

**UNIVERSIDAD NACIONAL DE SAN CRISTÓBAL
DE HUAMANGA**

FACULTAD DE CIENCIAS DE LA SALUD

ESCUELA PROFESIONAL DE FARMACIA Y BIOQUÍMICA



TESIS:

**Metabolitos bioactivos antiinflamatorios de la flora peruana
identificados por análisis quimioinformático comparativo**

Para optar el título profesional de:
QUÍMICO FARMACÉUTICO

PRESENTADO POR:
Bach. Yasmin Nieves CAYLLAHUA ORE

ASESORA:
MSc. Kirianova GODOY BAUTISTA

COASESOR:
Q.F. Hugo Jesús VALDIVIA OLARTE

AYACUCHO - PERÚ

2025

Dedicado a Dios, a mis padres y a todos los que me apoyaron en el proceso.

AGRADECIMIENTOS

A la Universidad Nacional de San Cristóbal de Huamanga, mi *Alma Máter* por brindarme gran conocimiento en los cinco años de enseñanza y plantar en mí el querer científico.

A la Facultad de Ciencias de la Salud, en especial a la Escuela Profesional de Farmacia y Bioquímica, que lograron formar un profesional competente y veraz.

A mi asesora: MSc. KIRIANOVA GODOY BAUTISTA, profesora adscrita al Departamento Académico de Medicina Humana de la Facultad de Ciencias de la Salud, por su permanente apoyo y consejería en el desarrollo de la investigación y al Q.F. HUGO VALDIVIA OLARTE, por su revisión metodológica y análisis crítico de los resultados.

A la Facultad de Química e Ingeniería Química de la Universidad Mayor de San Marcos por brindarme la herramienta computacional para realizar mis análisis.

A los profesores de la Escuela Profesional de Farmacia y Bioquímica quienes con sus amplios conocimientos me formaron y guiaron en este ámbito científico y profesional.

ÍNDICE GENERAL

	Página
CAPITULO I. INTRODUCCION	1
CAPITULO II. DESARROLLO DE LA PERSPECTIVA TEÓRICA	3
2.1. Marco Referencial	3
2.2. Marco teórico	12
2.2. Marco conceptual	29
2.3. Marco ético y legal	31
CAPÍTULO III. MATERIALES Y MÉTODOS	33
3.1. Alcance de investigación	33
3.2. Diseño de investigación	33
3.3. Unidad de análisis	33
3.4. Población de estudio	33
3.5. Muestra	33
3.6. Criterios de selección	33
3.7. Técnicas e Instrumentos de Recolección de Datos	33
3.8. Procedimiento para la recolección de datos.	34
CAPITULO IV. RESULTADOS	43
CAPITULO V. DISCUSIÓN	59
CAPITULO VI. CONCLUSIONES	69
CAPITULO VII. RECOMENDACIONES	71
BIBLIOGRAFÍA	73
ANEXOS	85

ÍNDICE DE SIGLAS

ADME/Tox	Absorción, Distribución, Metabolismo, Excreción y Toxicidad.
AINE	Antiinflamatorio no esteroideo.
API	Ingrediente Farmacéutico Activo.
BCUT	Propiedades de Burden (propiedades electrónicas/estructurales).
COX-1	Ciclooxigenasa-1.
COX-2	Ciclooxigenasa-2.
COX-3	Ciclooxigenasa-3.
CSR	Curvas de Recuperación de Sistemas Cíclicos (<i>Cyclic System Recovery</i>).
CSV	Valores separados por comas.
ECFP	<i>Extended-Connectivity Fingerprints</i> (huellas moleculares extendidas).
ESP	<i>Electrostatic Potential</i> (Potencial Electroestático).
ESP-sim	Índice de similitud basado en potencial electrostático 3D.
FDA	Administración de Alimentos y Medicamentos de EE. UU.
GAN	<i>Generative Adversarial Network</i> (Red Generativa Antagónica).
GRAS	Sustancias Generalmente Reconocidas como Seguras.
HBA	Aceptores de puentes de hidrógeno.
HBD	Donadores de puentes de hidrógeno.
IQR	Rango intercuartílico.
KNN	<i>K-Nearest Neighbors</i> (K vecinos más cercanos).
KNApSAcK	Base de datos de relación metabolito especie vegetal.
LOX	Lipooxigenasa.
MACCS	<i>Molecular ACCess System</i> (huellas moleculares de 166 bits).
MAPK	<i>Mitogen-Activated Protein Kinase</i> (Proteína quinasa activada por mitógenos).
MCS	<i>Maximum Common Substructure</i> (Subestructura mayor común).

MW	Peso molecular.
NF- κ B	Factor Nuclear Kappa Beta.
NuBBEDB	<i>Brazilian Natural Products Database</i> (Base de datos brasileña de productos naturales).
PCA	Análisis de componentes principales.
PDB	<i>Protein Data Bank</i> .
PGE ₂	Prostaglandina E ₂ .
PheSA Score	<i>Farmacóforo Shape Alignment Score</i> (Alineamiento sofisticado de forma y farmacóforo).
PLA ₂	Fosfolipasa A ₂ .
PSA / TPSA	Área superficial polar (topológica).
QSAR	Relación cuantitativa estructura–actividad.
RB	Enlaces rotables.
RDKit	Toolkit de química computacional de código abierto.
RENATI	Registro Nacional de Trabajos de Investigación del Perú.
ROC	<i>Receiver Operating Characteristic</i> (Característica operativa del receptor).
SANCDDB	Database of South African natural compounds (Base de datos de compuestos naturales sudafricanos).
SAR	Relación estructura–actividad.
SMILES	<i>Simplified Molecular Input Line Entry System</i> (Sistema simplificado de introducción de líneas de entrada molecular).
TNF- α	Factor de necrosis tumoral alfa.
TPSA	Área superficial polar topológica
UNPD	Base de datos universal de productos naturales
VAE	Autoencoder variacional.

ÍNDICE DE TABLAS

		Página
Tabla 1.	Tipos de metabolitos con actividad antiinflamatoria, sus grupos funcionales y mecanismo de acción.	14
Tabla 2.	Cuadro cronológico de la Quimioinformática.	17
Tabla 3.	Comparación de bases de datos en tipo, tamaño aproximado, contenido/enfoque y usos principales.	19
Tabla 4.	Tabla resumen de las métricas de diversidad y similitud.	27
Tabla 5.	Plantas peruanas con actividad antiinflamatoria registradas desde 1998 hasta el febrero del 2025 en la base de datos RENATI.	44
Tabla 6.	Resumen de las plantas y medicamentos con actividad antiinflamatoria recopiladas de las cuatro bases de datos RENATI, NuBBEDB, SANCDB y DrugBank.	48
Tabla 7.	Candidatos antiinflamatorios prometedores obtenidos mediante análisis de consenso que integra los puntajes de similaridad de Tanimoto, Phesa y ESP-sim. Teniendo como punto de corte como puntaje mínimo 0,105.	57

ÍNDICE DE FIGURAS

	Página
Figura 1. Estructura de los principales flavonoides.	16
Figura 2. Evolución de las estrategias para el descubrimiento de nuevos fármacos de forma más rápida y económica.	18
Figura 3. Esquema del Cribado Virtual: Enfoques Basados en Estructura y Ligandos.	22
Figura 4. Esquema visual de los parámetros de las Reglas de Lipinski: peso molecular (MW), área superficial polar (PSA), enlaces rotables (RB), aceptores de enlaces de hidrógeno (HBA), donantes de enlaces de hidrógeno (HBD), log P.	23
Figura 5. Ejemplo de huella dactilar de la Tirosina, representado como cadenas de bits.	27
Figura 6. Distribución geográfica de plantas peruanas con potencial antiinflamatorio recopiladas de la base de datos RENATI (del año 1998 hasta febrero 2025).	46
Figura 7. Cantidad de metabolitos encontrados en las plantas peruanas recopiladas de la base de datos RENATI.	47
Figura 8. Pesos moleculares (MW) de las cuatro bases de datos estudiadas (RENATI, DrugBank, SANCDB, NuBBEDB).	48
Figura 9. Área de superficie polar (TPSA) de las cuatro bases de datos estudiadas (RENATI, DrugBank, SANCDB, NuBBEDB).	49
Figura 10. Enlaces rotables (RB) de las cuatro bases de datos estudiadas (RENATI, DrugBank, SANCDB, NuBBEDB).	50
Figura 11. LogP de las cuatro bases de datos estudiadas (RENATI, DrugBank, SANCDB, NuBBEDB).	51
Figura 12. Número de donadores de puentes de hidrógeno (nHBDon) de las cuatro bases de datos estudiadas (RENATI, DrugBank, SANCDB, NuBBEDB).	52

Figura 13.	Número de aceptores de puentes de hidrógeno (nHB _{Acc}) de las cuatro bases de datos estudiadas (RENATI, DrugBank, SANCDB, NuBBEDB).	53
Figura 14.	Análisis de componentes principales (PCA): Visualización del espacio químico entre las cuatro bases de datos estudiadas (RENATI, DrugBank, SANCDB, NuBBEDB).	54
Figura 15.	A. Histograma de frecuencias de los andamiajes (scaffolds) encontrados en las cuatro bases. B. Estructuras químicas de los andamiajes encontrados en las cuatro bases de datos.	55
Figura 16.	Diagrama de Venn obtenido de los tres análisis realizados (ESPSim, PheSA, Tanimoto) a la base de datos RENATI comparando con las tres bases de referencia (SANCDB, NuBBEDB, DrugBank).	56

