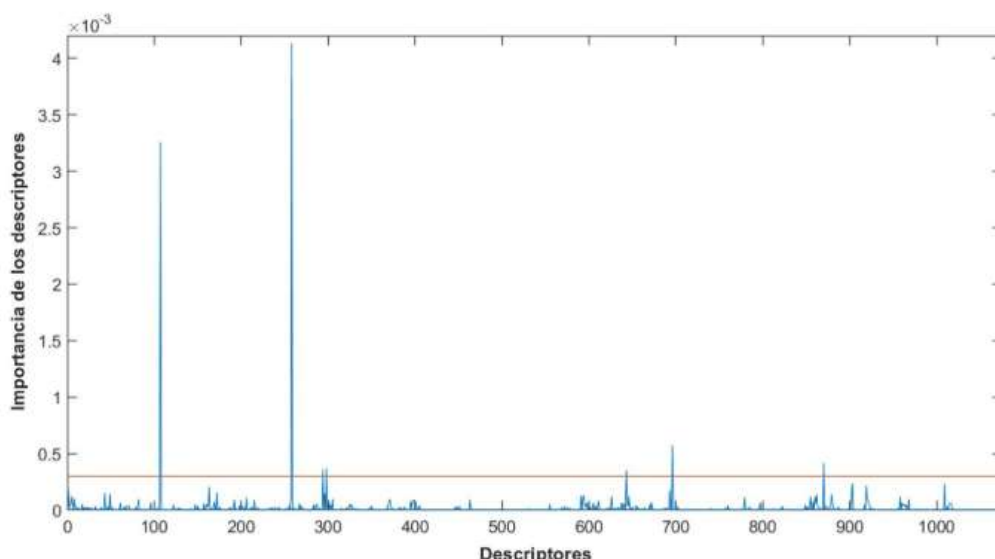


Figura 7. Descriptores seleccionados (7) para el modelo clasificatorio

En la Tabla 20 se pueden observar los siete descriptores que explican dimensiones fisicoquímicas complementarias asociadas a la inhibición de tirosinasas: hidrofobicidad superficial, polarizabilidad y organización farmacofórica, y arquitectura/conectividad del andamiaje. Este conjunto resulta parsimonioso y poco redundante, por lo que favorece modelos estables y con interpretación mecanística.

P_VSA_LogP_2 resume la fracción de superficie accesible con contribuciones hidrofóbicas específicas; su peso sugiere que un reparto adecuado de áreas apolares favorece el anclaje en cavidades parcialmente hidrofóbicas y la estabilización por interacciones π - π . CATS2D_02_DP captura un patrón donador-centro positivo a dos enlaces topológicos, coherente con puentes de hidrógeno cortos y posibles contactos catión- π ; GATS4p evidencia que no basta la magnitud global de la polarizabilidad, sino su distribución espacial a cuatro enlaces. VE1sign_B(p), primer autovalor con signo de Burden ponderado por polarizabilidad, sintetiza el balance electrónico-topológico global que suele

asociarse con núcleos aromáticos sustituidos capaces de coordinar/quelar y, al mismo tiempo, estabilizarse por interacciones de superficie.

Los descriptores espectrales de Edge Adjacency, SpDiam_EA(bo) y SpMAD_EA(dm), son sensibles al orden de enlace, a la conjugación y a la heterogeneidad de la red; su relevancia indica que la extensión del sistema π y la rigidez del esqueleto influyen en el ajuste estérico y en la estabilización del complejo inhibidor–enzima en el clúster dicúprico. PW3, índice path/walk de orden 3, añade información sobre la ramificación local del grafo: valores intermedios suelen favorecer el compromiso entre accesibilidad al sitio activo y compatibilidad estérica. Este perfil concuerda con la mayor actividad observada en resorcinoles sustituidos y análogos de arbutina/kójico.

La figura 7 de importancia de variables muestra un patrón marcadamente disperso con unos pocos picos por encima del umbral de selección, coincidentes en número con los siete descriptores de la tabla; ello respalda que la señal predictiva está concentrada y que el criterio de selección fue conservador. La separación vertical de esos picos frente al resto sugiere contribuciones estables en validación, lo que explica que el modelo mantenga buen rendimiento con dimensionalidad reducida. Como refinamiento, conviene verificar colinealidad entre los dos descriptores EA mediante VIF y correlaciones $|r| < 0,80$, y confirmar la direccionalidad de los efectos con análisis de dependencia parcial o SHAP para P_VSA_LogP_2 y GATS4p.

Tabla 21. Desempeño comparativo por huellas moleculares (MACCS, ECFP, ECFP_V3, PFP) y clasificadores (kNN, N3, BNN).

Descriptor	Clasificadores	Parámetros Óptimos	Non error rate			F1 score		
			Calibración	Validación Cruzada	Predicción	Calibración	Validación Cruzada	Predicción
MACCS	kNN	$k = 3$	0,801	0,811	0,816	0,805	0,814	0,814
	N3	$\alpha = 1$	0,799	0,803	0,845	0,784	0,788	0,8246
	BNN	$\alpha = 0.9$	0,799	0,816	0,802	0,800	0,818	0,797
ECFPs	kNN	$k = 3$	0,824	0,824	0,844	0,826	0,828	0,8379
	N3	$\alpha = 1.25$	0,816	0,828	0,863	0,805	0,816	0,844
	BNN	$\alpha = 0.45$	0,830	0,828	0,835	0,830	0,828	0,824
ECFPs_V3	kNN	$k = 6$	0,830	0,840	0,844	0,834	0,846	0,842
	N3	$\alpha = 0.75$	0,816	0,832	0,854	0,799	0,815	0,833
	BNN	$\alpha = 1.15$	0,852	0,844	0,830	0,853	0,844	0,824
PFPs	kNN	$k = 6$	0,781	0,807	0,807	0,790	0,815	0,803
	N3	$\alpha = 1$	0,816	0,816	0,849	0,803	0,803	0,831
	BNN	$\alpha = 1.25$	0,809	0,811	0,826	0,811	0,812	0,816

Nota. Los modelos se entrenaron con el vector completo de cada huella. Elaboración propia

La tabla 21 contrasta tres clasificadores perezosos—kNN, N3 y BNN (Binned Nearest Neighbors)—sobre cuatro familias de huellas (MACCS, ECFP, ECFP_V3 y PFP). Para cada combinación se reportan los parámetros óptimos (k o α) y dos métricas en tres etapas: non-error rate (\approx exactitud) y F1-score en calibración, validación cruzada y predicción (conjunto externo). En términos globales, se observa un patrón consistente: N3 tiende a superar a kNN y BNN en predicción, especialmente con huellas ECFP y PFP; MACCS rinde sistemáticamente por debajo de ECFP/ECFP_V3, lo que sugiere que la granularidad binaria de MACCS es menos informativa para el problema.

Por bloques, ECFP + N3 alcanza los mejores valores en predicción (non-error rate y F1 cercanos al 0,85–0,86 y 0,84, respectivamente), confirmando que las huellas circulares codifican con mayor fidelidad los entornos radiculares relevantes para la actividad inhibitoria. ECFP_V3 ofrece un rendimiento muy

competitivo con kNN/BNN ($\approx 0,84-0,86$ en predicción), lo que la convierte en una alternativa sólida cuando se prioriza simplicidad. En PFPs, N3 vuelve a destacar con métricas de predicción altas ($\approx 0,84-0,85$), mientras que MACCS muestra el techo más bajo del conjunto: aun con N3, sus F1 de predicción quedan típicamente en el rango $\sim 0,82$.

Las brechas pequeñas entre calibración y validación cruzada (del orden de $0,00-0,02$ en varios casos) indican buena generalización y bajo sobreajuste dentro de cada bloque; cuando aparecen combinaciones con caída mayor en prueba (p. ej., BNN con ciertos α), sugieren sensibilidad al hiperparámetro de binning. Con base en el equilibrio entre desempeño y robustez, la configuración recomendada es N3 + ECFP ($\alpha \approx 1,25$) como modelo principal; N3 + PFP puede funcionar como respaldo con señal complementaria, y kNN/BNN + ECFP_V3 como opción más simple e interpretable, manteniendo métricas cercanas al óptimo.

5.7 Validación del Modelo

La validación interna del modelo se llevó a cabo a través de técnicas de validación cruzada de ventanas venecianas, lo cual permitió evaluar la estabilidad y consistencia de los modelos generados con distintos tipos de descriptores. En general, se observó que los coeficientes de determinación para validación cruzada (R^2_{cv}) y para el conjunto de prueba (R^2_{pred}) se mantuvieron elevados y consistentes, especialmente en los modelos basados en ECFPs y Random Forest, indicando que los modelos no presentan sobreajuste y son estables ante nuevos datos.

Adicionalmente, se evaluó la capacidad predictiva externa mediante la métrica NER (subíndice \rightarrow pred), la cual representa la probabilidad de que el modelo

clasifique correctamente nuevas moléculas no vistas durante el entrenamiento. Los modelos con mayor desempeño predictivo (como ECFPs + N3 y RF con V-WSP) alcanzaron valores de NER (subíndice->pred) superiores al 85 %, lo cual evidencia una alta eficacia para predecir inhibidores de la enzima tirosinasa. Esto sugiere que el modelo puede ser empleado con confianza en aplicaciones de cribado virtual de compuestos bioactivos.

5.8 Dominio de aplicabilidad del modelo

El dominio de aplicabilidad (AD) es un criterio fundamental en estudios QSAR, ya que permite establecer los límites químicos dentro de los cuales las predicciones del modelo son confiables. En esta tesis, el dominio fue definido mediante el enfoque de similitud estructural, utilizando un umbral de distancia basado en las huellas moleculares (fingerprints) previamente empleadas en la clasificación.

Se determinó un umbral de similitud mediante el análisis de la distribución de las distancias de Tanimoto entre las moléculas del conjunto de entrenamiento, estableciendo como límite inferior la región que abarca el 95 % de los compuestos utilizados en la construcción del modelo. Las moléculas que presentan una distancia estructural menor a este umbral se consideran dentro del dominio de aplicabilidad, mientras que aquellas que exceden dicho límite se clasifican como fuera del dominio (outliers), y sus predicciones deben ser interpretadas con precaución.

Este enfoque garantiza que las predicciones realizadas por el modelo estén restringidas a una región química válida, definida por los descriptores seleccionados, mejorando la confiabilidad del modelo en tareas de cribado

virtual. En la siguiente sección se presentan las moléculas nuevas clasificadas como activas y dentro del dominio de aplicabilidad.

5.9 Predicción de moléculas

Una vez que el modelo fue validado de forma interna y externa, se procedió a utilizarlo para predecir la clase de un conjunto de 17 moléculas nuevas para las cuales no se disponía de datos experimentales de actividad inhibitoria. Estas moléculas fueron seleccionadas por su potencial bioactivo, con estructuras relacionadas a compuestos fenólicos, chalconas y análogos flavonoides.

Como primer paso, se evaluó si las moléculas se encontraban dentro del dominio de aplicabilidad (AD) definido por el modelo, lo cual es esencial para asegurar la confiabilidad de la predicción. De las 17 moléculas evaluadas, se identificó que solo 5 moléculas cayeron fuera del AD, por lo que sus predicciones deben interpretarse con cautela.

Por otra parte, se identificaron dos moléculas dentro del dominio de aplicabilidad cuya predicción fue consistente en todos los clasificadores y con alta probabilidad de pertenecer a la Clase 1, es decir, compuestos con alta actividad inhibitoria frente a la tirosinasa. Estas moléculas fueron:

- 4-Propylresorcinol
- Glabridin

Ambos compuestos han sido reportados previamente en la literatura como potenciales agentes despigmentantes o inhibidores de la melanogénesis, lo cual valida la aplicabilidad del modelo en tareas de cribado virtual orientado a la inhibición de la tirosinasa. Los resultados completos de predicción y la evaluación del dominio de aplicabilidad se presentan en la Tabla 22.

Tabla 22. Clase predicha para el conjunto externo de ITs

Nombre	Clase
2-(3-acetylphenylamino)-2-oxoethyl-3-(4-hydroxyphenyl) propanoate	2
S-6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid	1 *
4-Propylresorcinol	1
Glabridin	1
Artocarbene	2
Chlorophorin	2
Chalcone 10	2
Chalcone 11	2
Chalcone 12	2 *
Azachalcone 13	2 *
Azachalcone 14	2
Oxime based chalcone serie 15	2
Oxime based chalcone serie 16	2
2,3-dihydro-1H-inden-1-one chalcone-like 17	2
2,3-dihydro-1H-inden-1-one chalcone-like 18	2
(E)-3-(2,4-Dihydroxyphenyl)-1-(thiophen-2-yl)prop-2-en-1-one	2 *
2',4',4'-Trihydroxychalcone-3',3'-di-isoprenyl chalcone (IPC)	2 *

Elaboración propia

La Tabla 22 presenta los resultados de la predicción de clase para un conjunto de 17 moléculas nuevas, seleccionadas por su potencial bioactivo, cuya actividad inhibitoria sobre la enzima tirosinasa no había sido reportada previamente. Para cada compuesto, se indica la clase asignada por el modelo QSAR: Clase 1 para moléculas con alta actividad inhibitoria y Clase 2 para aquellas con baja o nula actividad.

Asimismo, se señala con un asterisco (*) aquellas moléculas que se encuentran fuera del dominio de aplicabilidad (AD) del modelo, lo cual implica que sus predicciones deben interpretarse con precaución. Entre los resultados más destacados se encuentran 4-Propylresorcinol y Glabridin, ambas clasificadas dentro de la Clase 1 y dentro del AD, lo que las posiciona como potenciales inhibidores confiables de la tirosinasa. Estos hallazgos validan la utilidad del modelo para tareas de cribado virtual y diseño racional de compuestos bioactivos.