ÍNDICE DE ANEXOS

	Página
Anexo 1. Enlace de Drive de las plantas encontradas en la base de datos RENATI.	85
Anexo 2. Enlace de Drive de las cuatro bases de datos generadas.	85
Anexo 3. Enlace de Drive del análisis de PheSA SCORE de las cuatro bases de datos.	85
Anexo 4. Enlace de Drive del análisis de ESPSim de las cuatro bases de datos.	85
Anexo 5. Enlace de Drive del análisis de Distancia de Tanimoto, de las cuatro bases de datos.	85
Anexo 6. Enlace de Drive del análisis de 5 reglas de Lipinski efectuado a las cuatro bases de datos.	85
Anexo 7. Enlace de Drive del análisis de PCA efectuado a las cuatro bases de datos.	85
Anexo 8. Toma de pantalla del script realizado en Jupyter Notebook usado para la curación de las cuatro bases de datos usadas.	86
Anexo 9. Toma de pantalla del script realizado en Python usado para la generación de los descriptores moleculares (Reglas de Lipinski) y PCA.	87
Anexo 10. Toma de pantalla del script realizado en Python usado para la generación scaffolds de las cuatro bases de datos.	88
Anexo 11. Toma de pantalla de los parámetros usados para la creación de confórmers de las cuatro bases de datos en Data Warrior.	89
Anexo 12. Toma de pantalla de la superposición realizada a las plantas peruanas vs las bases de datos de referencia en Data Warrior.	90
Anexo 13. Toma de pantalla del script para la conversión de SMILES a archivos SDF para el análisis de ESPSim realizado en Jupyter Notebook.	91
Anexo 14. Toma de pantalla del script para el análisis de la metodología ESPSim frente a las tres bases de datos de referencia.	92

Anexo 15.	Toma de pantalla del script realizado en Rstudio para el análisis de la metodología Distancia de Tanimoto frente a las tres bases de datos de referencia.	93
Anexo 16.	Toma de pantalla del script realizado en Rstudio para la creación del Diagrama de Venn de basado en las tres metodologías usadas.	93
Anexo 17.	Toma de pantalla del script realizado en Rstudio para el análisis del puntaje consenso obtenido del diagrama de Venn.	94
Anexo 18.	Toma de pantalla del script realizado en Rstudio para la validación de datos usando el diagrama de violín plot y percentil.	94
Anexo 19.	Histograma de frecuencia de la validación de datos realizadas de los candidatos de plantas peruanas versus los decoys (falsos positivos).	95
Anexo 20.	Obtención del umbral del puntaje consenso obtenido del análisis de las plantas peruanas, mediante el método del codo (Elbow method) para validación de datos.	95
Anexo 21.	Violin Plot de la distribución del Consensus Score entre falsos positivos (Decoys) y candidatos para validación del análisis consenso. Ilustración realizada en RStudio.	96
Anexo 22.	Tabla de enriquecimiento farmacológico de los 11 candidatos reportados.	97
Anexo 23.	Matriz de definición y operacionalización de variables.	98
Anexo 24.	Matriz de consistencia.	99

RESUMEN

La flora peruana es una fuente poco explorada de moléculas con potencial antiinflamatorio. Este estudio aborda esta brecha mediante un enfoque quimioinformático comparativo de cribado molecular tipo SAR. El objetivo fue identificar metabolitos secundarios de la flora peruana estructuralmente similares a compuestos antiinflamatorios de referencia. A partir de 56 plantas reportadas en RENATI, se recopilaron 1501 estructuras (SMILES) de KNApSAcK. Este conjunto se comparó estructuralmente con bases de datos de fármacos (DrugBank) y productos naturales (SANCDB, NuBBEDB). La evaluación incluyó la caracterización de propiedades fisicoquímicas, PCA, Núcleos de las estructuras de las moléculas (*Scaffolds*) y la cuantificación de la similitud molecular mediante tres métricas: Coeficiente de Tanimoto (2D), PheSA Score (farmacofórica 3D) y ESP-Sim (electrostática 3D). Un algoritmo de consenso integró los puntajes para priorizar candidatos. La robustez del método de selección fue validada estadísticamente (test de Wilcoxon, $p < 2.2e-16$), demostrando una separación significativa entre candidatos y compuestos de distracción. Los resultados confirman que estos metabolitos poseen características estructurales, fisicoquímicas y farmacocinéticas adecuadas para candidatos a fármacos, permitiendo identificar 11 candidatos de alto potencial antiinflamatorio, validando así el enfoque quimioinformático para explorar la biodiversidad peruana en el descubrimiento de fármacos.

Palabras clave: Quimioinformática, productos naturales, antiinflamatorio, similitud molecular, flora peruana.

ABSTRACT

Peruvian flora is an under-explored source of molecules with anti-inflammatory potential. This study addresses this gap through a comparative chemoinformatic approach to SAR molecular screening. The objective was to identify secondary metabolites from Peruvian flora that are structurally similar to reference anti-inflammatory compounds. From 56 plants reported in RENATI, 1,501 structures (SMILES) were collected from KNApSAcK. This set was structurally compared with databases of drugs (DrugBank) and natural products (SANCDB, NuBBEDB). The evaluation included the characterization of physicochemical properties, PCA, molecule structure scaffolds, and the quantification of molecular similarity using three metrics: Tanimoto coefficient (2D), PheSA Score (3D pharmacophore), and ESP-Sim (3D electrostatic). A consensus algorithm integrated the scores to prioritize candidates. The robustness of the selection method was statistically validated (Wilcoxon test, $p < 2.2e-16$), demonstrating a significant separation between candidates and distractor compounds. The results confirm that these metabolites have structural, physicochemical, and pharmacokinetic characteristics suitable for drug candidates, allowing the identification of 11 candidates with high anti-inflammatory potential, thus validating the chemoinformatic approach to exploring Peruvian biodiversity in drug discovery.

Keywords: Chemoinformatics, natural products, anti-inflammatory, molecular similarity, Peruvian flora.

CAPITULO I. INTRODUCCION

Perú es uno de los países más megadiversos del planeta, alberga una extensa variedad de recursos genéticos vegetales y es considerado como un Centro Vavilov,¹ lo cual lo convierte en un reservorio invaluable de metabolitos secundarios con propiedades bioactivas promisorias, aún por descubrir. La medicina tradicional peruana, forjada a través del tiempo por diversas culturas, ha sabido aprovechar la riqueza natural del país, posicionando a las plantas medicinales como eje fundamental de su saber ancestral.²

Muchas de estas plantas han sido usadas por los antiguos peruanos por sus propiedades antiinflamatorias; sin embargo, a pesar de ser poseedores de información etnofarmacológica valiosa, la mayoría de los estudios de actividad antiinflamatoria realizadas con plantas peruanas ha sido enfocado desde un punto de vista tradicional, sin ser capaces de identificar la molécula responsable de dicha actividad, lo cual es una estrategia ineficiente en tiempo y costo para el descubrimiento de nuevas drogas.³ Hecho que dificulta enormemente el descubrimiento de nuevos y prometedores agentes antiinflamatorios.

En la actualidad, las enfermedades inflamatorias continúan siendo un desafío para la salud pública global, por lo que la búsqueda de agentes antiinflamatorios más eficaces, selectivos y con menos efectos adversos son una prioridad en la investigación farmacéutica particularmente para el manejo de enfermedades inflamatorias crónicas, los cuales están fuertemente relacionados con la aparición y progresión de enfermedades neurodegenerativas, trastornos cardiovasculares, cáncer y enfermedades autoinmunes.⁴

La mayoría de los antiinflamatorios disponibles son útiles en el tratamiento de la inflamación aguda, sin embargo, muestran importantes reacciones adversas relacionadas con afecciones respiratorias, gastrointestinales y metabólicas,^{5,6} además de pérdida de eficacia con el tiempo de uso, asimismo, su eficacia y seguridad es limitada frente al tratamiento de procesos inflamatorios crónicos mediados por

mecanismos moleculares diversos. En tal sentido, esta investigación surgió de la necesidad de identificar moléculas antiinflamatorias prometedoras a partir de la vasta biodiversidad peruana empleando análisis quimioinformático y herramientas computacionales de vanguardia, teniendo como punto de partida todos los trabajos académicos de actividad antiinflamatoria de plantas medicinales reportadas en el país a través de RENATI y responde a la interrogante: ¿El análisis quimioinformático comparativo SAR (relación estructura actividad) basado en algoritmos de similaridad y diversidad, permitirá identificar metabolitos antiinflamatorios de la flora peruana con perfiles similares a fármacos y productos naturales de referencia? En consecuencia, los objetivos fueron:

Objetivo general: Identificar metabolitos antiinflamatorios de la flora peruana mediante análisis quimioinformático comparativo, basado en algoritmos de similaridad y diversidad, con perfiles similares a fármacos y productos naturales de referencia.

Objetivos específicos:

- Generar una base de datos de los metabolitos de la flora peruana que registran actividad antiinflamatoria en RENATI y que cuentan con información estructural reportada en la base de datos KNApSACk.
- Realizar análisis quimioinformático comparativo de similaridad y diversidad de los metabolitos de plantas peruanas que reportan actividad antiinflamatoria en RENATI con fármacos antiinflamatorios de referencia.
- Realizar análisis quimioinformático comparativo de similaridad y diversidad de los metabolitos de plantas peruanas que reportan actividad antiinflamatoria en RENATI con productos naturales de referencia.
- Identificar metabolitos de plantas peruanas con potencial actividad antiinflamatoria.

CAPITULO II. DESARROLLO DE LA PERSPECTIVA TEÓRICA

2.1. Marco Referencial

2.1.1. *Antecedentes internacionales*

Mwamatope *et al*⁷, realizaron en el año 2024 un estudio cuyo objetivo general fue identificar las plantas medicinales utilizadas tradicionalmente en Malawi para el tratamiento de síntomas alérgicos y evaluar, mediante análisis quimioinformático, el potencial biológico de los metabolitos asociados. Para ello, se aplicó una revisión sistemática en bases de datos científicas para recuperar estudios etnobotánicos, se recopilaron los metabolitos reportados desde PubChem y se depuró la información estructural mediante un flujo de trabajo con RDKit en KNIME, eliminándose duplicados y compuestos fuera de criterios fisicoquímicos relevantes; posteriormente, se calcularon descriptores moleculares como MW, HBA, HBD, LogP, TPSA y RB, y se evaluó la diversidad y la similaridad estructural mediante huellas MACCS y coeficiente de Tanimoto. Los resultados indicaron que se identificaron 60 especies de plantas asociadas al manejo de alergias, de las cuales 39 presentaron metabolitos registrados, obteniéndose un conjunto curado de 1757 compuestos; además, cerca del 49 % de los metabolitos cumplió la regla de Lipinski y el 45 % la regla de Veber, y el análisis de espacio químico mostró una amplia diversidad estructural, identificándose compuestos como la usambarina con alta similaridad ($T_c > 0,70$) respecto a antihistamínicos experimentales. Se concluyó que las plantas medicinales de Malawi constituyeron una fuente relevante de productos naturales con características fisicoquímicas compatibles con perfiles drug-like y con potencial actividad antihistamínica, lo que justificó su consideración como candidatos para investigaciones posteriores orientadas al descubrimiento de nuevos agentes terapéuticos.

Gómez *et al*⁸ en su estudio realizado en el 2023 titulado “Navegando por el espacio químico y el multiverso químico de una base de datos unificada de productos naturales latinoamericanos: LANaPDB”, tuvo como objetivo caracterizar el espacio químico de los productos naturales latinoamericanos mediante un análisis comparativo

con bases de datos consolidadas, con el fin de evaluar su diversidad estructural y potencial utilidad en el descubrimiento de fármacos. Para ello, se emplearon herramientas quimiométricas y técnicas de análisis computacional, como el análisis de componentes principales (PCA) para representar el espacio químico utilizando propiedades fisicoquímicas; además, se calcularon huellas digitales moleculares MACCS (166 bits) y ECFP4 para estimar la similitud estructural aplicando coeficientes de Tanimoto y Coseno, y se evaluó la diversidad de andamiajes moleculares mediante la metodología de Bemis–Murcko, analizando métricas derivadas de curvas CSR como el área bajo la curva (AUC) y la fracción de quimiotipos que conforman el 50 % del conjunto (F50). Se reportó que algunas bases de datos, como UPMA_2V y NUBBE_1V, presentaron mayor diversidad estructural, con F50 de 0,19 y 0,18, respectivamente, mientras que bases como SPECS_NP y AFRODB mostraron menor diversidad con altos valores de AUC (0,90 y 0,79). Asimismo, se observó que los compuestos de bases como IBSCREEN_NP y PRINCETON_NP presentaron mayor complejidad molecular, reflejada en altos promedios de carbonos hibridados sp³ (indicador de complejidad molecular) y fracción de carbonos quirállicos (FCC) con valores entre 0,22 y 0,35, mientras que bases como LAIPNUDELSAV y NPACT_NP mostraron menor complejidad con valores entre 0,15 y 0,17. Se concluyó que los productos naturales latinoamericanos representaron una fuente significativa de estructuras químicas diversas y complejas, con potencial prometedor para la exploración en descubrimiento de fármacos, especialmente por su variabilidad en andamiajes moleculares y perfiles fisicoquímicos distintivos.

Periwal et al⁹ efectuaron en 2022 una investigación denominada: “Evaluación de la bioactividad de compuestos naturales mediante modelos de aprendizaje automático entrenados en la similitud de objetivos entre fármacos”, que tuvo como objetivo evaluar el potencial bioactivo de compuestos naturales mediante la predicción de su capacidad de unión a proteínas humanas conocidas como dianas terapéuticas. La investigación se enfocó en el uso de modelos de aprendizaje automático entrenados con múltiples herramientas de análisis de similitud química, entre ellas métricas de similitud estructural como huellas moleculares (*fingerprints*), similitud de subestructura máxima común (MCS) y descriptores fisicoquímicos, con el fin de capturar relaciones complejas entre estructura y actividad biológica. La metodología consistió en entrenar

clasificadores utilizando un conjunto de 1410 fármacos aprobados por la FDA junto con sus respectivos blancos proteicos como etiquetas de clase, considerando que los compuestos que compartían una diana se clasificaron como “Match”. Para representar la similitud química entre pares de moléculas, se calcularon múltiples huellas moleculares como Morgan, Featmorgan, MACCS, RDKit y Atompair, además de descriptores estructurales y topológicos. Se aplicaron modelos de aprendizaje supervisado, incluyendo regresión logística, Naïve Bayes y bosque aleatorio, siendo este último el que mostró mejor desempeño tras un proceso de optimización de hiperparámetros. Los modelos de bosque aleatorio fueron los que lograron el mejor rendimiento, alcanzando un área bajo la curva ROC (Características Operativas del Receptor) de 0,90, un coeficiente de correlación de Matthews (MCC) de 0,35 y una puntuación F1 de 0,33 indicando una adecuada capacidad para capturar relaciones estructura-actividad. La ROC se definió como una curva que evalúa la relación entre sensibilidad y especificidad; el MCC como una métrica útil para evaluar rendimiento en conjuntos desbalanceados; y la puntuación F1 como el promedio armónico entre precisión y exhaustividad. Posteriormente, los modelos fueron aplicados para predecir posibles interacciones entre 11000 compuestos naturales y los fármacos entrenados, identificando varios compuestos con potencial terapéutico, algunos ya reportados en literatura y otros no explorados previamente. Como validación experimental, se comprobó la inhibición de la proteína Cox-1 por parte del compuesto natural 5-metoxisalicílico, lo cual respaldó la efectividad del enfoque predictivo empleado. En conclusión, los resultados demuestran la utilidad de un enfoque de aprendizaje automático que combina múltiples características químicas para descubrir el potencial de unión a proteínas de los compuestos naturales, contribuyendo a acelerar la identificación computacional de moléculas bioactivas.

Zhang *et al*¹⁰ en el año 2022 presentó el artículo titulado: “InflamNat: base de datos web y predictor de productos naturales antiinflamatorios” ha diseñado una plataforma en línea para investigar productos naturales con propiedades antiinflamatorias y facilitar el descubrimiento de nuevos fármacos. La base de datos contiene información sobre 1351 compuestos naturales, incluidas las propiedades fisicoquímicas y la actividad biológica de los compuestos siendo los más comunes los flavonoides. Se implementaron dos herramientas predictivas de aprendizaje automático

(AI-A y CT) utilizando el modelo Multi-Tokenization Transformer (MTT) para mejorar la representación de los datos, logrando valores de área bajo la curva (AUC) de 0,842 y 0,872, respectivamente, superiores a otros métodos existentes. En conclusión, se creó una base de datos de productos antiinflamatorios naturales.

Menezes *et al*¹¹ en el estudio publicado en el 2022 denominado: “Descubrimiento asistido por computadora de alcaloides con actividad esquistosomicida”, cuyo objetivo fue el desarrollo de nuevos fármacos contra *Schistosoma mansoni*, utilizó cribado virtual (basado en ligandos y estructuras) y acoplamiento molecular para identificar nuevos alcaloides esquistosomicidas. Se emplearon las bases de datos Sistemax y ChEMBL para generar modelos predictivos con una precisión superior al 84%. La evaluación *in silico* identificó cinco alcaloides prometedores, de los cuales dos, 11, 12 – metilen – dioxi – kopsaporina y metil – 3 – oxo – 12 – metoxi – n(1) – decarbometoxi – 14,15 - didehidrocanofruticosinato, mostraron perfiles toxicológicos y metabólicos favorables. Estos resultados establecen puntos de partida basados en productos naturales para el desarrollo de nuevos tratamientos.

Núñez *et al*¹², en su estudio realizado el 2021 titulado: “Bases de datos latinoamericanas de productos naturales: biodiversidad y descubrimiento de fármacos contra el SARS-CoV-2”, tuvieron como propósito analizar de manera integral el espacio químico, la diversidad estructural, la complejidad molecular y el potencial bioactivo de productos naturales latinoamericanos utilizando herramientas y métricas avanzadas de quimioinformática y bioinformática, sin mencionar el uso de nanopartículas. Para ello, evaluaron 15 bases de datos de compuestos naturales (cuatro latinoamericanas y once internacionales), que tras filtrado contenían 46413 moléculas. La metodología combinó diversas estrategias computacionales: análisis del espacio químico mediante PCA; evaluación de similitud molecular usando huellas digitales moleculares con coeficientes de similitud de Tanimoto y Cosine; estudio de diversidad de andamiajes estructurales mediante curvas de recuperación de sistemas cíclicos (CSR) y el armazón de Murcko; análisis de complejidad molecular mediante métricas como fracción de carbonos sp³, fracción de carbonos quirales, fracción de átomos y enlaces aromáticos, flexibilidad molecular, índice de forma y globularidad; además de simulaciones de acoplamiento molecular con la proteína NSP-15 del SARS-CoV-2. Los

resultados mostraron que las bases de datos LAIPNUDELSAV y UPMA_2V presentaron la mayor diversidad estructural, NUBBE_1V destacó por su complejidad molecular y UPMA_2V junto con NUBBE_1V registraron la mayor diversidad de andamiajes. Se identificaron andamiajes recurrentes como benzopirranonas, triterpenoides pentacíclicos y agarofuranos con alto potencial farmacológico. En los análisis de *docking*, los compuestos LAIPNUDELSAV_029, LAIPNUDELSAV_031 y UPMA_2V_0266 mostraron alta afinidad y estabilidad de interacción frente a la proteína NSP-15, superando incluso al ligando de referencia Tipiracil. Los autores concluyeron que el uso combinado de análisis de diversidad, métricas de similitud, complejidad molecular y acoplamiento permitió identificar compuestos naturales latinoamericanos con prometedor potencial antiviral, demostrando la utilidad de enfoques quimioinformáticos integrados para la exploración virtual de compuestos bioactivos.

Fatima *et al*¹³ Realizó en el 2020 el trabajo titulado: “Perfiles ADMET de fitoquímicos geográficamente diversos mediante metodologías quimioinformáticas”. realizaron un estudio quimioinformático para evaluar las propiedades de Absorción, Distribución, Metabolismo, Excreción y Toxicidad (ADMET) en 3,158 fitoquímicos de bases de datos geográficamente diversas (India, Brasil, Sudáfrica) y 38 fármacos de referencia. La metodología incluyó el preprocesamiento molecular y la detección virtual, lo que permitió identificar 24 compuestos con perfiles farmacocinéticos óptimos, cinco de los cuales ya se sabía que poseían actividad anticancerígena. En conclusión, el trabajo validó este enfoque para el descubrimiento de fármacos y resultó en la creación de la base de datos de acceso abierto ADMET-BIS para facilitar la identificación de moléculas prometedoras con perfiles ADMET favorables.