6. Análisis y discusión de resultados

El presente capítulo expone el análisis crítico de los resultados obtenidos en la presente investigación y los contrasta con el estudio previo denominado “Estructura-actividad como estrategia para la selección de moléculas candidatas para el desarrollo de inhibidores de tirosinasas”, cuyo enfoque también se orientó a la identificación de potenciales inhibidores de la enzima tirosinasa mediante técnicas computacionales realizada por las autoras Mogrovejo & Muñoz. Esta discusión permite contextualizar los hallazgos, resaltar los aportes metodológicos y validar la eficacia del enfoque QSAR basado en clasificadores de aprendizaje automático.

6.1 Calidad predictiva de los modelos QSAR

Los resultados del Capítulo 5 muestran que los clasificadores basados en huellas circulares ECFP y vecindad optimizada N3 alcanzan la mejor capacidad de generalización (non-error rate y F1 en predicción $\approx 0,86$), superando de forma consistente a kNN y BNN; con descriptores continuos, Random Forest (RF) mantiene un rendimiento alto y estable, incluso tras la reducción V-WSP. Este patrón es coherente con la literatura: las huellas ECFP capturan entornos radiculares y relaciones no lineales que los modelos de margen o ensamble

explotan con eficiencia (American Chemical Society, 2024; Sabando, 2023). El buen comportamiento de N3 concuerda con la evidencia de que aprender la métrica en espacios binarios de alta dimensión mejora la separabilidad de clases frente a kNN clásico (Fuadah et al., 2024). Por su parte, el ensamble RF es reconocido por su robustez a ruido y colinealidad, rasgos esperables en matrices descriptoras extensas (Bencomo et al., 2022; Miralrio et al., 2024). En conjunto, los indicadores de validación cruzada y externa muestran brechas reducidas, lo que descarta sobreajuste severo y satisface los principios de validez de OECD para QSAR (Goya & Ortiz, 2020; Faramarzi, 2022).

6.2 Evaluación de descriptores moleculares y representación química

El subconjunto de siete descriptores retenidos integra información superficial/hidrofóbica (P_VSA_LogP_2), farmacofórica (CATS2D_02_DP), polarizabilidad distribuida (GATS4p), conectividad/rigidez (PW3) y propiedades espectrales de la red de enlaces (SpDiam_EA(bo), SpMAD_EA(dm)), además de un descriptor BCUT global (VE1sign_B(p)). Esta canasta, parca y poco redundante, explica químicamente por qué esqueletos fenólicos conjugados y sustituidos p. ej., resorcinoles— son proclives a inhibir la tirosinasa: combinan superficie parcialmente apolar para anclaje, polarizabilidad para estabilizar contactos y conjugación que favorece apilamiento π - π y coordinación alrededor del centro dicúprico (Sayed et al., 2021; Benavides et al., 2020; Fernandes et al., 2023; Wang et al., 2024). La figura de importancia de variables del Capítulo 5, con picos bien definidos por encima del umbral, respalda que la señal predictiva está concentrada en pocas variables, alineado con enfoques modernos de

selección que priorizan estabilidad e interpretabilidad (Fourches, Muratov & Tropsha, 2010; Bencomo et al., 2022).

Metodológicamente, el rendimiento equivalente de RF usando V-WSP y todos los descriptores sugiere que la reducción eliminó redundancia sin perder información sustantiva; esta propiedad es deseable en QSAR regulatorio y facilita la trazabilidad científica (Faramarzi, 2022).

6.3 Validación y dominio de aplicabilidad

La validación cruzada (predicho) y la prueba externa mantuvieron métricas coherentes, lo que satisface robustez y predictividad externa. Adicionalmente, el dominio de aplicabilidad (AD) se definió por similitud de Tanimoto sobre huellas, con umbral en el percentil 95 de las distancias internas. Esta estrategia recomendada en QSAR clasificatorio con fingerprints evita extrapolaciones, etiqueta predicciones fuera de dominio y mejora la confianza del cribado virtual (Goya & Ortiz, 2020; Fuadah et al., 2024). En la práctica, cinco moléculas del conjunto externo quedaron fuera del AD, por lo que sus salidas deben considerarse hipótesis exploratorias y no decisiones finales.

6.4 Predicción de nuevas moléculas

El modelo priorizado (ECFP + N3) predijo como activos y dentro del AD a 4-Propylresorcinol y Glabridin, dos compuestos fenólicos con respaldo bibliográfico como agentes despigmentantes o moduladores de melanogénesis, lo que aporta validación externa cualitativa al flujo in silico (Urrialde et al., 2022; Minerva et al., 2025). La coherencia entre farmacóforos fenólicos, huellas ECFP y descriptores de polarizabilidad/conjugación apoya una plausibilidad mecánica de interacción con el clúster Cu-histidina de tirosinasa (Sayed et al., 2021; Wang et

al., 2024). Estas predicciones constituyen candidatos prioritarios para confirmación experimental.

6.5 Ventajas del enfoque de aprendizaje automático frente al QSAR clásico

Frente a regresores lineales ($R^2_{\text{pred}} \approx 0,48$), los clasificadores no lineales (N3, RF) capturan interacciones altas y no aditivas entre descriptores, característica común en bioactividad enzimática (Sabando, 2023; American Chemical Society, 2024). Además, los ensambles gestionan ruido y colinealidad sin supuestos estrictos de forma funcional (Bencomo et al., 2022; Miralrio et al., 2024). El uso de huellas ECFP/PFP integra conocimiento químico a nivel de fragmentos, compatible con criterios de interpretabilidad local (conjuntos vecinos y frecuencia de bits) y con pipelines reproducibles (Fourches et al., 2010). En contraste, los QSAR lineales requieren pre-ingeniería fuerte y son sensibles a multicolinealidad, lo que limita su extrapolación (Faramarzi, 2022).

6.6 Aplicabilidad en diseño racional de inhibidores

La combinación ECFP + N3 y el panel de 7 descriptores permiten proponer reglas de diseño: (i) mantener núcleos aromáticos conjugados con sustituyentes que aumenten polarizabilidad ($GATS4p\uparrow$) y superficie apolar distribuida ($P_VSA_LogP_2$ en rango intermedio), (ii) preservar ramificación/modularidad que no penalice el acceso estérico ($PW3$ intermedio), y (iii) favorecer motivos D–P a distancias topológicas cortas ($CATS2D_02_DP$) para habilitar puentes de H y posibles contactos catión– π . Estas pautas son coherentes con hallazgos estructurales de inhibidores fenólicos y chalconas reportados en la literatura (Sayed et al., 2021; Azad et al., 2023; Mohaya et al., 2024). Asimismo, el AD por

similitud acota el espacio químico viable, útil para bibliotecas virtuales y priorización (Goya & Ortiz, 2020).

6.7 Relevancia farmacológica de los compuestos predichos

Desde la perspectiva clínica y cosmética, modular tirosinasa tiene impacto en hiperpigmentaciones y en el pardeamiento de alimentos (Fernandes et al., 2023; Urrialde et al., 2022). Glabridin y 4-Propylresorcinol poseen perfiles de actividad melanogénica reportados, lo que facilita su transición a estudios in vitro dirigidos a potencia (IC_{50}), selectividad y estabilidad. La naturaleza fenólica de ambos candidatos demanda evaluación paralela de fotolabilidad y citotoxicidad, aspectos ya subrayados para inhibidores clásicos (Zambrano & Rodríguez, 2022). La incorporación temprana de predicciones ADME-Tox puede evitar callejones sin salida en etapas preclínicas (Azad et al., 2023; Mohaya et al., 2024).

6.8 Limitaciones y perspectivas futuras

Limitaciones. (i) Heterogeneidad experimental del IC_{50} en la literatura (medios, pH, sustrato) introduce variabilidad; aunque se aplicaron media aritmética y test de Dixon, persiste incertidumbre en etiquetas, típica en QSAR retrospectivo (Fourches et al., 2010). (ii) La mezcla de representaciones (descriptores continuos y huellas) complica la comparación absoluta entre modelos. (iii) El AD basado en Tanimoto puede subestimar similitud para series con transformaciones sutiles 3D.

Perspectivas. (i) Incorporar calibración probabilística y curvas de confiabilidad para probabilidades de clase (Fuadah et al., 2024). (ii) Explorar consensos entre N3, RF y AdaBoost para ganar robustez (American Chemical Society, 2024). (iii) Integrar docking guiado por metales y dinámica molecular en un ciclo activo de diseño**→predicción→**síntesis virtual (Azad et al., 2023; Benavides et al., 2020). (iv) Evaluar métodos de selección alternativos (Boruta/ReliefF) y verificar colinealidad ($VIF < 5$) entre descriptores EA. (v) Ampliar la base con mediciones uniformes y series con bioisostería sistemática para mejorar extrapolación.

7. Conclusiones y Recomendaciones

7.1 Conclusiones

- Se aplicaron métodos de química computacional avanzados, integrando el modelado QSAR con técnicas de aprendizaje automático, para establecer relaciones cuantitativas entre la estructura molecular y la actividad biológica de un conjunto de compuestos. Esta aproximación permitió construir modelos predictivos confiables para la identificación de inhibidores de tirosinasa, demostrando la efectividad de los algoritmos no lineales frente a enfoques tradicionales.
- Se actualizó una base de datos curada de 548 moléculas, partiendo de una recopilación inicial de 700 compuestos. Se eliminaron duplicados y valores atípicos mediante criterios estadísticos, y se asignaron descriptores moleculares calculados con alvaDesc, abarcando propiedades estructurales, electrónicas, topológicas y tridimensionales, lo cual enriqueció sustancialmente la calidad de la información química disponible para el modelado.

- Se generaron modelos matemáticos mediante clasificadores como Random Forest, AdaBoost, kNN y N3, utilizando tanto descriptores moleculares reducidos como huellas moleculares (ECFPs, PFPs, MACCS). Los modelos presentaron un desempeño sobresaliente, con precisiones superiores al 85 % en el conjunto de prueba. En especial, el modelo ECFPs + N3 ($\alpha = 1.25$) demostró una capacidad predictiva destacada, validando su utilidad para discriminar compuestos con actividad inhibitoria.
- Se estimó la potencia inhibitoria de 17 moléculas nuevas, permitiendo identificar a Glabridin y 4-Propylresorcinol como candidatos activos dentro del dominio de aplicabilidad del modelo. Esta predicción refuerza la aplicabilidad del modelo como herramienta de cribado virtual en las primeras etapas del descubrimiento de fármacos despigmentantes y antienzimáticos.
- Se investigaron métodos alternativos para la reducción de variables, aplicando con éxito el algoritmo Variable-Wise Sequential Projection (V-WSP), el cual permitió disminuir la dimensionalidad del conjunto de datos sin afectar la precisión de los modelos. Esta estrategia contribuyó a optimizar la interpretación química y la eficiencia computacional del proceso de modelado.

7.2 Recomendaciones

- Se recomienda profundizar en la validación externa de los modelos desarrollados mediante ensayos experimentales in vitro o in silico complementarios, tales como estudios de acoplamiento molecular (docking) o simulaciones de dinámica molecular, con el fin de reforzar la

evidencia sobre la afinidad y modo de interacción de los compuestos predichos como inhibidores de tirosinasa.

- Se sugiere ampliar la base de datos de entrenamiento, incorporando compuestos con estructuras novedosas y actividades medidas bajo condiciones experimentales uniformes, para mejorar la diversidad química y reducir la variabilidad asociada a diferencias metodológicas en la medición del IC₅₀.
- Se recomienda aplicar los modelos desarrollados a bibliotecas moleculares más extensas, incluyendo bases de datos naturales y sintéticas, con el objetivo de realizar campañas de cribado virtual que permitan identificar nuevos candidatos bioactivos de forma eficiente, acelerando así el proceso de descubrimiento de ingredientes funcionales en áreas cosméticas y farmacéuticas.
- Se aconseja complementar los modelos con predicciones ADME-Tox (absorción, distribución, metabolismo, excreción y toxicidad), con el fin de filtrar compuestos que, pese a su actividad enzimática, puedan presentar limitaciones en su uso terapéutico por problemas de biodisponibilidad o seguridad.
- Se propone explorar métodos adicionales de reducción de variables, como algoritmos de selección basados en importancia relativa (ReliefF, Boruta) o estrategias de embedding supervisado, que podrían optimizar aún más la interpretabilidad de los modelos sin comprometer su precisión predictiva.
- Se recomienda incorporar modelos de predicción multicategoría o regresión no lineal, que permitan estimar de forma continua la potencia

inhibitoria ($\log I_{C_{50}}$) en lugar de clasificarla en categorías discretas, aumentando así la resolución del modelo y su utilidad para decisiones de priorización molecular.

Referencias

- Abril, M. (2023). *Aplicación del aprendizaje supervisado para la predicción del gusto de diversas moléculas*. Cuenca: Universidad del Azuay. <https://dspace.uazuay.edu.ec/bitstream/datos/13219/1/18745.pdf>
- Aguilera, G. (2021). *Estudio Computacional de blancos moleculares relevantes en Vitiligo: Activación de los receptores CXCR3, y modulación Alostérica de la Tirosina*. Mexico : Universidad Autonoma de Queretaro. <https://ri-ng.uaq.mx/bitstream/123456789/3440/1/RI006481.pdf>
- American Chemical Society. (2024). *QSAR Modeling of TBK1 Kinase Inhibitors for Enhanced Drug Discovery*. <https://pubs.acs.org/doi/10.1021/acs.jcim.4c00864>
- Arbito, J., & Guamán, A. (2024). *EVALUACIÓN DEL POTENCIAL CARDIOTÓXICO DE LOS FITOQUÍMICOS DE NERIUM OLEANDER (ADELFA) EMPLEANDO ANÁLISIS BASADOS EN MODELOS QSAR*. Cuenca: Universidad Católica de Cuenca. <https://dspace.ucacue.edu.ec/server/api/core/bitstreams/63d2ccf3-49e1-4c1c-9fc5-d7a7f77a6b98/content>
- Azad, I., Khan, T., Ahmad, N., Khan, A. R., & Akhter, Y. (2023). Updates on drug designing approach through computational strategies. *future-science OA* , 40. <https://doi.org/https://pmc.ncbi.nlm.nih.gov/articles/PMC10167725/pdf/fsoa-09-862.pdf>

- Benavides, I., Peña, J., Cutiño, L. M., & Morris, H. J. (2020). Modelación por homología de la tirosinasa hipotética en pleurotus ostreatus . *Cubana Quím*, 32(2), 21. <https://doi.org/http://scielo.sld.cu/pdf/ind/v32n2/2224-5421-ind-32-02-172.pdf>
- Bencomo, A., Álvarez, Y. M., Sánchez, A. T., Sablón, M., & Rodríguez, C. (2022). Diseño relacional de fármacos contra la enfermedad de Alzheimer: estudios QSAR y de acoplamiento molecular. *Cubana Quím*, 34(3), 34. <https://doi.org/http://scielo.sld.cu/pdf/ind/v34n3/2224-5421-ind-34-03-369.pdf>
- Benzama, K. (2021). *FORMULACIÓN Y CARACTERIZACIÓN DE LAS PROPIEDADES REOLÓGICAS Y DIGESTIVAS DE UNA CREMA DE CHAMPIÑONES ENRIQUECIDA EN PROTEÍNA Y CON TEXTURA MODIFICADA*. Valencia: Universida de Valencia . <https://riunet.upv.es/server/api/core/bitstreams/e524ea92-bd2a-4d8a-ac5f-ceb7571779bc/content>
- Botella, M. B., Quintas, P., & Gonzales, R. (2023). 1008 DETERMINACIÓN ENANTIOMÉRICA DE TIROSINA Y TRIPTÓFANO EN MUESTRAS DE CALABAZA. *CyTAL*, 3. https://doi.org/https://ri.conicet.gov.ar/bitstream/handle/11336/250209/CONICET_Digital_Nro.07dd37ed-cc88-4982-ac40-d9b1afb3f877_C.pdf?sequence=5&isAllowed=y
- Burón, M. (2021). *Efecto fotoprotector del medio acondicionado por las hASCs pre-acondicionadas con H2O2 sobre fibroblastos dérmicos sometidos a la radiación UV-B*. Leioa: Universidad del país Vasco.

https://addi.ehu.es/bitstream/handle/10810/54164/TESIS_BURON_AIZPI_RI_MARIA.pdf?sequence=1&isAllowed=y

Caballero, E. (2024). *EVALUACIÓN DE LA FEOMELANINA COMO FOTOSENSIBILIZADOR EN LA TERAPIA FOTODINÁMICA CONTRA CÉLULAS DE MELANOMA*. Tonantzintla, Puebla: INAOE. https://inaoe.repositorioinstitucional.mx/jspui/bitstream/1009/2578/1/CABALLEROLE_MTB.pdf

Caicedo, L. (2024). *Inhibición en el pardeamiento de bananas (Cavendish) de exportación causada por la actividad de la enzima polifenol oxidasa utilizando la combinación de métodos físicos y químicos*. Milagros: UNEMI. <http://repositorio.unemi.edu.ec/bitstream/123456789/7194/1/CAICEDO%20IBAN%CC%83EZ%20%20LEANDRO%20AGUST%C3%8DN.pdf>

Calderón, V., Garza, D., Zaragoza, E., Reul, E., Velazco, G. G., Ochoa, H., & Lechuga, M. (2023). Epidemiology of alopecia areata in the Hispanic/Latinx community. *Centro dermatologico Pascua*, 5. <https://doi.org/https://www.medigraphic.com/pdfs/derma/cd-2024/cd242e.pdf>

Coutant, K., Magne, B., Ferland, K., Rodríguez, A. F., Chancy, O., Mitchell, A., . . . Landreville, S. (2024). Melanocitos en aplicaciones de medicina regenerativa y modelado de enfermedades. *Medicina Traslacional*(336), 22. <https://doi.org/https://translational-medicine.biomedcentral.com/articles/10.1186/s12967-024-05113-x#citeas>

- Faramarzi, S. (2022). New developments in regulatory QSAR modeling: a new QSAR model for predicting blood brain barrier permeability. *Frontiers in Pharmacology*. <https://www.fda.gov/drugs/regulatory-science-action/new-developments-regulatory-qsar-modeling-new-qsar-model-predicting-blood-brain-barrier-permeability>
- Fernandes, B., Cavaco, A., & Matamá, T. (2023). Una revisión exhaustiva de la pigmentación de los mamíferos: allanando el camino para la innovación en cosméticos que modifican el color del cabello. *Biology*, 2, 12. <https://doi.org/https://www.mdpi.com/2079-7737/12/2/290>
- Fuadah, Y. N., Pramudito, M. A., Firdaus, L., & Vanheusden, F. J. (2024). QSAR Classification Modeling Using Machine Learning with a Consensus-Based Approach for Multivariate Chemical Hazard End Points. *ACS OMEGA*, 13. https://doi.org/https://pubs.acs.org/doi/pdf/10.1021/acsomega.4c09356?ref=article_openPDF
- Goode, G., Aguayo, R., & Domínguez, L. (2019). RELACIONES CUANTITATIVAS ESTRUCTURA/ACTIVIDAD/PROPIEDAD EN DOS DIMENSIONES EMPLEANDO EL PROGRAMA R. *Educacion Quimica*, 30(2), 14. <https://doi.org/https://www.scielo.org.mx/pdf/eq/v30n2/0187-893X-eq-30-02-27.pdf>
- Goya, J., & Ortiz, J. (2020). Revisión de los modelos computacionales que relacionan la estructura química con la disrupción del sistema endocrino. *Centro Europeo de Empresas Innovadoras*, 37, 14. <https://doi.org/https://rev.aetox.es/wp/wp-content/uploads/2020/06/vol-37.1-1-75-59-72.pdf>

- Hernández, B. (2023). *Estudio Computacional del canal de calcio dependiente de voltaje de tipo T (CACNA1G) y su interacción con anticonvulsivos*. Queretaro: Universidad Autónoma de Querétaro. <https://ri-ng.uaq.mx/bitstream/123456789/9056/1/FQMAC-309155.pdf>
- Ivanov, J. M., Tenchov, R., Ralhan, K., Iyer, K. A., & Agarwal, S. (2024). In Silico Insights: QSAR Modeling of TBK1 Kinase Inhibitors for Enhanced Drug Discovery. *Chemical Information and Modeling*, 15. https://doi.org/https://pubs.acs.org/doi/pdf/10.1021/acs.jcim.4c00864?ref=article_openPDF
- Jiménez, V. (2023). *USO DE HERRAMIENTAS QUIMIOINFORMÁTICAS PARA LA PREDICCIÓN DE POTENCIALES ACTIVIDADES BIOLÓGICAS DE MOLÉCULAS CON NÚCLEO CHALCONA*. Mexico: Universidad Autónoma Metropolitana. <https://repositorio.xoc.uam.mx/jspui/retrieve/6f637650-2291-4780-8ab0-f1eee5ade73c/250995.pdf>
- Jiyan, S. (2024). The biochemistry of melanogenesis: an insight into the function and mechanism of melanogenesis-related proteins. *Sec. Cellular Biochemistry*, 11. <https://www.frontiersin.org/journals/molecular-biosciences/articles/10.3389/fmolb.2024.1440187/full>
- Maddaleno, A. S., Camargo, J., Mitjans, M., & Vinardell, M. P. (2021). Melanogenesis and Melasma Treatment. *MDPI*, 8. <https://doi.org/https://www.mdpi.com/2079-9284/8/3/82>
- Minerva, F. F., Amelia, R., Maidawilis, M., Rita, R. S., & Alimuddin, A. (2025). Phenolic Content, Antioxidant Activity, and Tyrosinase Inhibitor Potential

for both Skin and Physical Health: Test of Extract, Skin Content and Content of Garcinia Cowa Fruit. *Retos*, 62, 7. <https://doi.org/https://www.proquest.com/docview/3134051732?fromopenview=true&pq-origsite=gscholar&sourcetype=Scholarly%20Journals>

Miralrio, A., Beltrán, C., Abundis, A., & Villada, M. (2024). MODELADO QSAR DE LA EFICIENCIA EN LA INHIBICIÓN DE LA CORROSIÓN MEDIANTE FÁRMACOS Y FITOQUÍMICOS. *CIMAT*, 164. <https://doi.org/https://supercomputo.cimat.mx/sites/default/files/supercomputo/LibroProySupercomputo.pdf#page=36>

Mohaya, M., Mesut, B., Kurt, A., & Selim, Y. (2024). In silico approaches which are used in pharmacy. *Applied Pharmaceutical Science*, 14, 15. https://doi.org/https://japsonline.com/admin/php/uploads/4213_pdf.pdf

Morgado, D., Piquero, J., Trullas, C., & Granger, C. (2023). Fotoprotección en la piel oscura Photoprotection in dark-skinned phenotypes. *ELSEVIER*, 38(1), 7. <https://doi.org/https://pdf.sciencedirectassets.com/277686/1-s2.0-S0213925122X00129/1-s2.0-S0213925122000776/main.pdf?X-Amz-Security-Token=IQoJb3JpZ2luX2VjEFwaCXVzLWVhc3QtMSJGMEQCIAcQgc0HyNMYBAJAHDHJx7hWJMetPnrVjIguF8c45e7ZAIArhP1MSA05m1VHjBI6iktnM36RpPhT0p3EqLsa7iB%2Bci>

Niazi, S. (2025). Protein Catalysis Through Structural Dynamics: A Comprehensive Analysis of Energy Conversion in Enzymatic Systems and Its Computational Limitations. *Pharmaceuticals*, 18. <https://doi.org/https://www.mdpi.com/1424-8247/18/7/951>

- Ponphaiboon, J., Krongrawa, W., Aung, W. W., Chinatangkul, N., Limmatvapirat, S., & Limmatvapirat, C. (2023). Advances in Natural Product Extraction Techniques, Electrospun Fiber Fabrication, and the Integration of Experimental Design: A Comprehensive Review. *Molecules* , 23. <https://doi.org/https://doi.org/10.3390/molecules28135163>
- Reyes, R., Barajas, M., Flores, J. d., Hernández, G., & Xoca, L. Á. (2023). Biopéptidos derivados de los pseudocereales: Amaranto, Quinoa, Chía y Trigo sarraceno. *TIP*, 26(1), 27. <https://doi.org/https://www.scielo.org.mx/pdf/tip/v26/1405-888X-tip-26-e616.pdf>
- Rodrigues, A., Iost, J., Moraes, M., Goulart, T., Silva, C., Ferreira, V., . . . Figueiredo, H. S. (2022). Biomarcadores y preeclampsia. *Acervo Médico*, 8. <https://doi.org/https://acervomais.com.br/index.php/medico/article/view/10389/6239>
- Romanos, I. A. (2024). *Impacto del procesado y de la fermentación con Pleurotus ostreatus sobre la composición nutricional, contenido de anti-nutrientes y digestibilidad de la judía de herradura (Phaseolus vulgaris L.)*. Valencia: UNIVERSITAT POLITÈCNICA DE VALÈNCIA. <https://riunet.upv.es/server/api/core/bitstreams/1de0ddc0-3a63-48ae-96cb-d17b3e3aebf2/content>
- Sabando, V. (2023). *Estrategias de aprendizaje profundo aplicadas al descubrimiento de fármacos : representación molecular, modelado de bioactividad y analítica visual para cribado virtual*. Argentina: Universidad

del Sur.
https://repositoriodigital.uns.edu.ar/bitstream/handle/123456789/6737/SABANDO%20V_TESIS.pdf?sequence=1&isAllowed=y

Sanchez, A. (2024). *Marcadores genéticos de susceptibilidad y pronóstico del melanoma cutáneo*. Vasco: Universidad del País Vasco.
<https://addi.ehu.eus/handle/10810/70910>

Santana, H. G. (2022). Biomarcador FLT3-ITD: perspectivas y retos. *Cubana de Hematología, Inmunología y Hemoterapia*, 38(1), 11.
<https://doi.org/http://scielo.sld.cu/pdf/hih/v38n1/1561-2996-hih-38-01-e1484.pdf>

Sayed, A., Maamoun, H. S., Rabie, G. H., Shaker, I., Alaidaroos, B. A., Ali, M. G., & Rady, A. M. (2021). Microbial Tyrosinase: Biochemical, Molecular Properties and Pharmaceutical Applications . *Biomedical and Pharmacology* , 569.
<https://doi.org/https://biomedpharmajournal.org/vol14no3/microbial-tyrosinase-biochemical-molecular-properties-and-pharmaceutical-applications/>

Singh, D., & Chaturvedi, V. (2025). The amazing world of biological pigments: A review on microbial melanins. *Science Direct*, 237.
<https://www.sciencedirect.com/science/article/abs/pii/S0143720825000816>

Singh, D., Deepshikha, Chaturvedi, V., & Verma, P. (2025). El asombroso mundo de los pigmentos biológicos: una revisión sobre las melaninas microbianas. *ELSEVIER*, 237, 112.