Chen y Kirchmair¹⁴, en su estudio publicado el 2020: “La quimioinformática en el descubrimiento de fármacos basados en productos naturales” su objetivo fue investigar el uso de la quimioinformática en el desarrollo de fármacos de productos naturales, con énfasis en cómo las herramientas computacionales pueden facilitar la identificación y optimización de compuestos biológicamente activos. La metodología incluyó el uso de herramientas como el análisis de propiedades fisicoquímicas, PCA, evaluación de *scaffolds* y métodos de similitud 2D y 3D permiten caracterizar el espacio químico y priorizar compuestos con potencial bioactividad. Los autores concluyeron

que la integración de datos y el uso de aproximaciones tridimensionales son esenciales para comprender la complejidad estructural de los productos naturales y acelerar la identificación de candidatos farmacológicos. Este aporte se vinculó directamente con la presente investigación, ya que se emplearon los mismos enfoques análisis fisicoquímico, PCA, *scaffolds* y métricas de similitud como PheSA score y ESP-Sim para comparar metabolitos antiinflamatorios de plantas peruanas con fármacos y productos naturales de bases de datos internacionales.

Zhang *et al*¹⁵, en el artículo realizado en el 2019 nombrado: “Espacio químico y red de dianas biológicas de productos naturales antiinflamatorios” tuvieron como objetivo explorar el espacio químico, la diversidad estructural y los blancos biológicos de compuestos naturales con actividad antiinflamatoria, mediante herramientas de quimioinformática que permitieron comprender su potencial farmacológico multiobjetivo (*multitarget*). Los autores recopilaron 665 compuestos naturales a partir de 362 artículos científicos y construyeron la base de datos InflammNat. La metodología incluyó el análisis de propiedades fisicoquímicas según los criterios de Lipinski, la evaluación del espacio químico mediante clustering jerárquico, análisis de componentes principales (PCA) y la generación de andamiajes de Bemis–Murcko. Para cuantificar la similitud estructural y la diversidad molecular, emplearon huellas digitales moleculares MACCS y ECFP-6, junto con los coeficientes de similitud Tanimoto y Cosine, donde el primero midió la similitud estructural basada en la intersección de características moleculares presentes en los *fingerprints*, permitiendo identificar compuestos altamente similares o estructuralmente diversos. Además, se construyó una red compuesto-blanco usando datos de PubChem BioAssay para analizar los mecanismos de acción y los blancos moleculares asociados. Los resultados indicaron que los compuestos presentaron alta diversidad estructural, destacando los flavonoides y triterpenoides; además, el 66,4 % cumplió con tres o más criterios de Lipinski, mostrando buen potencial farmacológico. Se evidenció que el 72 % de los compuestos estaba asociado con más de un blanco molecular, demostrando un marcado carácter *multitarget* frente a proteínas como COX, HSD, VDR y ROR γ . Se concluyó que los productos naturales con actividad antiinflamatoria representan una rica fuente para el diseño de fármacos dirigidos a múltiples blancos terapéuticos y que todavía

existe un importante espacio químico sin explorar, ofreciendo oportunidades para el descubrimiento de nuevos andamiajes bioactivos.

En el 2019 Saldívar *et al*¹⁶, en su investigación titulada “Espacio químico y diversidad de la base de datos NuBBE: una caracterización quimioinformática” se tuvo como objetivo caracterizar cuantitativamente la diversidad química y la complejidad estructural de la base de datos NuBBEDB, que contenía compuestos naturales provenientes de la biodiversidad brasileña, y compararla con otras bases de datos de productos naturales y fármacos aprobados. Para ello, se empleó una metodología quimioinformática basada en la evaluación de propiedades fisicoquímicas relevantes, análisis del espacio químico mediante componentes principales (PCA), determinación de la complejidad molecular utilizando descriptores tales como el FCsp³ (fracción de carbonos sp³, que representa el grado de tridimensionalidad y complejidad estructural de una molécula) y el FCC (fracción de carbonos quirales, que indica el porcentaje de carbonos estereogénicos asociados con la complejidad y actividad biológica potencial), además del peso molecular, TPSA y número de enlaces aromáticos. Asimismo, se analizó la diversidad de andamiajes moleculares usando la entropía de Shannon y las curvas CSR, y se evaluó la similitud química mediante huellas moleculares ECFP4 y MACCS Keys con el coeficiente de Tanimoto. También se identificaron *scaffolds* únicos mediante comparación cruzada con bases como ChEMBL y *Dictionary of Natural Products*. Los resultados mostraron que NuBBEDB presentó una amplia diversidad estructural basada en huellas moleculares, aunque baja diversidad de andamiajes; además, el 12,28 % de sus *scaffolds* no fue encontrado en ChEMBL y 36 no habían sido reportados previamente en productos naturales, evidenciando un alto potencial para el descubrimiento de nuevos fármacos. Se concluyó que NuBBEDB fue una fuente prometedora para la investigación en química medicinal y diseño de bibliotecas químicas, debido a su diversidad estructural, complejidad molecular y presencia de núcleos base únicos compatibles con criterios de oralidad y biodisponibilidad.

Naveja *et al*¹⁷ en el trabajo de investigación publicado en el 2018: “Análisis de una gran base de datos de sustancias químicas alimentarias: espacio químico, diversidad y complejidad”, cuyo objetivo fue evaluar el espacio químico, diversidad y la complejidad estructural del repositorio FooDB, que almacena casi 24000 compuestos

químicos. Para lograrlo, emplearon la herramienta ChemMaps para analizar propiedades fisicoquímicas, complejidad molecular y la base estructural de los compuestos. Los resultados mostraron que FooDB presenta una notable diversidad y alta complejidad química, con un tercio de las moléculas siendo acíclicas y la mayoría de las cíclicas siendo monocíclicas, cuyas estructuras son similares a productos naturales de otras bibliotecas. El estudio concluyó que la alta diversidad y las similitudes estructurales de los compuestos de FooDB con otros productos naturales sugieren un nuevo y prometedor campo de investigación para el descubrimiento de potenciales agentes terapéuticos.

Paguigan et al,¹⁸ en el 2017 presentó el estudio titulado: “Fluoración quimioselectiva y análisis quimioinformático de griseofulvina: metabolitos fúngicos naturales vs. Fluorados”. Se realizó con el objetivo de aislar la griseofulvina y sus análogos relacionados a *Xylaria cubensis*, además, de sintetizar sus análogos fluoronados, para poder comparar su espacio químico mediante quimioinformática para así evaluar sus actividades anticancerígenas y antifúngicas. Se aislaron y caracterizaron 6 compuestos naturales (incluyendo 4 inéditos), lo análogos naturales y fluorados ocuparon regiones complementarias según PCA. Dos derivados fluorados se encontraron en el mismo espacio químico de la griseofulvina original. En conclusión, los compuestos fluorados amplían la investigación en PCA y al ser los dos compuestos con más activos los que ocuparon el mismo lugar químico sugiere que las modificaciones realizadas deben preservar las áreas relevantes del compuesto para expresar su actividad.

En el 2016 González et al¹⁹, en su publicación denominada “Exploración quimioinformática del espacio químico de productos fúngicos”. realizaron un estudio quimioinformático orientado a caracterizar el espacio químico de 207 metabolitos fúngicos aislados en un proyecto de descubrimiento de compuestos citotóxicos. El objetivo principal consistió en evaluar la complejidad estructural, diversidad molecular y propiedades fisicoquímicas de estos metabolitos, comparándolos con diversos conjuntos de referencia, incluyendo fármacos aprobados (oncológicos y no oncológicos), compuestos en ensayos clínicos, moléculas de screening general y sustancias Generalmente Reconocidas como Seguras (GRAS). Para ello, emplearon métricas de complejidad molecular (fracción de carbonos sp³, fracción de centros

quirales y globularidad), huellas moleculares (MACCS y ECFP4) para determinar diversidad estructural, y seis propiedades fisicoquímicas relevantes (HBA, HBD, logP, MW, TPSA y número de enlaces rotables). Los datos fueron analizados mediante El Diagrama de Caja y bigotes (*boxplots*), pruebas estadísticas de Kruskal Nemenyi y análisis de componentes principales. Los resultados mostraron que los metabolitos fúngicos presentaron mayor complejidad estructural que los fármacos aprobados, compuestos en etapa clínica y bibliotecas de screening, incluso alcanzando niveles comparables a los compuestos GRAS. Además, demostraron una diversidad estructural considerable, similar a la observada en otras bases de datos de productos naturales. En cuanto al perfil fisicoquímico, los metabolitos analizaron exhibieron propiedades cercanas al espacio drug-like (parecido a un fármaco en sus propiedades fisicoquímicas), especialmente en comparación con los medicamentos oncológicos aprobados. Los autores concluyeron que los metabolitos fúngicos eran buenos candidatos para ampliar el espacio químico medicinal, porque presentaban alta complejidad, buena diversidad y propiedades tipo fármaco. También identificaron diferencias claras entre fármacos oncológicos y no oncológicos y plantearon que una mayor complejidad molecular podría relacionarse con mayor selectividad y menor toxicidad.

2.1.2. Antecedentes nacionales

Barazorda *et al*²⁰. En la investigación publicada el 2023 titulada “PeruNPDB: la base de datos de productos naturales peruanos para el análisis *in silico* de fármacos”, diseña una plataforma para recolectar productos naturales (PN) del Perú para apoyar la investigación sobre el desarrollo de fármacos y el uso sostenible de la biodiversidad. Mediante una revisión sistemática de 399 publicaciones relevantes (1950-2021) y el procesamiento de datos de bases de datos como PubChem y DrugBank, se recolectaron 280 compuestos, de los cuales el 95% eran de origen vegetal. Se analizaron las propiedades fisicoquímicas y las similitudes de los medicamentos y se encontró que muchos compuestos eran estructuralmente similares a los medicamentos aprobados, destacando su potencial terapéutico. Además, la visualización del espacio químico mostró una rica diversidad y conexiones con otras bases de datos.