<https://doi.org/https://www.sciencedirect.com/science/article/abs/pii/S0143720825000816>

Ugaldea, M. J., b, A. C., Fernándezc, M. M., & Tamayo, E. (2024). Valor del biomarcador tirosina quinasa 1 soluble tipo fms (sFLT-1) en el diagnóstico y pronóstico de la sepsis. *ELSEVIER*, 163, 8. <https://doi.org/https://pdf.sciencedirectassets.com/277687/1-s2.0-S0025775324X00147/1-s2.0-S0025775324002884/main.pdf?X-Amz-Security-Token=IQoJb3JpZ2luX2VjEFwaCXVzLWVhc3QtMSJIMEYCIQC3ZEgBf42bstpQrc%2FvUb7DtNYqliJQqYfusrJ0JT9nGAIhAKnd6FoifSq27%2FgWo54INNtaDyWbiYE8AO4oZqef>

Urrialde, R., Gómez, A., Pintos, B., Gómez, M. A., & Cifuentes, B. (2022). Compuestos bioactivos de origen vegetal: desarrollo de nuevos alimentos . *ARAN*, 4. <https://doi.org/https://scielo.isciii.es/pdf/nh/v39nspe3/0212-1611-nh-39-nspe3-8.pdf>

Vargas, J., Ruíz, M., Vera, B. C., Briano, M. A., & Barrientos, L. (2023). OBTENCIÓN DE COMPUESTOS BIOACTIVOS DE PLANTAS MEDICINALES . *CNIC PRONAT*, 54, 12. <https://doi.org/https://www.redalyc.org/journal/1816/181676104010/html/>

Wang, F., Ma, W., Ma, W., Fan, D., Fan, D., Hu, J., . . . A, X. (2024). The biochemistry of melanogenesis: an insight into the function and mechanism of melanogenesis-related proteins. *Frontiers*, 11, 15. <https://doi.org/https://www.readcube.com/articles/10.3389/fmolb.2024.1440187>

Zambrano, N., & Rodríguez, A. (2022). NUEVA ERA COSMETOLÓGICA: MICROORGANISMOS Y SUS METABOLITOS EN LA FORMULACIÓN DE FOTOPROTECTORES Y ACLARADORES. *REDIUPAC*, 14. <https://doi.org/https://rdicuap.buap.mx/index.php/rdicuap/article/view/683/896>

Zou, H., Xiao, Q., Li, G., Wei, X., Tian, X., Zhu, L., . . . Li, M. (2025). Revisiting the advancements in plant polyphenol oxidases research. *Elsevier*, 11. <https://doi.org/https://pdf.sciencedirectassets.com/271177/1-s2.0-S0304423825X00029/1-s2.0-S0304423825000111/main.pdf?X-Amz-Security-Token=IQoJb3JpZ2luX2VjEFsaCXVzLWVhc3QtMSJHMEUCIQD1eAyoMUgqG%2FFUO1JXedgkZ6Aeg75iBvofBYBQIILxglgJeHm9oX%2F0uZDIAjzJeFX7IkBqeQEQ47BPdZC0dIU>

Anexos

Anexo 1. Métricas de desempeño (precisión, sensibilidad y especificidad) de los modelos de clasificación con descriptores completos

Modelos de clasificación	de Clase	Calibración			Validación Cruzada			Predicción			
		Pr	Sn	Sp	Pr	Sn	Sp	Pr	Sn	Sp	
Descriptores completos	PLSDA	Activo	0,746	0,758	0,871	0,746	0,758	0,871	0,700	0,778	0,836
		Inactivo	0,878	0,871	0,758	0,878	0,871	0,758	0,885	0,837	0,778
	KNN	Activo	0,741	0,758	0,867	0,739	0,773	0,8633	0,690	0,741	0,836
		Inactivo	0,758	0,867	0,758	0,884	0,863	0,773	0,868	0,836	0,741
	RF	Activo	1	1	1	0,798	0,680	0,914	0,786	0,815	0,891
		Inactivo	1	1	1	0,851	0,914	0,680	0,907	0,891	0,815
	AdaBoost	Activo	1	1	1	0,856	0,742	0,938	0,764	0,778	0,882
		Inactivo	1	1	1	0,879	0,938	0,742	0,890	0,882	0,778
VWS	RF	Activo	1	1	1	0,821	0,719	0,922	0,786	0,815	0,891

	Inactivo	1	1	1	0,868	0,922	0,719	0,907	0,891	0,815
AdaBoost	Activo	1	1	1	0,826	0,742	0,922	0,746	0,759	0,873
	Inactivo	1	1	1	0,877	0,922	0,742	0,881	0,873	0,759

Anexo 2. Métricas de desempeño de modelos de clasificación basados en huellas moleculares (fingerprints)

Modelos de clasificación	de Clase	Calibración			Validación Cruzada			Predicción			
		Pr	Sn	Sp	Pr	Sn	Sp	Pr	Sn	Sp	
ECFPs	kNN	Activo	0,754	0,719	0,883	0,764	0,734	0,887	0,746	0,759	0,873
		Inactivo	0,863	0,883	0,719	0,870	0,887	0,734	0,881	0,873	0,759
	N3	Activo	0,637	0,836	0,762	0,640	0,844	0,762	0,646	0,944	0,746
		Inactivo	0,901	0,762	0,826	0,907	0,762	0,844	0,965	0,746	0,944
	BNN	Activo	0,738	0,727	0,871	0,762	0,750	0,883	0,707	0,759	0,845
		Inactivo	0,864	0,871	0,727	0,876	0,883	0,750	0,877	0,845	0,759
	kNN	Activo	0,776	0,758	0,891	0,787	0,750	0,898	0,759	0,815	0,873

		Inactivo	0,880	0,891	0,758	0,878	0,898	0,750	0,906	0,873	0,815
	N3	Activo	0,686	0,820	0,813	0,699	0,836	0,820	0,694	0,926	0,800
		Inactivo	0,900	0,813	0,820	0,909	0,820	0,836	0,957	0,800	0,926
	BNN	Activo	0,773	0,773	0,887	0,767	0,773	0,883	0,714	0,833	0,836
		Inactivo	0,887	0,887	0,773	0,886	0,883	0,773	0,911	0,836	0,833
	kNN	Activo	0,795	0,758	0,902	0,817	0,766	0,914	0,782	0,796	0,890
		Inactivo	0,882	0,902	0,758	0,886	0,914	0,766	0,899	0,890	0,796
	N3	Activo	0,638	0,883	0,750	0,663	0,891	0,773	0,6500	0,963	0,745
		Inactivo	0,928	0,750	0,883	0,934	0,773	0,891	0,976	0,745	0,963
	BNN	Activo	0,810	0,797	0,906	0,795	0,790	0,898	0,741	0,796	0,863
		Inactivo	0,900	0,906	0,797	0,895	0,898	0,790	0,896	0,863	0,796
	kNN	Activo	0,754	0,672	0,891	0,785	0,711	0,902	0,719	0,759	0,854
		Inactivo	0,844	0,891	0,672	0,862	0,902	0,711	0,879	0,854	0,759
	N3	Activo	0,673	0,836	0,797	0,673	0,836	0,797	0,681	0,907	0,791

	Inactivo	0,907	0,797	0,836	0,907	0,797	0,836	0,946	0,791	0,907
BNN	Activo	0,759	0,734	0,883	0,754	0,742	0,879	0,710	0,815	0,836
	Inactivo	0,869	0,883	0,734	0,871	0,879	0,742	0,902	0,836	0,815

Anexo 3. Base de datos de las 565 moléculas utilizadas para el estudio QSAR

Nº	Nombre	IC 50	Unidad	CAS number	Referencia
1	Kojic acid	27.0	uM	501-30-4	Khatib 2005
2	N-aryl S-alkylthiocarbamate derivative 1	100.0	uM	nd	Khan 2007
3	N-aryl S-alkylthiocarbamate derivative 2	100.0	uM	nd	Khan 2007
4	N-aryl S-alkylthiocarbamate derivative 3	38.1	uM	nd	Khan 2007
5	N-aryl S-alkylthiocarbamate derivative 4	18.4	uM	nd	Khan 2007
6	N-aryl S-alkylthiocarbamate derivative 5	7.3	uM	nd	Khan 2007
7	N-aryl S-alkylthiocarbamate derivative 6	46.6	uM	nd	Khan 2007
8	N-aryl S-alkylthiocarbamate derivative 7	48.5	uM	nd	Khan 2007
9	N-aryl S-alkylthiocarbamate derivative 8	87.5	uM	nd	Khan 2007
10	N-aryl S-alkylthiocarbamate derivative 9	47.4	uM	nd	Khan 2007
11	N-aryl S-alkylthiocarbamate derivative 10	100.0	uM	nd	Khan 2007
12	N-aryl S-alkylthiocarbamate derivative 11	100.0	uM	nd	Khan 2007
13	Diethylthiocarbamate	61.2	uM	392-74-5	Khan 2007
14	Oxyresveratrol	27.4	uM	4721-07-7	Khan 2007
15	N,N-unsubstituted selenourea derivative S4	170.0	uM	nd	Khan 2007
16	Benzaldehyde derivative 1	367.0	uM	100-52-7	Khan 2007
17	Benzaldehyde analog B2	3772.1	uM	nd	Khan 2007
18	Benzaldehyde analog B3	721.4	uM	nd	Khan 2007
19	Benzaldehyde analog B4	144.5	uM	nd	Khan 2007
20	Benzaldehyde analog B5 (2-hydroxy-4-methoxybenzaldehyde)	1115.0	uM	nd	Khan 2007
21	Benzaldehyde analog B6 (2-hydroxy-4-isopropylbenzaldehyde)	2.3	uM	nd	Khan 2007
22	Chalcone derivative C1	68.8	uM	nd	Khan 2007
23	Chalcone derivative C2	53.8	uM	nd	Khan 2007
24	Chalcone derivative C4	90.0	uM	nd	Khan 2007
25	4-Hydroxychalcone	46.4	uM	nd	Khan 2007
26	2',4',4'-Trihydroxychalcon	37.6	uM	nd	Khan 2007
27	4,4' Biphenyl	1.9	uM	92-52-4	Khan 2007
28	1,3-selenazol-4-one derivative A	333.2	uM	nd	Khan 2007
29	1,3-selenazol-4-one derivative B	384.3	uM	nd	Khan 2007
30	1,3-selenazol-4-one derivative C	500.0	uM	nd	Khan 2007
31	1,3-selenazol-4-one derivative D	478.1	uM	nd	Khan 2007
32	1,3-selenazol-4-one derivative E	498.0	uM	nd	Khan 2007
33	1,3-selenazol-4-one derivative F	500.0	uM	nd	Khan 2007
34	Sildenafil analog 14	20.0	uM	nd	Khan 2005a; 2007
35	Sildenafil analog 15	8.7	uM	nd	Khan 2005a; 2007
36	Sildenafil analog 16	54.4	uM	nd	Khan 2005a; 2007
37	Sildenafil analog 17	3.5	uM	nd	Khan 2005a; 2007
38	Sildenafil analog 6	8.8	uM	nd	Khan, 2005b; 2007
39	Sildenafil analog 7	16.4	uM	nd	Khan, 2005b; 2007
40	Sildenafil analog 8	10.4	uM	nd	Khan, 2005b; 2007
41	Sildenafil analog 9	21.4	uM	nd	Khan, 2005b; 2007
42	Sildenafil analog 10	3.6	uM	nd	Khan, 2005b; 2007

43	Sildenafil analog 11	9.7	uM	nd	Khan, 2005b; 2007
44	Sildenafil analog 12	2.2	uM	nd	Khan, 2005b; 2007
45	Sildenafil analog 13	12.3	uM	nd	Khan, 2005b; 2007
46	Sildenafil analog 14'	19.4	uM	nd	Khan, 2005b; 2007
47	L-mimosine	3.7	uM	500-44-7	Khan, 2005 (a,b,c); 2007
48	Oxadiazole derivative 3a	6.9	uM	nd	Khan 2005c; 2007
49	Oxadiazole derivative 3b	3.2	uM	nd	Khan 2005c; 2007
50	Oxadiazole derivative 3c	5.2	uM	nd	Khan 2005c; 2007
51	Oxadiazole derivative 3d	6.6	uM	nd	Khan 2005c; 2007
52	Oxadiazole derivative 3e	5.0	uM	nd	Khan 2005c; 2007
53	Oxadiazole derivative 3f	3.3	uM	nd	Khan 2005c; 2007
54	Oxadiazole derivative 3g	4.2	uM	nd	Khan 2005c; 2007
55	Oxadiazole derivative 3h	5.6	uM	nd	Khan 2005c; 2007
56	Oxadiazole derivative 3i	5.1	uM	nd	Khan 2005c; 2007
57	Oxadiazole derivative 3j	8.4	uM	nd	Khan 2005c; 2007
58	Oxadiazole derivative 3l	3.2	uM	nd	Khan 2005c; 2007
59	Oxadiazole derivative 4m	7.8	uM	nd	Khan 2005c; 2007
60	4 vinylbenzoic acid	0.3	uM	1075-49-6	Khan 2007
61	4 vinylbenzaldehyde	23.0	uM	nd	Khan 2007
62	3,4-Dihydroxybenzonitrile	45.0	uM	17345-61-8	Khatib 2005
63	Steppogenin	7.5	uM	56486-94-3	Nguyen 2016
64	Coumarin derivative 5	2070.0	uM	nd	Matos 2011
65	Coumarin derivative 6	5000.0	uM	nd	Matos 2011
66	Coumarin derivative 7	5000.0	uM	nd	Matos 2011
67	Coumarin derivative 8	1360.0	uM	nd	Matos 2011
68	Coumarin derivative 9	5000.0	uM	nd	Matos 2011
69	Coumarin derivative 10	302.0	uM	nd	Matos 2011
70	Coumarin derivative 11	1300.0	uM	nd	Matos 2011
71	Coumarin derivative 12	215.0	uM	nd	Matos 2011
72	Umbelliferone	420.0	uM	93-35-6	Matos 2011
73	Esculetin	43.0	uM	305-01-1	Masamoto 2003
74	Cumarin	8100.0	uM	91-64-5	Masamoto 2003
75	Esculin	14000.0	uM	531-75-9	Masamoto 2003
76	Scopoletin	2600.0	uM	92-61-5	Masamoto 2003
77	N-Benzylbezamide derivative 1	1990.0	uM	nd	Cho 2006
78	N-Benzylbezamide derivative 2	780.0	uM	nd	Cho 2006
79	N-Benzylbezamide derivative 3	1180.0	uM	nd	Cho 2006
80	N-Benzylbezamide derivative 4	555.0	uM	nd	Cho 2006
81	N-Benzylbezamide derivative 5	17.0	uM	nd	Cho 2006
82	N-Benzylbezamide derivative 7	2000.0	uM	nd	Cho 2006
83	N-Benzylbezamide derivative 8	2000.0	uM	nd	Cho 2006
84	N-Benzylbezamide derivative 9	280.0	uM	nd	Cho 2006
85	N-Benzylbezamide derivative 10	11.0	uM	nd	Cho 2006
86	N-Benzylbezamide derivative 12	700.0	uM	nd	Cho 2006