En el año 2021 Goyzueta *et al*.²¹ En el estudio nombrado “Actividad antiviral de metabolitos de plantas peruanas contra el SARS-CoV-2: un enfoque *in silico*” Este

trabajó tuvo como objetivo seleccionar plantas peruanas con actividad antiviral contra tres blancos de SARS-CoV-2. Se usó como metodología *in silico*, cribado virtual, acoplamiento molecular, simulación de dinámica molecular y estimación de MM/GBSA para evaluar la interacción de los compuestos con sus dianas. Tuvo como resultado el sistema de Mpro-Rutina que mostró una alta energía libre de Gibbs a comparación de los otros sistemas. En conclusión, el estudio aporta evidencia del potencial terapéutico de las plantas del Perú contra el SARS-Cov-2.

2.2. Marco teórico

2.2.1. *Productos naturales*

Los productos naturales son compuestos químicos producidos por organismos vivos, como plantas, hongos, bacterias o animales, que no realizan funciones básicas necesarias para el crecimiento, desarrollo o reproducción, productos que actúan como metabolitos secundarios. Estas moléculas suelen tener configuraciones quirales específicas y una considerable diversidad estructural. Se producen de forma natural, pero también pueden ser sintetizados o producidos químicamente por organismos genéticamente modificados si son químicamente equivalentes a los compuestos naturales originales.¹³ Los productos naturales se distinguen por sus propiedades biológicas y farmacológicas, lo que los convierte en una fuente fundamental para el descubrimiento de fármacos y nutracéuticos. Su investigación es fundamental para el campo de la farmacología, examinando sus propiedades medicinales, ecológicas y funcionales.²²

2.2.2. *Actividad antiinflamatoria*

La inflamación es un proceso natural del organismo ante alguna lesión o afección, sin embargo, cuando se vuelve persistente produce daños a los tejidos y órganos. La terapia antiinflamatoria actual tiene como objetivo controlar los síntomas inflamatorios clave al contrarrestar o bloquear los mediadores inflamatorios liberados al inicio de la respuesta inflamatoria aguda.⁵ Los AINE generalmente reducen la inflamación y el dolor asociado al inhibir la enzima ciclooxigenasa involucrada en la producción de prostaglandinas. Estas enzimas existen en dos isoformas (COX-1 y COX-2).⁶ Por lo tanto, los compuestos que inhiben las enzimas COX pueden considerarse fármacos antiinflamatorios potenciales. Sin embargo, Sin embargo, hoy se considera también a la COX-3, una variante de COX-1 que retiene un

intrón y se expresa en tejidos como el cerebro y el corazón, sensible a analgésicos como el paracetamol, aunque su rol en la inflamación crónica es menos dominante que el de COX-2 y podría explicar efectos antitérmicos con menor impacto gastrointestinal.²³ Muchos medicamentos antiinflamatorios de uso común son cada vez menos aceptables debido a los efectos secundarios graves, como la intolerancia gástrica, la supresión de la médula ósea y la retención de agua y sal debido al uso a largo plazo.⁷ En este sentido, es de fundamental importancia buscar sustancias que puedan contribuir a la resolución de la inflamación, es decir, homeostasis y regulación, efectivas y toleradas por el organismo. Existen numerosas investigaciones que muestran dicha función en animales de experimentación. Empero, en pocas plantas se ha logrado elucidar los compuestos activos y sus blancos terapéuticos.

2.2.3. Plantas peruanas con actividad antiinflamatoria

Se conoce que en el Perú existen muchos usos tradicionales de las plantas de esta región, además que constituyen una fuente rica de metabolitos con actividad bioactiva que muestran efectos antiinflamatorios e inmunomodulares comprobados en estudios *in vitro*, *in vivo*. Estas suelen estar relacionadas con compuestos fenólicos (Flavonoides, taninos, ácidos fenólicos), terpenoides y alcaloides que modulan vías inflamatorias.²⁴ En los últimos años fue incrementado las investigaciones científicas sobre estas plantas y los beneficios que traen. A continuación, se detalla una selección de plantas peruanas con actividad antiinflamatoria conocida.

- *Curcuma longa L.*: Esta especie, ampliamente estudiada, contiene curcuminoides (curcumina, demetoxicurcumina, bisdemetoxicurcumina)²⁵ cuya acción moduladora de la inflamación ha sido documentada por numerosos autores. Por ejemplo, se ha demostrado que este extracto inhibe la vía de señalización NF- κ B, disminuye la expresión de COX-2, iNOS y citocinas proinflamatorias (IL-1 β , IL-6, TNF- α) en modelos celulares y animales.²⁶
- *Punica granatum L.*: Cultivada en valles peruanos, sus cáscaras y semillas contienen punicalaginas y elálgico ácido, con efectos antiinflamatorios por inhibición de COX-2 y NF- κ B en modelos de edema.²⁷
- *Hibiscus sabdariffa L.*: Común en regiones costeras peruanas, sus cálices exhiben antocianinas que reducen TNF- α e IL-6, con potencial en inflamación respiratoria.²⁸

- *Solanum tuberosum L.*: Es una planta nativa de los Andes peruanos, variedades pigmentadas como Vitelotte muestran actividad antiinflamatoria por antocianinas que inhiben COX-2 y iNOS.²⁹
- *Plantago lanceolata L.*: Usada en la sierra peruana para heridas y bronquitis, sus iridoides (aucubina) modulan PGE2 y TXA2, y un notable contenido fenólico, así como actividad antioxidante, antiinflamatoria (potencial de inhibición de COX-1 y 12-LOX) y citotóxica. con efectos en inflamación respiratoria.³⁰
- *Syzygium aromaticum (L.) Merr. & L.M. Perry*: Introducido en Perú, su aceite esencial (eugenol) inhibe NF-κB y citoquinas, útil para dolor dental y artritis.³¹ Además tradicionalmente los botones florales del *Syzygium aromaticum* (clavo de olor) se utilizan machacados como enjuague bucal o masticados para aliviar el dolor de muela. Se les atribuyen propiedades analgésicas, anestésicas, antiinflamatorias, antioxidantes, antisépticas, digestivas y antieméticas.³²
- *Psidium guajava L.*: Nativa de regiones tropicales peruanas, sus hojas ricas en quercetina exhiben inhibición de COX-2 y TNF-α, efectiva contra diarrea inflamatoria.³³ También es usada en la medicina popular por sus componentes activos que ayudan a tratar diversas enfermedades. Diferentes partes de la planta se emplean tradicionalmente para aliviar afecciones digestivas, infecciones, heridas y dolores como diarrea, úlceras, dolor de muelas, tos e inflamaciones.³⁴

2.2.4. Metabolitos antiinflamatorios y sus grupos funcionales

La actividad antiinflamatoria de un compuesto está íntimamente ligada a su estructura química, un concepto conocido como Relación Estructura-Actividad (SAR). Los grupos funcionales definen esta actividad, ya sea en fármacos sintéticos (AINE) o en metabolitos secundarios naturales.

Tabla 1

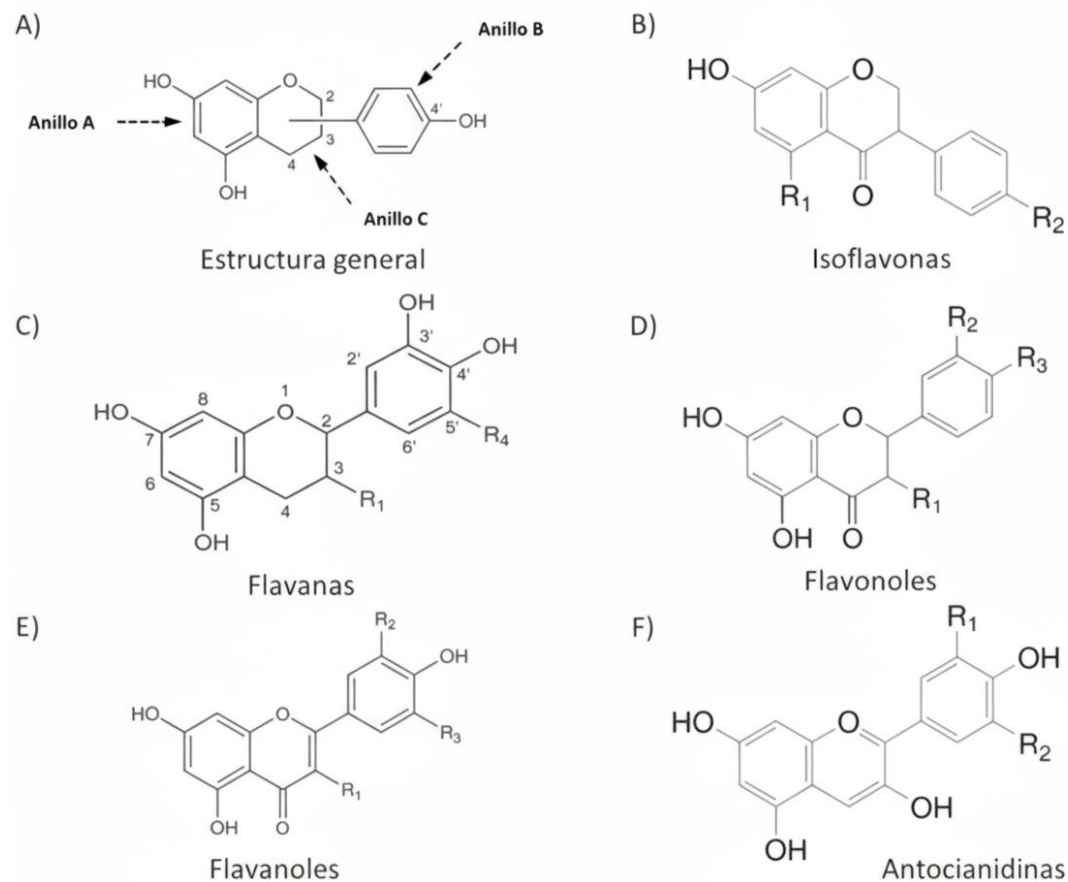
Tipos de metabolitos con actividad antiinflamatoria, sus grupos funcionales y mecanismo de acción.

Clase de Metabolito	Metabolito Representativo	Grupos Funcionales y Actividad Clave
Compuestos Fenólicos	Proantocianidinas	Compuestos derivados de catequinas mediante enlaces C-C. La actividad antiinflamatoria se atribuye principalmente a

Clase de Metabolito	Metabolito Representativo	Grupos Funcionales y Actividad Clave
(Táninos condensados)		su fuerte capacidad antioxidante y captadora de radicales libres, además de ser antagonistas de receptores hormonales o inhibidores de enzimas como COX.
Flavonoides (Flavonoles)	Quercetina	Polifenoles con múltiples grupos Hidroxilo (-OH). Se encontró que su actividad antiinflamatoria es significativa, especialmente cuando contienen un anillo B de tipo catecol o guayacol en su estructura. Bloquean tanto las vías de la ciclooxigenasa (COX) como de la lipooxigenasa (LOX), y modulan citocinas y el factor nuclear NF- κ B. ³⁵
Flavonoides (Flavanona)	Hesperidina	Posee un grupo rutinósido (un disacárido) unido a la estructura base. Exhibe actividad antiinflamatoria en modelos de edema sin producir los efectos secundarios de otras clases de antiinflamatorios. ³⁵
Cumarinas	Columbianetina y Libanoridina	Estructuras que contienen un grupo lactona cíclica y pueden poseer otros grupos funcionales (ej. Hidroxilo, Éster). Se ha reportado que poseen actividad antioxidante (capturando radicales superóxidos) y pueden inhibir las vías de la lipooxigenasa y ciclooxigenasa del metabolismo del ácido araquidónico. ³⁵
Alcaloides (Isoquinolinas)	Berberina	Compuestos que contienen Nitrógeno dentro de un anillo, se ha demostrado que las isoquinolinas (como la berberina) tienen una buena actividad antiinflamatoria. Otros como los bisbencilisquinolina (Tetrandrina) previenen la síntesis o acción de citocinas proinflamatorias. ³⁵
Saponinas	Saponinas triterpenoides y esteroidales	Son glucósidos de esteroides o triterpenos. Actúan a través de mecanismos como la actividad corticomimética (indirecta o directa) y la inhibición de la formación y liberación de mediadores de la inflamación. La estructura triterpenoide es clave para su actividad. ³⁵

Figura 1

Estructura de los principales flavonoides.



Nota: La diferencia entre ellos radica principalmente en la presencia y posición del grupo $-OH$ y en el grado de saturación del anillo C. tal como se observan en la figura 1. Tomado de Salih *et al.*³⁵ (2014).

2.1.1. Historia de la Quimiinformática

La quimiinformática nació como una respuesta a la necesidad de analizar y organizar grandes volúmenes de información química mediante computadoras. Sus orígenes se remontan a mediados del siglo XX, cuando los avances en mecánica cuántica y la digitalización de estructuras permitieron realizar cálculos moleculares que antes eran impracticables. Con el tiempo, se desarrollaron métodos como QSAR, bases de datos químicas, descriptores moleculares y algoritmos de similitud, consolidando la disciplina. Actualmente, la quimiinformática integra estadística, modelado molecular y aprendizaje automático, convirtiéndose en una herramienta esencial para el diseño de

fármacos y la exploración del espacio químico.³⁶ A continuación se muestra un cuadro cronológico de los avances de la quimioinformática a través del tiempo.

Tabla 2

Cuadro cronológico de la Quimioinformática.

Año / Década	Hito Principal	Descripción Científica
1964	Nacimiento del QSAR	Hansch y Fujita, publican el primer modelo QSAR lineal múltiple ($\log P + \sigma + E_s$). ³⁷
1970s–1980s	Fingerprints y búsqueda de subestructuras	Desarrollo de los primeros fingerprints (MACCS, Daylight) y algoritmos de similitud (Tanimoto). ³⁸
1988	Creación del SMILES	Weininger, publica el sistema SMILES: primera notación lineal universal para estructuras químicas. ³⁹
1996	Definición formal de <i>scaffolds</i> (Murcko)	Bemis & Murcko, definen “molecular frameworks” (scaffolds sin cadenas laterales). ⁴⁰
1997	Regla de los Cinco (Lipinski)	Se establece el concepto de “drug-likeness” para filtrar bibliotecas químicas. ⁴¹
2004	Lanzamiento de PubChem	Mayor base de datos pública de compuestos químicos (actualmente >110 millones). ⁴²
2006	RDKit open-source	Greg Landrum, libera RDKit: la herramienta quimioinformática más usada actualmente. ⁴³
2011	ChEMBL versión pública	Base de datos curada de bioactividad (hoy >2,4 millones de compuestos). ⁴⁴
2015	Tanimoto como estándar de similitud	Revisión que valida el coeficiente de Tanimoto como métrica óptima para fingerprints. ⁴⁵
2018	Deep Learning en quimioinformática	Redes neuronales para generación de moléculas (VAE, GAN) y predicción de actividad. ⁴⁶
2022	ESP-sim: similitud electrostática 3D	Nuevo índice basado en mapas de potencial electrostático. ⁴⁷
2023–2025	AlphaFold más Quimioinformática	Integración de estructuras 3D predichas por IA en cribado virtual y docking masivo. ⁴⁸

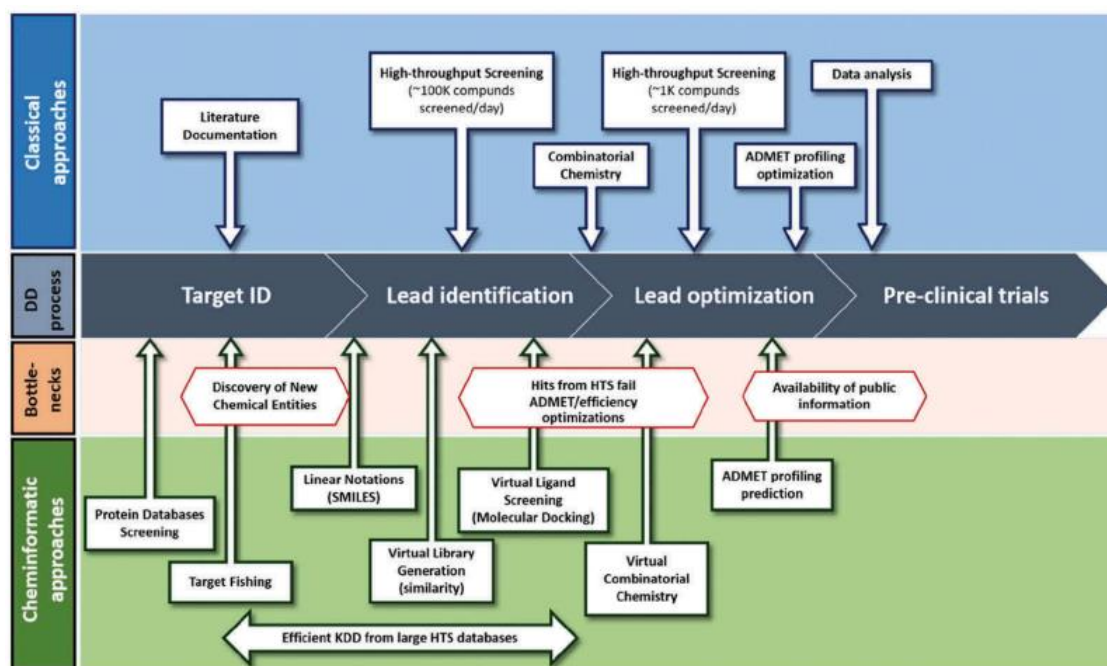
2.2.5. Quimioinformática

Combina recursos informáticos y datos químicos para poder gestionar, visualizar y analizar información química de forma sistemática.¹³ Facilita el procesamiento eficiente de una gran cantidad de datos, lo que permite la exploración del espacio químico, la extracción de datos y las relaciones matemáticas entre estructura y actividad. El desarrollo de modelos de relación cuantitativa estructura-actividad (QSAR, por sus siglas en inglés).²⁰ Además, la quimioinformática ayuda a encontrar patrones en datos

químicos, biológicos y clínicos, permitiendo realizar predicciones más detalladas en sistemas complejos sin la necesidad de simplificar demasiado, lo que permite un estudio más completo para el diseño de medicamentos. Asimismo, permite el uso de técnicas y metodologías innovadoras para mejorar las representaciones farmacofóricas basadas en grafos, el diseño de fármacos basado en fragmentos, en *novo*, las predicciones de energía de enlace y los modelos de consenso.¹⁹ Los temas más usados en la quimioinformática son: las propiedades fisicoquímicas, la diversidad y similaridad estructural, la complejidad estructural, los descriptores moleculares, además de la distribución en el espacio químico.⁴⁹

Figura 2

*Evolución de las estrategias para el descubrimiento de nuevos fármacos de forma más rápida y económica.*⁵⁰



Nota: Tomado de Olmedo et al.⁵¹ (2017).

El desarrollo y la mejora continua de metodologías y tecnologías de vanguardia que tienen como objetivo acelerar la identificación y validación de compuestos terapéuticos y de forma simultánea reducir significativamente los altos costos y el tiempo necesarios para que un nuevo medicamento llegue desde la fase de laboratorio hasta el mercado.⁵⁰

2.1.2. *Smiles (Simplified Molecular Input Line Entry System)*

Es un lenguaje de notación química lineal que permite representar la estructura molecular completa de un compuesto mediante una secuencia compacta de caracteres alfanuméricos y símbolos, codificando átomos, enlaces, ramificaciones, ciclos, aromaticidad y estereoquímica sin requerir coordenadas espaciales o gráficos.³⁹Desarrollado para facilitar el almacenamiento, recuperación y manipulación computacional de estructuras químicas en bases de datos, un SMILES canónico es único para cada molécula, mientras que versiones isoméricas incluyen información de configuración, por ejemplo, las moléculas lineales. En quimioinformática, los SMILES son la entrada estándar para generar *fingerprints moleculares*, calcular similitud de Tanimoto y realizar cribado virtual⁵¹.