87	N-Benzylbezamide derivative 13	710.0	uM	nd	Cho 2006
88	N-Benzylbezamide derivative 14	705.0	uM	nd	Cho 2006
89	N-Benzylbezamide derivative 15	2.2	uM	nd	Cho 2006
90	N-Benzylbezamide derivative 17	1660.0	uM	nd	Cho 2006
91	N-Benzylbezamide derivative 18	1820.0	uM	nd	Cho 2006
92	N-Benzylbezamide derivative 19	550.0	uM	nd	Cho 2006
93	N-Benzylbezamide derivative 20	29.0	uM	nd	Cho 2006
94	Hydroquinone	70.0	uM	123-31-9	Zolghadry 2019
95	3-aryl coumarin derivate 1	408.4	uM	nd	Zolghadry 2019
96	3-aryl coumarin derivate 2	1.1	uM	nd	Zolghadry 2019
97	3-aryl coumarin derivate 3	0.2	uM	nd	Zolghadry 2019
98	3-aryl coumarin derivate 4	0.2	uM	nd	Zolghadry 2019
99	3-aryl coumarin derivate 5	1000.0	uM	nd	Zolghadry 2019
100	3-aryl coumarin derivate 6	1000.0	uM	nd	Zolghadry 2019
101	3-aryl coumarin derivate 7	0.4	uM	nd	Zolghadry 2019
102	3-aryl coumarin derivate 8	31.8	uM	nd	Zolghadry 2019
103	3-aryl coumarin derivate 9	1000.0	uM	nd	Zolghadry 2019
104	3-aryl coumarin derivate 10	259.0	uM	nd	Zolghadry 2019
105	Umbelliferone analog 15	118.5	uM	nd	Zolghadry 2019
106	Umbelliferone analog 16	9.0	uM	nd	Zolghadry 2019
107	Oxindole-based chalcone derivate 1	85.3	uM	nd	Zolghadry 2019
108	Oxindole-based chalcone derivate 2	145.2	uM	nd	Zolghadry 2019
109	Oxindole-based chalcone derivate 3	241.9	uM	nd	Zolghadry 2019
110	Oxindole-based chalcone derivate 4	180.3	uM	nd	Zolghadry 2019
111	Oxindole-based chalcone derivate 5	59.7	uM	nd	Zolghadry 2019
112	Oxindole-based chalcone derivate 6	232.3	uM	nd	Zolghadry 2019
113	Oxindole-based chalcone derivate 7	99.1	uM	nd	Zolghadry 2019
114	Oxindole-based chalcone derivate 8	107.2	uM	nd	Zolghadry 2019
115	Resveratrol (3,5,4-trihydroxy-trans-stilbene)	113.9	uM	501-36-0	Zolghadry 2019
116	Resveratrol analog 2	44.9	uM	nd	Zolghadry 2019
117	Resveratrol analog 3	72.6	uM	nd	Zolghadry 2019
118	Resveratrol analog 4	160.1	uM	nd	Zolghadry 2019
119	Resveratrol analog 5	28.6	uM	nd	Zolghadry 2019
120	Resveratrol analog 6	49.5	uM	nd	Zolghadry 2019
121	Resveratrol analog 7	148.0	uM	nd	Zolghadry 2019
122	Resveratrol analog 8	61.7	uM	nd	Zolghadry 2019
123	Resveratrol analog 9	51.5	uM	nd	Zolghadry 2019
124	Resveratrol analog 10	17.2	uM	nd	Zolghadry 2019
125	Resveratrol analog 11	140.2	uM	nd	Zolghadry 2019
126	Resveratrol analog 12	417.7	uM	nd	Zolghadry 2019
127	Resveratrol analog 13	50.2	uM	nd	Zolghadry 2019
128	Resveratrol analog 14	36.3	uM	nd	Zolghadry 2019
129	Resveratrol analog 20	200.8	uM	nd	Zolghadry 2019
130	Resveratrol analog 21	17.9	uM	nd	Zolghadry 2019
131	Resveratrol analog 22	44.1	uM	nd	Zolghadry 2019
132	Resveratrol analog 23	49.1	uM	nd	Zolghadry 2019

133	4-(4-fluorobenzyl) piperidine derivative 1	159.2	uM	nd	Zolghadry 2019
134	4-(4-fluorobenzyl) piperidine derivative 2	111.0	uM	nd	Zolghadry 2019
135	4-(4-fluorobenzyl) piperidine derivative 3	27.9	uM	nd	Zolghadry 2019
136	4-(4-fluorobenzyl) piperidine derivative 4	39.6	uM	nd	Zolghadry 2019
137	4-(4-fluorobenzyl) piperidine derivative 5	83.0	uM	nd	Zolghadry 2019
138	Indole derivative 6	252.0	uM	nd	Zolghadry 2019
139	Indole derivative 7	197.5	uM	nd	Zolghadry 2019
140	Indole derivative 8	168.6	uM	nd	Zolghadry 2019
141	Indole derivative 9	133.0	uM	nd	Zolghadry 2019
142	Indole derivative 10	8.8	uM	nd	Zolghadry 2019
143	Indole derivative 11	14.2	uM	nd	Zolghadry 2019
144	Indole derivative 12	18.0	uM	nd	Zolghadry 2019
145	Indole derivative 13	7.6	uM	nd	Zolghadry 2019
146	Amine compound	286.8	uM	nd	Zolghadry 2019
147	N-ethyl	116.0	uM	nd	Zolghadry 2019
148	4,5,6,7-tetrahydro-2- [[[(phenylamino)thioxomethyl]amino]- benzo[b]thiophene-3-carboxylic acid derivative 22	53.3	uM	nd	Zolghadry 2019
149	4,5,6,7-tetrahydro-2- [[[(phenylamino)thioxomethyl]amino]- benzo[b]thiophene-3-carboxylic acid derivative 24	44.3	uM	nd	Zolghadry 2019
150	4,5,6,7-tetrahydro-2- [[[(phenylamino)thioxomethyl]amino]- benzo[b]thiophene-3-carboxylic acid derivative 25	44.7	uM	nd	Zolghadry 2019
151	2-(1,3,4-thiadiazol-2-yl) thio acetic acid derivative 26	27.7	uM	nd	Zolghadry 2019
152	2-(1,3,4-thiadiazol-2-yl) thio acetic acid derivative 27	54.4	uM	nd	Zolghadry 2019
153	2-(1,3,4-thiadiazol-2-yl) thio acetic acid derivative 28	20.9	uM	nd	Zolghadry 2019
154	2-(1,3,4-thiadiazol-2-yl) thio acetic acid derivative 29	6.1	uM	nd	Zolghadry 2019
155	1,3,4-thiadiazole derivative 1	49.2	uM	nd	Zolghadry 2019
156	1,3,4-thiadiazole derivative 2	0.8	uM	nd	Zolghadry 2019
157	1,3,4-thiadiazole derivative 3	4.8	uM	nd	Zolghadry 2019
158	1,3,4-thiadiazole derivative 4	2.5	uM	nd	Zolghadry 2019
159	1,3,4-thiadiazole derivative 5	3.6	uM	nd	Zolghadry 2019
160	1,3,4-thiadiazole derivative 6	1.7	uM	nd	Zolghadry 2019
161	1,3,4-thiadiazole derivative 7	2.0	uM	nd	Zolghadry 2019
162	1,3,4-thiadiazole derivative 8	1.4	uM	nd	Zolghadry 2019
163	1,3,4-thiadiazole derivative 9	0.7	uM	nd	Zolghadry 2019
164	1,3,4-thiadiazole derivative 10	0.7	uM	nd	Zolghadry 2019
165	1,3,4-thiadiazole derivative 11	5.4	uM	nd	Zolghadry 2019
166	1,3,4-thiadiazole derivative 12	0.8	uM	nd	Zolghadry 2019
167	1,3,4-thiadiazole derivative 13	0.3	uM	nd	Zolghadry 2019
168	1,3,4-thiadiazole derivative 14	0.0	uM	nd	Zolghadry 2019
169	1,3,4-thiadiazole derivative 15	0.5	uM	nd	Zolghadry 2019

170	1,3,4-thiadiazole derivative 16	0.9	uM	nd	Zolghadry 2019
171	1,3,4-thiadiazole derivative 17	247.4	uM	nd	Zolghadry 2019
172	Thiazolidinone derivative 1	21.6	uM	nd	Zolghadry 2019
173	Thiazolidinone derivative 2	0.6	uM	nd	Zolghadry 2019
174	Thiazolidinone derivative 3	2.4	uM	nd	Zolghadry 2019
175	Thiazolidinone derivative 4	4.4	uM	nd	Zolghadry 2019
176	Thiazolidinone derivative 5	342.5	uM	nd	Zolghadry 2019
177	Thiazolidinone derivative 6	7.7	uM	nd	Zolghadry 2019
178	Thiazolidinone derivative 7	500.0	uM	nd	Zolghadry 2019
179	Thiazolidinone derivative 8	2.8	uM	nd	Zolghadry 2019
180	kojic acid analog 1	5.8	uM	nd	Zolghadry 2019
181	kojic acid analog 2	10.2	uM	nd	Zolghadry 2019
182	kojic acid analog 3	7.6	uM	nd	Zolghadry 2019
183	kojic acid analog 4	5.9	uM	nd	Zolghadry 2019
184	kojic acid analog 5	12.5	uM	nd	Zolghadry 2019
185	kojic acid analog 6	1.4	uM	nd	Zolghadry 2019
186	kojic acid analog 7	3.8	uM	nd	Zolghadry 2019
187	kojic acid analog 8	2.5	uM	nd	Zolghadry 2019
188	kojic acid analog 9	1.7	uM	nd	Zolghadry 2019
189	kojic acid analog 10	1.5	uM	nd	Zolghadry 2019
190	kojic acid analog 11	5.2	uM	nd	Zolghadry 2019
191	kojic acid analog 12	8.5	uM	nd	Zolghadry 2019
192	kojic acid analog 13	9.6	uM	nd	Zolghadry 2019
193	kojic acid analog 14	17.5	uM	nd	Zolghadry 2019
194	Benzaldehyde derivative 2	70.5	uM	nd	Zolghadry 2019
195	Benzaldehyde derivative 7	318.0	uM	nd	Zolghadry 2019
196	Benzaldehyde derivative 8	112.0	uM	nd	Zolghadry 2019
197	Benzaldehyde derivative 9	237.4	uM	nd	Zolghadry 2019
198	Benzaldehyde derivative 10	133.4	uM	nd	Zolghadry 2019
199	Benzaldehyde derivative 12	3604.2	uM	nd	Zolghadry 2019
200	Benzaldehyde derivative 13	287.8	uM	nd	Zolghadry 2019
201	Benzaldehyde derivative 14	940.5	uM	nd	Zolghadry 2019
202	Benzaldehyde derivative 15	59.0	uM	nd	Zolghadry 2019
203	Thiosemicarbazone a1	18.0	uM	nd	Zhu 2013
204	Thiosemicarbazone b1	213.0	uM	nd	Zhu 2013
205	Thiosemicarbazone c1	108.0	uM	nd	Zhu 2013
206	Thiosemicarbazone d1	6.0	uM	nd	Zhu 2013
207	Thiosemicarbazone e1	65.0	uM	nd	Zhu 2013
208	Thiosemicarbazone a2	387.0	uM	nd	Zhu 2013
209	Thiosemicarbazone b2	1294.0	uM	nd	Zhu 2013
210	Thiosemicarbazone c2	703.0	uM	nd	Zhu 2013
211	Thiosemicarbazone d2	345.0	uM	nd	Zhu 2013
212	Thiosemicarbazone a2	1090.0	uM	nd	Zhu 2013
213	2'-hydroxy-4'-methoxy acetophenone	207.0	uM	552-41-0	Zhu 2013
214	Acetophenone	916.0	uM	98-86-2	Zhu 2013
215	2'-hydroxy acetophenone	3096.0	uM	118-93-4	Zhu 2013