2.1.3. *Colecciones virtuales de productos naturales bases de datos*

La última década mostró un gran aumento de los datos que dan acceso a datos químicos, biológicos, toxicológicos y estructurales de los productos naturales (PN). La mayoría de las bases de datos de estructuras también ofrecen descargas masivas gratuitas, lo que permite la selección y las aplicaciones virtuales. De estos estudios se deduce que el número total de PN cuyas estructuras pueden obtenerse mediante descargas masivas. Entre las bases de datos de productos naturales más conocidas se encuentran:

Tabla 3

Comparación de bases de datos en tipo, tamaño aproximado, contenido/enfoque y usos principales.

Base de datos	Tipo (Global/Regional/Especializada)	Tamaño aproximado	Contenido / Enfoque	Usos principales
NuBBE_DB	Regional (Brasil) / Especializada	2000–5000 compuestos (actualizaciones periódicas). ⁵²	Metabolitos de la biodiversidad brasileña, datos químicos y bioactividad. ⁵²	Quimiografía de biodiversidad, búsqueda de <i>leads</i> naturales, análisis de propiedades fisicoquímicas. ⁵²
SANCDDB	Regional (Sudáfrica)	1012 (actualizado). ⁵³	Compuestos naturales aislados de flora y vida	Investigación de metabolitos regionales, cribado in silico y

Base de datos	Tipo (Global/Regional/Especializada)	Tamaño aproximado	Contenido / Enfoque	Usos principales
			marina sudafricana, curada y referenciada. ⁵⁴	estudios etnobotánicos. ⁵⁴
DrugBank	Global/Farmacéutica (no exclusivamente NP)	Más de 30000 entradas (incluye fármacos aprobados y metabolitos). ⁵⁵	Base de datos de fármacos con información química, farmacológica y de dianas. ⁵⁵	Comparación fármaco-NP, punto de referencia de propiedades ADME/Tox, identificación de fármacos similares. ⁵⁵
KNAPSAcK	Global / Especializada (metabolito-especie). ⁵⁶	≥ 100000 relaciones metabolito-especie; 20000 metabolitos registrados. ⁵⁶	Relación metabolito-especie vegetal; útil en metabolómica y taxonomía química. ⁵⁶	Metabolómica, identificación de biomarcadores, estudios etnobotánicos. ⁵⁶
SuperNatural II	Global	≈ 326000 estructuras (bulk database). ⁵⁷	Colección extensa de productos naturales y análogos; diseñada para cribado virtual. ⁵⁷	Cribado virtual, generación de bibliotecas NP-like, análisis de espacio químico. ⁵⁷
Natural Products Atlas	Global	≈ 25000 microbially derivados NPs (bacterias y hongos), ampliada en versiones recientes. ⁵⁸	Catálogo curado de NPs microbianos con referencias y estructuras. ⁵⁸	Estudios de metabolitos microbianos, dereplicación, búsqueda de líderes antimicrobianos. ⁵⁸
TCM Database@Taiwan	Regional (TM Chino)/Especializada	≈ 20,000 compuestos (TCM). ⁵⁹	Compuestos aislados de ingredientes de la Medicina Tradicional	Cribado in silico dirigido a compuestos tradicionales,

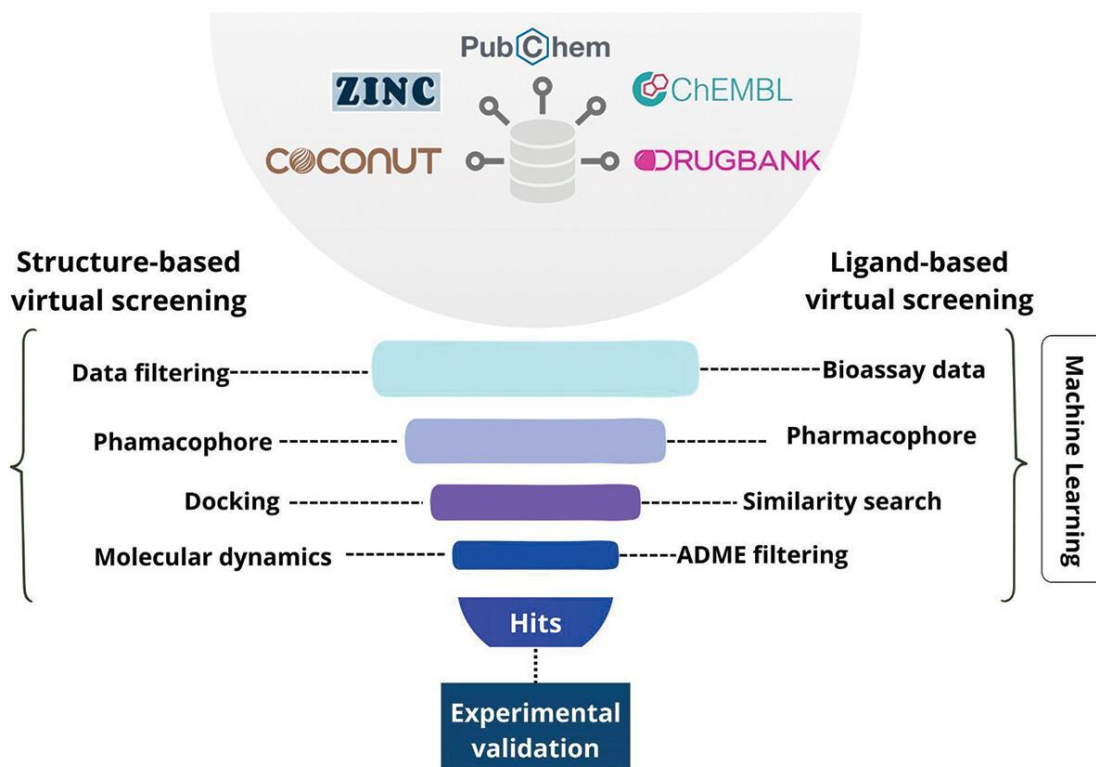
Base de datos	Tipo (Global/Regional/Especializada)	Tamaño aproximado	Contenido / Enfoque	Usos principales
			China; formatos 2D/3D disponibles. ⁵⁹	virtual screening. ⁵⁹
UNPD/COCONUT	Global	UNPD ≈ 153000; COCONUT y > 200000 agregados. ⁶⁰	Bases de datos agregadas de NPs de múltiples fuentes; grandes colecciones descargables. ⁶⁰	Mapeo amplio del espacio NP, de estudios estadísticos, minería de <i>scaffolds</i> . ⁶⁰
ZINC (NP-like subsets)	Global	Millones de Compuestos en ZINC; subconjunto variable tipo NP (10000– ⁶¹	Catálogo de compuestos comercialmente disponibles, incluye subcolecciones NP. ⁶¹	Fuente para adquisición de compuestos para pruebas experimentales; cribado virtual con compuestos obtenibles. ⁶¹

2.1.4. Productos naturales y quimioinformática

La recopilación sistemática de datos químicos y biológicos precisos es fundamental para el descubrimiento de fármacos. Las bases de datos científicas actúan como repositorios esenciales, pero su curación rigurosa es indispensable para garantizar la fiabilidad, la coherencia y la reproducibilidad de los estudios.^{62,63} Una curación de datos adecuada es vital porque permite una aplicación quimioinformática robusta, reduce los errores en la identificación de compuestos y acelera la validación biológica. En última instancia, la elección de la biblioteca de compuestos en un proyecto de Virtual screening (cribado virtual) depende del objetivo específico: explorar nuevas áreas del espacio químico, acceder a sustancias sintetizadas, o validar candidatos farmacéuticos.⁶⁴

Figura 3

*Esquema del Cribado Virtual: Enfoques Basados en Estructura y Ligandos.*⁶⁴



Nota: Tomado de Marbán *et al.*⁶⁴ (2025).

2.1.5. Métricas de Diversidad química

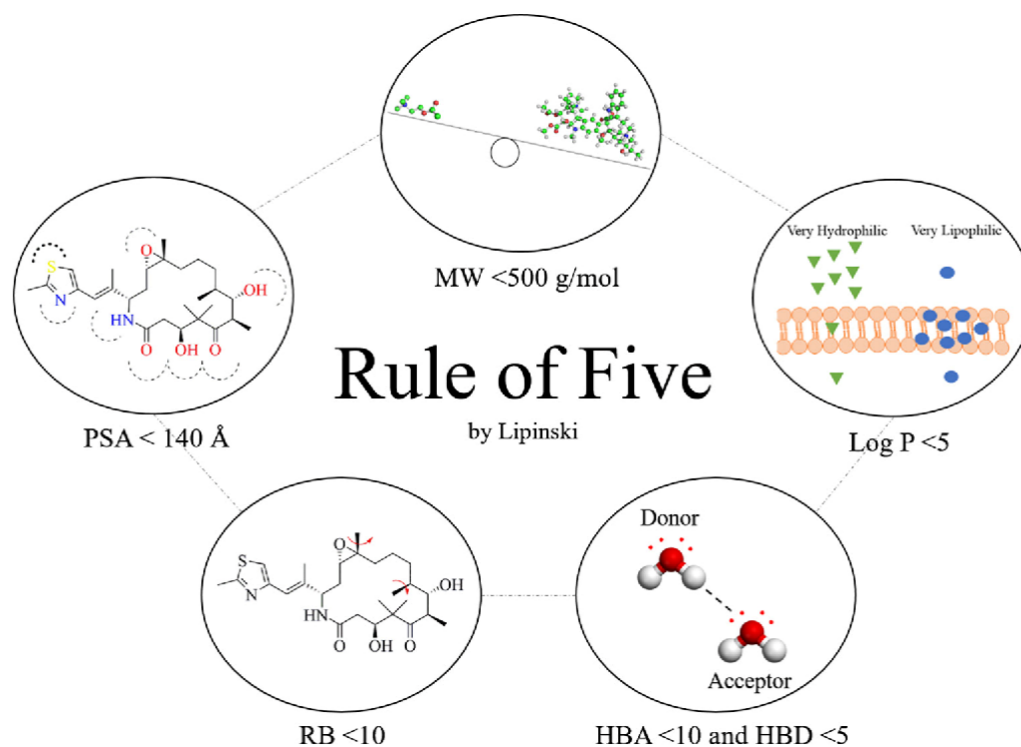
La diversidad química se define como la variabilidad estructural y funcional de los compuestos químicos presentes en un sistema biológico, esta diversidad puede evaluarse mediante descriptores moleculares como las propiedades fisicoquímicas, esqueletos carbonados, grupos funcionales, estereoquímica y grados de oxidación, lo que genera un amplio espacio químico con potencial farmacológico diferenciado.^{65,66}