216	2'-4'-dihydroxy acetophenone	1248.0	uM	nd	Zhu 2013
217	4'-methoxy acetophenone	1674.0	uM	100-06-1	Zhu 2013
218	4-dimethylaminobenzaldehyde	899.6	uM	100-10-7	Maghsoudi 2014
219	2,4-dihydroxybenzaldehyde	1131.1	uM	95-01-2	Maghsoudi 2014
220	3,4-dihydroxybenzaldehyde	251.1	uM	139-85-5	Maghsoudi 2014
221	Rifampicin	90.0	uM	13292-46-1	Chai 2017
222	Arbutin	105.4	uM	497-76-7	Chai 2017
223	Amoxicillin	900.0	uM	26787-78-0	Chai 2017
224	Rhodiolinin	250.0	uM	nd	Zhou 2016
225	Kaempferol	520.0	uM	520-18-3	Zhou 2016
226	Ascorbic acid	375.0	uM	50-81-7	Zhou 2016
227	Catechin	57.1	uM	154-23-4	Tang 2018
228	Epicatechin gallate	22.6	uM	863-03-6	Tang 2018
229	Epigallocatechin gallate	142.4	uM	989-51-5	Tang 2018
230	Thiosemicarbazone 1	210.2	uM	nd	Soares 2017
231	Thiosemicarbazone 2	177.8	uM	nd	Soares 2017
232	Thiosemicarbazone 3	181.0	uM	nd	Soares 2017
233	Thiosemicarbazone 4	220.7	uM	nd	Soares 2017
234	Thiosemicarbazone 5	820.0	uM	nd	Soares 2017
235	Thiosemicarbazone 6	710.1	uM	nd	Soares 2017
236	Thiosemicarbazone 7	420.9	uM	nd	Soares 2017
237	Pyridine 2 carboxaldehyde thiosemicarbazone	1190.0	uM	nd	Soares 2017
238	Methyl 4-(1-thiosemicarbazidoethyl)phenylacetate	0.3	uM	nd	Song 2017
239	Ethyl 4-(1-thiosemicarbazidoethyl)phenylacetate	0.4	uM	nd	Song 2017
240	Propyl 4-(1-thiosemicarbazidoethyl)phenylacetate	0.5	uM	nd	Song 2017
241	Butyl 4-(1-thiosemicarbazidoethyl)phenylacetate	0.4	uM	nd	Song 2017
242	Methyl 3-(4-(1-thiosemicarbazidoethyl)phenyl)propanoate	0.2	uM	nd	Song 2017
243	Ethyl 3-(4-(1-thiosemicarbazidoethyl)phenyl)propanoate	0.2	uM	nd	Song 2017
244	Propyl 3-(4-(1-thiosemicarbazidoethyl)phenyl)propanoate	0.1	uM	nd	Song 2017
245	3-(4-(1-thiosemicarbazidoethyl)phenyl)propanoic acid	0.4	uM	nd	Song 2017
246	Methyl 4-(4-(1-thiosemicarbazidoethyl)phenyl)butanoate	0.2	uM	nd	Song 2017
247	Ethyl 4-(4-(1-thiosemicarbazidoethyl)phenyl)butanoate	0.3	uM	nd	Song 2017
248	Butyl 4-(4-(1-thiosemicarbazidoethyl)phenyl)butanoate	5.9	uM	nd	Song 2017
249	2-(1-(40 -Acetoxy-[1,10 -biphenyl]-4-yl)ethylidene)hydrazine1-carbothio-amide	1.5	uM	nd	Song 2017
250	2-(1-(40 -iso-Butyroyloxy-[1,10 -biphenyl]-4-yl)ethylidene)hydrazine-1-carbo-thioamide	2.0	uM	nd	Song 2017
251	2-(1-(40 -Pivaloyloxy-[1,10 -biphenyl]-4-yl)ethylidene)hydrazine-1-carbo-thioamide	2.5	uM	nd	Song 2017

252	2-(1-(40 -Benzoyloxy-[1,10 -biphenyl]-4-yl)ethylidene)hydrazine-1-carbo-thioamide	50.0	uM	nd	Song 2017
253	2-(1-(4'-(4-Methoxybenzoyloxy)-[1,10 -biphenyl]-4-yl)ethylidene)-hydrazine-1-carbothioamide	50.0	uM	nd	Song 2017
254	2-(1-(40 -Hydroxy-[1,10 -biphenyl]-4-yl)ethylidene)hydrazine1-carbothio- amide	13.9	uM	nd	Song 2017
255	2-(1-(40 -Ethoxy-[1,10 -biphenyl]-4-yl)ethylidene)hydrazine-1- carbo-thioamide	17.6	uM	nd	Song 2017
256	2-(1-(40 -Propoxy-[1,10 -biphenyl]-4-yl)ethylidene)hydrazine1-carbo-thioamide	50.0	uM	nd	Song 2017
257	2-(1-(40 -Butoxy-[1,10 -biphenyl]-4-yl)ethylidene)hydrazine-1- carbothio- amide	50.0	uM	nd	Song 2017
258	2-(1-(40 -Pentoxy-[1,10 -biphenyl]-4-yl)ethylidene)hydrazine1-carbo- thioamide	50.0	uM	nd	Song 2017
259	2-(1-(4'-(Methoxycarbonylmethyl)-[1,10 -biphenyl]-4-yl)ethylidene)- hydrazine-1-carbothioamide	0.3	uM	nd	Song 2017
260	2-(1-(4'-(Ethoxycarbonylmethyl)-[1,10 -biphenyl]-4-yl)ethylidene)- hydrazine-1-carbothioamide	0.4	uM	nd	Song 2017
561	2-(1-(4'-(Propoxycarbonylmethyl)-[1,10-biphenyl]-4-yl)ethylidene)-hydrazine-1-carbothioamide	0.5	uM	nd	Song 2017
262	2-(1-(4'-(Butoxycarbonylmethyl)-[1,10-biphenyl]-4-yl)ethylidene) hydrazine-1-carbothioamide	0.6	uM	nd	Song 2017
263	2-(40-(1-(2-carbamothioylhydrazono)ethyl)-[1,10-biphenyl]-4-yl)acetates	2.0	uM	nd	Song 2017
264	5-Hydroxy-2-((4-(phenoxymethyl)-1H-1,2,3-triazol-1-yl)methyl)-4H-pyran-4-one	1.3	uM	nd	Ashooriha 2019
265	2-((4-((4-Fluorophenoxy)methyl)-1H-1,2,3-triazol-1-yl)methyl)-5-hydroxy-4H-pyran-4-one	0.9	uM	nd	Ashooriha 2019
266	2-((4-((4-Chlorophenoxy)methyl)-1H-1,2,3-triazol-1-yl)methyl)-5-hydroxy-4H-pyran-4-one	0.7	uM	nd	Ashooriha 2019
267	2-((4-((4-Bromophenoxy)methyl)-1H-1,2,3-triazol-1-yl)methyl)-5-hydroxy-4H-pyran-4-one	6.8	uM	nd	Ashooriha 2019
268	2-((4-((3-Fluorophenoxy)methyl)-1H-1,2,3-triazol-1-yl)methyl)-5-hydroxy-4H-pyran-4-one	1.1	uM	nd	Ashooriha 2019
269	2-((4-((3-Chlorophenoxy)methyl)-1H-1,2,3-triazol-1-yl)methyl)-5-hydroxy-4H-pyran-4-one	1.0	uM	nd	Ashooriha 2019
270	2-((4-((2,4-Dichlorophenoxy)methyl)-1H-1,2,3-triazol-1-yl)methyl)-5-hydroxy-4H-pyran-4-one	1.1	uM	nd	Ashooriha 2019
271	4-((1-((5-Hydroxy-4-oxo-4H-pyran-2-yl)methyl)-1H-1,2,3-triazol-4-yl)methoxy)benzotrile	6.3	uM	nd	Ashooriha 2019
272	5-Hydroxy-2-((4-((4-nitrophenoxy)methyl)-1H-1,2,3-triazol-1-yl)methyl)-4H-pyran-4-one	0.5	uM	nd	Ashooriha 2019
273	N-(4-((1-((5-Hydroxy-4-oxo-4H-pyran-2-yl)methyl)-1H-1,2,3-triazol-4-yl)methoxy)phenyl)acetamide	2.6	uM	nd	Ashooriha 2019

274	5-Hydroxy-2-((4-((4-(hydroxymethyl)phenoxy)methyl)-1H-1,2,3-triazol-1-yl)methyl)-4H-pyran-4-one	1.3	uM	nd	Ashooriha 2019
275	5-Hydroxy-2-((4-((4-methoxyphenoxy)methyl)-1H-1,2,3-triazol-1-yl)methyl)-4H-pyran-4-one	1.2	uM	nd	Ashooriha 2019
276	5-Hydroxy-2-((4-((3,4,5-trimethoxyphenoxy)methyl)-1H-1,2,3-triazol-1-yl)methyl)-4H-pyran-4-one	0.9	uM	nd	Ashooriha 2019
277	2-((4-((3,5-Dimethylphenoxy)methyl)-1H-1,2,3-triazol-1-yl)methyl)-5-hydroxy-4H-pyran-4-one	0.7	uM	nd	Ashooriha 2019
278	5-Hydroxy-2-((4-((naphthalen-1-yloxy)methyl)-1H-1,2,3-triazol-1-yl)methyl)-4H-pyran-4-one	0.1	uM	nd	Ashooriha 2019
279	5-Hydroxy-2-((4-((naphthalen-2-yloxy)methyl)-1H-1,2,3-triazol-1-yl)methyl)-4H-pyran-4-one	0.3	uM	nd	Ashooriha 2019
280	Benzaldehyde thiosemicarbazone derivative 2	0.4	uM	nd	Haldys 2019
281	Benzaldehyde thiosemicarbazone derivative 3	3.9	uM	nd	Haldys 2019
282	Benzaldehyde thiosemicarbazone derivative 7	0.3	uM	nd	Haldys 2019
283	Benzaldehyde thiosemicarbazone derivative 8	0.2	uM	nd	Haldys 2019
284	Benzaldehyde thiosemicarbazone derivative 9	6.5	uM	nd	Haldys 2019
285	Benzaldehyde thiosemicarbazone derivative 10	30.0	uM	nd	Haldys 2019
286	Benzaldehyde thiosemicarbazone derivative 11	40.0	uM	nd	Haldys 2019
287	Benzaldehyde thiosemicarbazone derivative 13	4.9	uM	nd	Haldys 2019
288	Benzaldehyde thiosemicarbazone derivative 14	50.0	uM	nd	Haldys 2019
289	2-chlorobenzaldehyde thiosemicarbazone 15	1.2	uM	nd	Haldys 2019
290	1-(3-Phenylallylidene)thiosemicarbazide	2.7	uM	nd	Liu 2009
291	4-dimethylaminobenzaldehyde thiosemicarbazone 18	2.0	uM	nd	Haldys 2019
292	4-dimethylaminobenzaldehyde-N-phenyl-thiosemicarbazone 19	0.8	uM	nd	Haldys 2019
293	4-O-substituted benzaldehyde thiosemicarbazone 20	0.5	uM	nd	Haldys 2019
294	4-O-substituted benzaldehyde thiosemicarbazone 21	0.9	uM	nd	Haldys 2019
295	4-O-substituted benzaldehyde thiosemicarbazone 22	0.3	uM	nd	Haldys 2019
296	4-O-substituted benzaldehyde thiosemicarbazone 23	0.4	uM	nd	Haldys 2019
297	4-O-substituted benzaldehyde thiosemicarbazone 24	0.8	uM	nd	Haldys 2019
298	4-O-substituted benzaldehyde thiosemicarbazone 25	0.4	uM	nd	Haldys 2019
299	4-O-substituted benzaldehyde thiosemicarbazone 26	0.9	uM	nd	Haldys 2019
300	4-O-substituted benzaldehyde thiosemicarbazone 28	0.9	uM	nd	Haldys 2019
301	4-O-substituted benzaldehyde thiosemicarbazone 29	3.0	uM	nd	Haldys 2019
302	4-O-substituted benzaldehyde thiosemicarbazone 30	1.3	uM	nd	Haldys 2019
303	4-O-substituted benzaldehyde thiosemicarbazone 31	2.2	uM	nd	Haldys 2019
304	4-O-substituted benzaldehyde thiosemicarbazone 32	1.3	uM	nd	Haldys 2019

305	4-O-substituted benzaldehyde thiosemicarbazone 33	1.1	uM	nd	Haldys 2019
306	4-O-substituted benzaldehyde thiosemicarbazone 34	1.0	uM	nd	Haldys 2019
307	4-O-substituted benzaldehyde thiosemicarbazone 35	1.5	uM	nd	Haldys 2019
308	Aminoacetophenone thiosemicarbazone derivative 41	2.6	uM	nd	Haldys 2019
309	Aminoacetophenone thiosemicarbazone derivative 42	0.5	uM	nd	Haldys 2019
310	Aminoacetophenone thiosemicarbazone derivative 43	1.3	uM	nd	Haldys 2019
311	Aminoacetophenone thiosemicarbazone derivative 44	0.4	uM	nd	Haldys 2019
312	Aminoacetophenone thiosemicarbazone derivative 45	0.4	uM	nd	Haldys 2019
313	Aminoacetophenone thiosemicarbazone derivative 46	0.6	uM	nd	Haldys 2019
314	Aminoacetophenone thiosemicarbazone derivative 47	0.8	uM	nd	Haldys 2019
315	Aminoacetophenone thiosemicarbazone derivative 48	0.5	uM	nd	Haldys 2019
316	Aminoacetophenone thiosemicarbazone derivative 49	0.8	uM	nd	Haldys 2019
317	Aminoacetophenone thiosemicarbazone derivative 50	0.4	uM	nd	Haldys 2019
318	Aminoacetophenone thiosemicarbazone derivative 51	0.5	uM	nd	Haldys 2019
319	Aminoacetophenone thiosemicarbazone derivative 52	0.3	uM	nd	Haldys 2019
320	Aminoacetophenone thiosemicarbazone derivative 53	0.6	uM	nd	Haldys 2019
321	Aminoacetophenone thiosemicarbazone derivative 54	1.1	uM	nd	Haldys 2019
322	Aminoacetophenone thiosemicarbazone derivative 55	0.6	uM	nd	Haldys 2019
323	4-alkoxy and 4-acyloxyacetophenone thiosemicarbazones derivate 56	0.5	uM	nd	Haldys 2019
324	4-alkoxy and 4-acyloxyacetophenone thiosemicarbazones derivate 57	0.2	uM	nd	Haldys 2019
325	4-alkoxy and 4-acyloxyacetophenone thiosemicarbazones derivate 58	0.2	uM	nd	Haldys 2019
326	4-alkoxy and 4-acyloxyacetophenone thiosemicarbazones derivate 59	0.3	uM	nd	Haldys 2019
327	4-alkoxy and 4-acyloxyacetophenone thiosemicarbazones derivate 60	0.3	uM	nd	Haldys 2019
328	4-alkoxy and 4-acyloxyacetophenone thiosemicarbazones derivate 61	0.4	uM	nd	Haldys 2019
329	4-alkoxy and 4-acyloxyacetophenone thiosemicarbazones derivate 62	0.5	uM	nd	Haldys 2019
330	4-alkoxy and 4-acyloxyacetophenone thiosemicarbazones derivate 63	0.8	uM	nd	Haldys 2019

331	4-alkoxy and 4-acyloxyacetophenone thiosemicarbazones derivate 64	0.9	uM	nd	Haldys 2019
332	4-alkoxy and 4-acyloxyacetophenone thiosemicarbazones derivate 65	0.4	uM	nd	Haldys 2019
333	4-alkoxy and 4-acyloxyacetophenone thiosemicarbazones derivate 66	0.2	uM	nd	Haldys 2019
334	4-alkoxy and 4-acyloxyacetophenone thiosemicarbazones derivate 67	0.4	uM	nd	Haldys 2019
335	4-alkoxy and 4-acyloxyacetophenone thiosemicarbazones derivate 68	0.8	uM	nd	Haldys 2019
336	4-alkoxy and 4-acyloxyacetophenone thiosemicarbazones derivate 69	0.5	uM	nd	Haldys 2019
337	4-alkoxy and 4-acyloxyacetophenone thiosemicarbazones derivate 70	0.3	uM	nd	Haldys 2019
338	4-alkoxy and 4-acyloxyacetophenone thiosemicarbazones derivate 71	0.1	uM	nd	Haldys 2019
339	4-alkoxy and 4-acyloxyacetophenone thiosemicarbazones derivate 72	0.3	uM	nd	Haldys 2019
340	4-alkoxy and 4-acyloxyacetophenone thiosemicarbazones derivate 73	0.2	uM	nd	Haldys 2019
341	4-alkoxy and 4-acyloxyacetophenone thiosemicarbazones derivate 74	0.6	uM	nd	Haldys 2019
342	4-alkoxy and 4-acyloxyacetophenone thiosemicarbazones derivate 75	0.7	uM	nd	Haldys 2019
343	4-alkoxy and 4-acyloxyacetophenone thiosemicarbazones derivate 76	2.7	uM	nd	Haldys 2019
344	4-alkoxy and 4-acyloxyacetophenone thiosemicarbazones derivate 77	0.4	uM	nd	Haldys 2019
345	4-alkoxy and 4-acyloxyacetophenone thiosemicarbazones derivate 78	0.6	uM	nd	Haldys 2019
346	4-alkoxy and 4-acyloxyacetophenone thiosemicarbazones derivate 79	0.3	uM	nd	Haldys 2019
347	4-alkoxy and 4-acyloxyacetophenone thiosemicarbazones derivate 80	0.5	uM	nd	Haldys 2019
348	4-substituted acetophenones thiosemicarbazones derivate 81	0.3	uM	nd	Haldys 2019
349	4-substituted acetophenones thiosemicarbazones derivate 82	0.6	uM	nd	Haldys 2019
350	4-substituted acetophenones thiosemicarbazones derivate 83	1.0	uM	nd	Haldys 2019
351	4-substituted acetophenones thiosemicarbazones derivate 84	0.5	uM	nd	Haldys 2019
352	4-substituted acetophenones thiosemicarbazones derivate 85	0.5	uM	nd	Haldys 2019
353	4-substituted acetophenones thiosemicarbazones derivate 86	0.6	uM	nd	Haldys 2019
354	4-substituted acetophenones thiosemicarbazones derivate 87	0.3	uM	nd	Haldys 2019
355	4-substituted acetophenones thiosemicarbazones derivate 88	0.3	uM	nd	Haldys 2019

356	4-substituted acetophenones thiosemicarbazones derivate 89	0.4	uM	nd	Haldys 2019
357	Substitution of thiosemicarbazone scaffold with alkyl chain in R2 position 95	1.2	uM	nd	Haldys 2019, Haldys 2018
358	Substitution of thiosemicarbazone scaffold with alkyl chain in R2 position 96	1.0	uM	nd	Haldys 2019, Haldys 2018
359	Substitution of thiosemicarbazone scaffold with alkyl chain in R2 position 97	2.3	uM	nd	Haldys 2019, Haldys 2018
360	Substitution of thiosemicarbazone scaffold with alkyl chain in R2 position 98	1.0	uM	nd	Haldys 2019, Haldys 2018
361	Substitution of thiosemicarbazone scaffold with alkyl chain in R2 position 99	1.0	uM	nd	Haldys 2019, Haldys 2018
362	Substitution of thiosemicarbazone scaffold with alkyl chain in R2 position 100	1.2	uM	nd	Haldys 2019, Haldys 2018
363	Substitution of thiosemicarbazone scaffold with alkyl chain in R2 position 101	1.0	uM	nd	Haldys 2019, Haldys 2018
364	Substitution of thiosemicarbazone scaffold with alkyl chain in R2 position 102	1.0	uM	nd	Haldys 2019, Haldys 2018
365	Substitution of thiosemicarbazone scaffold with alkyl chain in R2 position 103	1.0	uM	nd	Haldys 2019, Haldys 2018
366	Thiosemicarbazone derivatives with 5-membered aromatic heterocyclic substituents 104	4.0	uM	nd	Haldys 2019, Haldys 2018
367	Thiosemicarbazone derivatives with 5-membered aromatic heterocyclic substituents 105	2.4	uM	nd	Haldys 2019, Haldys 2018
368	Thiosemicarbazone derivatives with 5-membered aromatic heterocyclic substituents 106	0.4	uM	nd	Haldys 2019, Haldys 2018
369	Thiosemicarbazone derivatives with 5-membered aromatic heterocyclic substituents 107	4.1	uM	nd	Haldys 2019, Haldys 2018
370	Thiosemicarbazone derivatives with 5-membered aromatic heterocyclic substituents 108	0.8	uM	nd	Haldys 2019, Haldys 2018
371	Thiosemicarbazone derivatives with 5-membered aromatic heterocyclic substituents 109	0.8	uM	nd	Haldys 2019, Haldys 2018
372	Thiophene-2-carbaldehyde thiosemicarbazone derivative 110	0.7	uM	nd	Haldys 2019, Dong 2017
373	Thiophene-2-carbaldehyde thiosemicarbazone derivative 111	0.7	uM	nd	Haldys 2019, Dong 2017
374	Thiophene-2-carbaldehyde thiosemicarbazone derivative 112	0.5	uM	nd	Haldys 2019, Dong 2017
375	Thiophene-2-carbaldehyde thiosemicarbazone derivative 113	1.4	uM	nd	Haldys 2019, Dong 2017
376	Thiophene-2-carbaldehyde thiosemicarbazone derivative 114	0.2	uM	nd	Haldys 2019, Dong 2017
377	Thiophene-2-carbaldehyde thiosemicarbazone derivative 115	0.3	uM	nd	Haldys 2019, Dong 2017
378	Thiophene-2-carbaldehyde thiosemicarbazone derivative 116	0.2	uM	nd	Haldys 2019, Dong 2017

379	Thiophene-2-carbaldehyde thiosemicarbazone derivative 117	0.1	uM	nd	Haldys 2019, Dong 2017
380	Thiophene-2-carbaldehyde thiosemicarbazone derivative 118	0.1	uM	nd	Haldys 2019, Dong 2017
381	Thiophene-2-carbaldehyde thiosemicarbazone derivative 119	0.1	uM	nd	Haldys 2019, Dong 2017
382	Thiophene-2-carbaldehyde thiosemicarbazone derivative 120	0.1	uM	nd	Haldys 2019, Dong 2017
383	Thiophene-2-carbaldehyde thiosemicarbazone derivative 121	0.1	uM	nd	Haldys 2019, Dong 2017
384	Thiophene-2-carbaldehyde thiosemicarbazone derivative 122	0.7	uM	nd	Haldys 2019, Dong 2017
385	Artocae-pin E	6.7	uM	nd	Nguyen 2016
386	Artocae-pin F	50.0	uM	nd	Nguyen 2016
387	Norartocarpetin	50.0	uM	520-30-9	Nguyen 2016
388	Artocarpanone	2.0	uM	nd	Nguyen 2016
389	Liquiritigenin	22.0	uM	578-86-9	Nguyen 2016
390	Dihydromorin	50.0	uM	nd	Nguyen 2016
391	4-(4-Propoxyphenyl)butan-2-ylidenethiosemicarbazide	0.3	uM	nd	You 2015
392	1-(4-(3-Ethoxycarbonyl)propoxyphenyl)propan-2-ylidenethiosemicarbazide	0.5	uM	nd	You 2015
393	4-(4-Hydroxyphenyl)butan-2-ylidenethiocarbonohydrazide	125.6	uM	nd	You 2015
394	4-(4-Methoxyphenyl)butan-2-ylidenethiocarbonohydrazide	208.5	uM	nd	You 2015
395	4-(4-Butoxyphenyl)butan-2-ylidenethiocarbonohydrazide	231.5	uM	nd	You 2015
396	Thiocarbonohydrazide	178.3	uM	2231-57-4	You 2015
397	1-(4-Methoxyphenyl)propan-2-ylidenethiocarbonohydrazide	158.9	uM	nd	You 2015
398	1-(4-Butoxyphenyl)propan-2-ylidenethiocarbonohydrazide	224.1	uM	nd	You 2015
399	1-(4-Methoxyphenyl)ethylidenethiocarbonohydrazide	109.6	uM	nd	You 2015
400	30-geranyl-5,7,20,40-tetrahydroxyisoflavanone	128.2	uM	nd	Tan 2015
401	30-geranyl-5,7,20,50-tetrahydroxyisoflavone	200.0	uM	nd	Tan 2015
402	Neorauflavane	0.5	uM	nd	Tan 2015
403	(E)-3-(3-(3,7-dimethylocta-2,6-dienyl)-2,4-dihydroxyphenyl)-3,5,7-trihydroxy-chroman-4-one	144.0	uM	nd	Tan 2015
404	1-Ethylidenethiosemicarbazide	0.2	uM	nd	Liu 2009
405	1-Propylidenethiosemicarbazide	0.2	uM	nd	Liu 2009
406	1-(3-Methylbutylidene)thiosemicarbazide	0.6	uM	nd	Liu 2009
407	1-(But-2-enylidene)thiosemicarbazide	1.0	uM	nd	Liu 2009
408	1-(Propan-2-ylidene)thiosemicarbazide	0.1	uM	1752-30-3	Liu 2009
409	1-(Butan-2-ylidene)thiosemicarbazide	0.3	uM	1752-40-5	Liu 2009
410	1-(4-Methylpent-3-en-2-ylidene)thiosemicarbazide	11.5	uM	nd	Liu 2009
411	1-Cyclopentylidenethiosemicarbazide	0.2	uM	7283-39-8	Liu 2009
412	1-Cyclohexylidenethiosemicarbazide	1.0	uM	5351-77-9	Liu 2009
413	1-(3-Oxocyclohexylidene)thiosemicarbazide	15.1	uM	nd	Liu 2009

414	1-(2,5-Dimethyl-1H-pyrrol-1-yl)thiourea	0.9	uM	nd	Liu 2009
415	4-Methoxycinnamic acid	410.0	uM	830-09-1	Liu 2009
416	Tropolone	0.4	uM	533-75-5	Liu 2009
417	Dihydrostilbenes derivate 6	31.6	uM	nd	Vontzalidou 2012
418	Dihydrostilbenes derivate 7	8.4	uM	nd	Vontzalidou 2012
419	Dihydrostilbenes derivate 8	49.6	uM	nd	Vontzalidou 2012
420	Dihydrostilbenes derivate 9	33.8	uM	nd	Vontzalidou 2012
421	Dihydrostilbenes derivate 10	30.6	uM	nd	Vontzalidou 2012
422	3,5-Dihydroxy-4'-methoxystilbene	252.0	uM	nd	Shin 1998
423	3,4-Dimethoxy-5-hydroxystilbene	490.0	uM	58436-29-6	Shin 1998
424	Trimethylresveratrol	500.0	uM	nd	Shin 1998
425	Piceid	500.0	uM	27208-80-6	Shin 1998
426	Thiamidol	108.0	uM	1428450-95-6	Mann 2018
427	4-Butylresorcinol	0.6	uM	18979-61-8	Mann 2018
428	4-Hexylresorcinol	1.2	uM	136-77-6	Mann 2018
429	4-Phenylethylresorcinol	0.3	uM	nd	Mann 2018
430	Hydroquinone	1.1	uM	123-31-9	Mann 2018
431	Dimethoxytolylpropyl resorcinol	0.2	uM	nd	Mann 2018
432	α -Thujaplicin	9.5	uM	nd	Yoshimori 2014
433	β -Thujaplicin	0.1	uM	499-44-5	Yoshimori 2014
434	γ -Thujaplicin	0.1	uM	672-76-4	Yoshimori 2014
435	6,7,4 -Trihydroxyisoflavone	0.0	uM	17817-31-1	Chang 2005
436	Daidzein	9.2	uM	486-66-8	Chang 2005
437	Glycitein	237.0	uM	40957-83-3	Chang 2005
438	Genistein	264.0	uM	446-72-0	Chang 2005
439	(E)-3-(4-Hydroxyphenyl)-1-(thiophen-2-yl)prop-2-en-1-one	60.1	uM	nd	Kim 2018
440	(E)-3-(3,4-Dihydroxyphenyl)-1-(thiophen-2-yl)prop-2-en-1-one	103.4	uM	nd	Kim 2018
441	(E)-3-(2,4-Dihydroxyphenyl)-1-(thiophen-2-yl)prop-2-en-1-one	0.9	uM	nd	Kim 2018
442	(E)-3-(4-Hydroxy-3-methoxyphenyl)-1-(thiophen-2-yl)prop-2-en-1-one	200.0	uM	nd	Kim 2018
443	(E)-3-(3-Hydroxy-4-methoxyphenyl)-1-(thiophen-2-yl)prop-2-en-1-one	28.7	uM	nd	Kim 2018
444	(E)-3-(4-Hydroxy-3,5-dimethoxyphenyl)-1-(thiophen-2-yl)prop-2-en-1-one	200.0	uM	nd	Kim 2018
445	(E)-3-(3-Bromo-4-hydroxyphenyl)-1-(thiophen-2-yl)prop-2-en-1-one	112.1	uM	nd	Kim 2018
446	(E)-3-(3,5-Dibromo-4-hydroxyphenyl)-1-(thiophen-2-yl)prop-2-en-1-one	200.0	uM	nd	Kim 2018
447	Dopastin	20.0	uM	37134-80-8	Shiino 2001
448	Cupferron	1.1	uM	135-20-6	Shiino 2001
449	Neocupferron	46.0	uM	1013-20-3	Shiino 2001

450	Compound 4	3.0	uM	nd	Shiino 2001
451	<i>N</i> -Nitroso- <i>N</i> -(3-phenylpropyl)hydroxylamine ammonium salt	1.3	uM	nd	Shiino 2001
452	<i>N</i> -Nitroso- <i>N</i> -(α -methylbenzyl)hydroxylamine ammonium salt	23.5	uM	nd	Shiino 2001
453	<i>N</i> -Nitroso- <i>N</i> -(α -propylbenzyl)hydroxylamine ammonium salt	273.0	uM	nd	Shiino 2001
454	<i>N</i> -Nitroso- <i>N</i> -(diphenylmethyl)hydroxylamine ammonium salt	139.0	uM	nd	Shiino 2001
455	<i>N</i> -cyclopentyl- <i>N</i> -nitroso-hydroxylamine	0.6	uM	nd	Shiino 2001
456	<i>N</i> -Cycloalkyl derivate 10	1.2	uM	nd	Shiino 2001
457	<i>N</i> -Cycloalkyl derivatives 11	1.1	uM	nd	Shiino 2001
458	<i>N</i> -Nitroso- <i>N</i> -hexylhydroxylamine ammonium salt	1.5	uM	nd	Shiino 2001
459	<i>N</i> -Nitroso- <i>N</i> -dodecylhydroxylamine ammonium salt	2.2	uM	nd	Shiino 2001
460	Compound 14	19.3	uM	nd	Shiino 2001
461	<i>N</i> -(2'-hydroxybenzyl)- <i>N</i> -nitrosohydroxylamine ammonium salt	19.2	uM	nd	Shiino 2003
462	<i>N</i> -(3'-hydroxybenzyl)- <i>N</i> -nitrosohydroxylamine, ammonium salt	20.3	uM	nd	Shiino 2003
463	<i>N</i> -(4'-hydroxybenzyl) - <i>N</i> -nitrosohydroxylamine, ammonium salt	13.7	uM	nd	Shiino 2003
464	<i>N</i> -dihydroxybenzyl derivative 7	11.6	uM	nd	Shiino 2003
465	<i>N</i> -dihydroxybenzyl derivative 8	23.8	uM	nd	Shiino 2003
466	4'-chlorobenzyl derivative 9	13.1	uM	nd	Shiino 2003
467	4'-methoxybenzyl derivate 10	9.7	uM	nd	Shiino 2003
468	4'-tert-butylbenzyl derivative 11	15.8	uM	nd	Shiino 2003
469	1-(4-(iso-Propyl)benzylidene)thiosemicarbazide	2.8	uM	nd	Song 2017
470	1-(1-(3'-Methylphenyl)ethylidene)thiosemicarbazide	2.0	uM	nd	Song 2017
471	1-(1-(3',4'-Dimethylphenyl)ethylidene)thiosemicarbazide	0.4	uM	nd	Song 2017
472	1-(1-(4-tert-Butylphenyl)-2-hydroxy-2-methylpropylidene) thiosemicarbazide	200.0	uM	nd	Song 2017
473	1-(1-(Naphthalen-8-yl)ethylidene)thiosemicarbazide	1.9	uM	nd	Song 2017
474	1-(1-(1,2,3,4-Tetrahydronaphthalen-5-yl)ethylidene)thiosemicarbazide	2.4	uM	nd	Song 2017
475	2-([1,10-biphenyl]-4-ylmethylene)hydrazine-1-carbothioamide	2.8	uM	nd	Song 2017
476	2-(1-([1,10-Biphenyl]-4-yl)ethylidene)hydrazine-1-carbothioamide	0.4	uM	nd	Song 2017
477	1-(4-Styrylbenzylidene)thiosemicarbazide	3.1	uM	nd	Song 2017
478	1-(1-(4-Methylphenyl)butylidene)thiosemicarbazide	35.4	uM	nd	Song 2017
479	1-(1-(4-Methylphenyl)-2-methylpropylidene)thiosemicarbazide	115.2	uM	nd	Song 2017
480	1-(1-(4-Methylphenyl)-2-phenylethylidene)thiosemicarbazide	100.0	uM	nd	Song 2017
481	1-(1-(4-iso-Propylphenyl)-2-methylpropylidene)thiosemicarbazide	89.0	uM	nd	Song 2017
482	1-(1-(4-Butylphenyl)propylidene)thiosemicarbazide	9.5	uM	nd	Song 2017

483	1-(1-(4-Butylphenyl)butylidene)thiosemicarbazide	28.8	uM	nd	Song 2017
484	1-(1-(4-Cyclohexylphenyl)propylidene)thiosemicarbazide	8.1	uM	nd	Song 2017
485	1-(1-(4-Cyclohexylphenyl)butylidene)thiosemicarbazide	56.7	uM	nd	Song 2017
486	1-(1-(4-Cyclohexylphenyl)-2-methylpropylidene)thiosemicarbazide	98.3	uM	nd	Song 2017
487	1-(1-(4-Cyclohexylphenyl)-2-phenylethylidene)thiosemicarbazide	100.0	uM	nd	Song 2017
488	1-(1,2-Diphenylethylidene)thiosemicarbazide	100.0	uM	nd	Song 2017
489	1-(1-(4-Hydroxyphenyl)ethylidene)thiosemicarbazide	0.3	uM	nd	Liu 2008
490	1-(1-(2,4-Dihydroxyphenyl)ethylidene)thiosemicarbazide	0.6	uM	nd	Liu 2008
491	1-(1-(2,4,6-Trihydroxyphenyl)ethylidene)thiosemicarbazide	22.0	uM	nd	Liu 2008
492	1-(1-(4-Isopropylphenyl)ethylidene)thiosemicarbazide	1.0	uM	nd	Liu 2008
493	1-(1-(4-Methoxyphenyl)ethylidene)thiosemicarbazide	0.1	uM	nd	Liu 2008
494	1-(1-(Pyrazin-2-yl)ethylidene)thiosemicarbazide	0.9	uM	nd	Liu 2008
495	1-(1-(Pyridin-3-yl)ethylidene)thiosemicarbazide	0.8	uM	nd	Liu 2008
496	1-(2-Hydroxy-1,2-diphenylethylidene)thiosemicarbazide	55.5	uM	nd	Liu 2008
497	1-(2-Oxo-1,2-diphenylethylidene)thiosemicarbazide	0.1	uM	nd	Liu 2008
498	1-(1,4-Diacetylphenyl)dithiosemicarbazide	0.2	uM	nd	Liu 2008
499	2-Hydroxytyrosol (2-HT)	571.0	uM	10597-60-1	Uchida 2013
500	3'-geranyl-5,7,2',4'-tetrahydroxyisoflavanone	128.2	uM	nd	Tan,2015
501	3'-geranyl-5,7,2',5'-tetrahydroxyisoflavone	200.0	uM	nd	Tan,2015
502	Cyclocarposide (C1)	102.4	uM	nd	Casañola, 2006
503	Cyclosieversioside (C2)	92.3	uM	nd	Casañola, 2006
504	Askendoside D(C3)	48.6	uM	nd	Casañola, 2006
505	Askendoside B(C4)	14.0	uM	nd	Casañola, 2006
506	Cycloorbicoside G (C5)	54.6	uM	nd	Casañola, 2006
507	Bryoamaride (C6)	85.0	uM	nd	Casañola, 2006
508	Chalcone	3.0	uM	94-41-7	Casañola, 2006
509	4'-Hydroxychalcone (4'-HC)	11.0	uM	nd	Nerya, 2004
510	2'-Hydroxychalcone (2'-HC)	4.0	uM	nd	Nerya, 2004
511	2',4'-Dihydroxychalcone (2',4'-HC)	14.0	uM	nd	Nerya, 2004
512	2',4-Dihydroxychalcone (2',4-HC)	10.0	uM	nd	Nerya, 2004
513	Benzaldehyde Derivative 1	4517.1	uM	nd	Li, 2004
514	Benzaldehyde Derivative 10	2381.1	uM	nd	Li, 2004
515	Benzaldehyde Derivative 11	1541.5	uM	nd	Li, 2004
516	Benzaldehyde Derivative 12	2538.9	uM	nd	Li, 2004
517	Benzaldehyde Derivative 14	1182.5	uM	nd	Li, 2004
518	Benzaldehyde Derivative 15	37907.6	uM	nd	Li, 2004
519	Benzaldehyde Derivative 16	10821.6	uM	nd	Li, 2004
520	Benzaldehyde Derivative 17	6119.5	uM	nd	Li, 2004

521	Benzaldehyde Derivative 18	1809.6	uM	nd	Li, 2004
522	3,5 dimethylphenyl N-tosyl substituted indole-based thiosemicarbazone	20.6	uM	nd	Batool, 2024
523	4-methoxyphenyl N-tosyl substituted indole-based thiosemicarbazone	29.4	uM	nd	Batool, 2024
524	4-chlorobenzyl N-tosyl substituted indole-based thiosemicarbazone	13.7	uM	nd	Batool, 2024
525	2-methylphenyl N-tosyl substituted indole-based thiosemicarbazone	24.3	uM	nd	Batool, 2024
526	Cyclohexyl N-tosyl substituted indole-based thiosemicarbazone	61.8	uM	nd	Batool, 2024
527	2,3-dichlorophenyl N-tosyl substituted indole-based thiosemicarbazone	11.5	uM	nd	Batool, 2024
528	4-methylbenzyl N-tosyl substituted indole-based thiosemicarbazone	28.6	uM	nd	Batool, 2024
529	Phenyl N-tosyl substituted indole-based thiosemicarbazone	42.2	uM	nd	Batool, 2024
530	2,6-dimethylphenyl N-tosyl substituted indole-based thiosemicarbazone	9.5	uM	nd	Batool, 2024
531	4-methylphenyl N-tosyl substituted indole-based thiosemicarbazone	14.3	uM	nd	Batool, 2024
532	4-fluorophenyl N-tosyl substituted indole-based thiosemicarbazone	21.4	uM	nd	Batool, 2024
533	4-chlorophenyl N-tosyl substituted indole-based thiosemicarbazone	8.3	uM	nd	Batool, 2024
534	Benzyl N-tosyl substituted indole-based thiosemicarbazone	37.6	uM	nd	Batool, 2024
535	2-fluorophenyl N-tosyl substituted indole-based thiosemicarbazone	10.5	uM	nd	Batool, 2024
536	4-bromophenyl N-tosyl substituted indole-based thiosemicarbazone	25.3	uM	nd	Batool, 2024
537	2,4-dimethylphenyl N-tosyl substituted indole-based thiosemicarbazone	19.4	uM	nd	Batool, 2024
538	4-nitrophenyl N-tosyl substituted indole-based thiosemicarbazone	7.3	uM	nd	Batool, 2024
539	3-nitrophenyl N-tosyl substituted indole-based thiosemicarbazone	6.4	uM	nd	Batool, 2024
540	2-(3-acetylphenylamino)-2-oxoethyl-3-hydroxybenzoate	322.7	uM	nd	Nazir, 2022
541	2-(3-acetylphenylamino)-2-oxoethyl-4-hydroxybenzoate	287.5	uM	nd	Nazir, 2022
542	2-(3-acetylphenylamino)-2-oxoethyl 2,4-dihydroxybenzoate	27.4	uM	nd	Nazir, 2022
543	2-(3-acetylphenylamino)-2-oxoethyl 3,4-dihydroxybenzoate	127.7	uM	nd	Nazir, 2022
544	2-(3-acetylphenylamino)-2-oxoethyl 3,5-dihydroxybenzoate	255.3	uM	nd	Nazir, 2022
545	2-(3-acetylphenylamino)-2-oxoethylcinnamate	46.4	uM	nd	Nazir, 2022
546	2-(3-acetylphenylamino)-2-oxoethyl(E)-3-(4-hydroxyphenyl) acrylate	309.2	uM	nd	Nazir, 2022

547	2-(3-acetylphenylamino)-2-oxoethyl(E)-3-(2,4-dihydroxyphenyl)acrylate	0.0	uM	nd	Nazir, 2022
548	2-(3-acetylphenyl amino)-2-oxoethyl(E)-3-(4-chlorophenyl)acrylate	61.6	uM	nd	Nazir, 2022
549	2-(3-acetylphenylamino)-2-oxoethyl3-(4-hydroxyphenyl) propanoate	nd	-	nd	Nazir, 2022
550	5-6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid	nd	-	53174-06-4	Khan 2007
551	4-Propylresorcinol	nd	-	18979-60-7	Khatib 2005
552	Glabridin	nd	-	59870-68-7	Khatib 2005
553	Artocarbene	nd	-	194144-02-0	Khatib 2005
554	Chlorophorin	nd	-	537-41-7	Khatib 2005
555	Chalcone 10	nd	-	nd	Zolghadry 2019
556	Chalcone 11	nd	-	nd	Zolghadry 2019
557	Chalcone 12	nd	-	nd	Zolghadry 2019
558	Azachalcone 13	nd	-	nd	Zolghadry 2019
559	Azachalcone 14	nd	-	nd	Zolghadry 2019
560	Oxime based chalcone serie 15	nd	-	nd	Zolghadry 2019
561	Oxime based chalcone serie 16	nd	-	nd	Zolghadry 2019
562	2,3-dihydro-1H-inden-1-one chalcone-like 17	nd	-	nd	Zolghadry 2019
563	2,3-dihydro-1H-inden-1-one chalcone-like 18	nd	-	nd	Zolghadry 2019
564	(E)-3-(2,4-Dihydroxyphenyl)-1-(thiophen-2-yl)prop-2-en-1-one	nd	-	nd	Kim 2018
565	2',4',4'-Trihydroxychalcone-3',3'-di-isoprenyl chalcone (IPC)	nd	-	nd	Nerya, 2004