2.1.2.1. Caracterización de las propiedades fisicoquímicas

Para la identificación y comparación espacial de propiedades de fármacos, propiedades de compuestos bioactivos y otras bases de datos moleculares aprobadas, se utilizan con frecuencia las propiedades fisicoquímicas, como las reglas de Lipinski, este consta de 6 descriptores moleculares: Donantes de enlace de hidrógeno (HBD), aceptores de enlaces de hidrógeno (HBA), coeficiente de partición (SlogP), área de superficie polar (TPSA), números de enlaces rotables (RB) y el peso molecular (MW).⁶⁷

Figura 4

Esquema visual de los parámetros de las Reglas de Lipinski: peso molecular (MW), área superficial polar (PSA), enlaces rotables (RB), aceptores de enlaces de hidrógeno (HBA), donantes de enlaces de hidrógeno (HBD), log P.⁶⁸



Nota: Tomado de Chagas.⁶⁸ (2018).

2.1.2.2. Propiedades fisicoquímicas y su relación con la biodisponibilidad

La biodisponibilidad oral de un fármaco depende en gran medida de sus propiedades fisicoquímicas, lo cual se relaciona estrechamente con las reglas de Lipinski o Regla de los Cinco, propuestas como un conjunto de criterios orientados a predecir una adecuada absorción intestinal.⁶⁹ Sin embargo, análisis posteriores realizados sobre más de 1100 compuestos en *SmithKline Beecham Pharmaceuticals* una compañía farmacéutica en EE.UU mostraron que otras propiedades, como el número de enlaces rotables (flexibilidad molecular) y el área superficial polar (TPSA), pueden ser predictores aún más robustos. Este estudio demostró que moléculas con ≤ 10 enlaces rotables y $TPSA \leq 140 \text{ \AA}$ (o ≤ 12 donadores + aceptores de hidrógeno) presentan mayor probabilidad de lograr buena biodisponibilidad oral en ratas, independientemente del peso molecular. Además, se observó que una menor TPSA correlaciona mejor con una mayor permeabilidad a membranas artificiales que la

lipofilicidad (cLogP), mientras que la alta flexibilidad molecular disminuye la permeabilidad. Estos hallazgos complementan la Regla de Lipinski al resaltar que la rigidez molecular y la reducción de la polaridad son factores críticos para asegurar la absorción oral.⁶⁹

2.1.2.3. Visualización del Espacio Químico

El espacio químico se define como el conjunto de todas las moléculas reales o posibles que pueden generarse bajo reglas químicas específicas, representado como un espacio multidimensional donde cada dimensión corresponde a una propiedad molecular.⁷⁰ Por ello, usa técnicas como el Análisis de Componentes Principales (PCA) para representar gráficamente la diversidad de compuestos en un espacio químico, permitiendo a los usuarios observar agrupaciones y patrones en los datos.⁷¹ Para explorar y comparar este espacio, se emplean propiedades fisicoquímicas como el peso molecular, la superficie polar topológica (TPSA), y el número de donadores y aceptores de puentes de hidrógeno, las cuales determinan en gran medida el comportamiento biológico y farmacocinético de una molécula.⁷²

2.1.2.4. Scaffolds

Se definen como los esqueletos moleculares centrales que definen la arquitectura química de una clase de compuestos, consistiendo en anillos cíclicos, sistemas fusionados o cadenas lineales que sirven como base para la unión de grupos funcionales y la interacción con dianas biológicas, permitiendo la exploración sistemática del espacio químico en el descubrimiento de fármacos.⁴⁰

2.1.3. Métricas de similitud

Es una metodología química e informática que evalúa el grado de similitud entre sustancias químicas en función de características estructurales, funcionales o biológicas. Este análisis puede ser general, centrándose en similitudes visuales, estructurales o específicas, centrándose en propiedades asociadas con una actividad biológica o toxicidad específica. La elección del enfoque depende del contexto y los objetivos del estudio, lo cual es importante para la agrupación química, la predicción de la toxicidad y otras aplicaciones *in silico*.⁷³ El principio de la similitud química, es la probabilidad de que los compuestos con estructuras similares tengan actividades biológicas similares, es una suposición clave del descubrimiento virtual basado en la similitud. La similitud estructural de dos moléculas generalmente se evalúa calculando

el coeficiente de Tanimoto (Tc) de la huella química, es una métrica que mide la similitud entre dos conjuntos de datos, evaluando la relación entre los elementos comunes (intersección) y el total de elementos únicos (unión).⁴⁵

2.1.3.1. Confórmeros y PheSA Score (Alineación avanzada de formas farmacóforas)

Un confórmero es una configuración diferente (disposición de átomos en una molécula), en la superficie de energía potencial de la molécula.⁴⁶ PheSA es un algoritmo que utiliza funciones gaussianas para calcular y optimizar el volumen total entre confórmeros de dos moléculas integrando al mismo tiempo la superposición de funciones cinemáticas farmacéuticas. Esta capacidad permite un análisis y una comparación precisos de las conformaciones moleculares, optimizando el proceso de desarrollo de fármacos al examinar cómo las diferentes disposiciones atómicas (conformadores) afectan la interacción biológica objetivo. De esta forma, PheSA combina el acoplamiento molecular con el análisis de propiedades farmacológicas, maximizando su aplicación en el diseño de estructuras de base química, para una adecuada similitud lo ideal es que el valor de PheSA sea mayor a 0,6.⁷⁴ Para realizar la generación y superposición de confórmeros se hace uso de herramientas computacionales como Data Warrior; Es un software gratuito y multiplataforma diseñado para analizar y visualizar datos tabulares, con un enfoque especial en química. Permite explorar conjuntos de datos, descubrir patrones y analizar relaciones entre estructuras químicas y propiedades. Incluye herramientas para filtrar datos dinámicamente, predecir propiedades moleculares y manejar estructuras químicas de manera intuitiva.⁶¹

2.1.3.2. Potencial electrostático

La energía electrostática es la energía potencial generada por la interacción entre cargas eléctricas en un campo eléctrico. Puede ser positiva si las cargas son del mismo signo (se repelen) o negativa si son de signos opuestos (se atraen).⁷⁵ El potencial también se ha utilizado con éxito para estudiar interacciones que involucran determinaciones de alguna orientación relativa óptima de electrófilos, los cuales tienen afinidad con los electrones, y es particularmente útil como indicador de los sitios o regiones de la molécula a los que se atrae inicialmente un electrófilo que se aproxima.⁷⁵

Se entiende de forma matemática, la energía electrostática U entre dos cargas puntuales Q_1 y Q_2 separadas por una distancia r se expresa mediante la fórmula:

$$U = k \frac{Q_1 Q_2}{r}$$

Dónde k es la constante de Coulomb. Esta energía juega un papel crucial en química y física, ya que afecta las interacciones entre átomos y moléculas, lo que a su vez influye en las propiedades químicas y físicas de las sustancias.⁴¹

- **ESP-Sim**

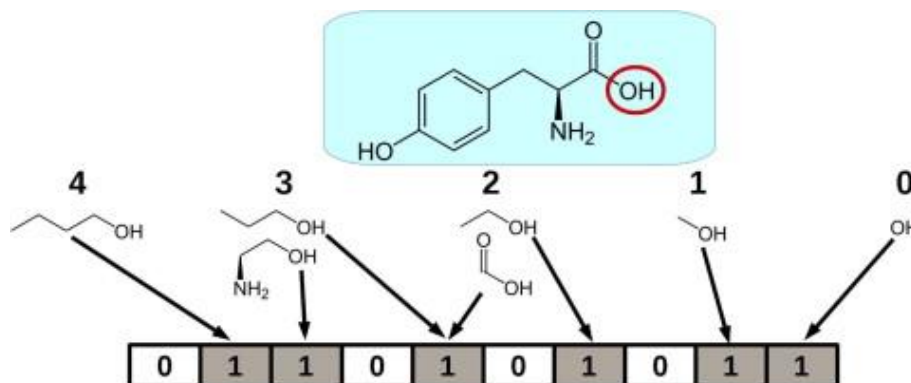
Es un índice de similitud diseñado para comparar mapas de potencial electrostático (ESP) de moléculas en quimioinformática. El ESP representa la distribución de cargas superficiales de una molécula, crucial para predecir interacciones moleculares como unión a proteínas o reactividad.⁴⁷ ESP-Sim utiliza la máxima discrepancia media (MMD), un estadístico de kernel de aprendizaje automático, para cuantificar la similitud entre dos mapas ESP, considerando tanto la amplitud como la forma de las distribuciones electrostáticas.⁴⁷ A diferencia de métricas tradicionales como Tanimoto (que se enfoca en estructura 2D), ESP-Sim captura similitudes 3D funcionales, siendo especialmente útil para evaluar análogos en diseño de fármacos donde las interacciones electrostáticas son clave.⁴⁷

2.1.3.3. Fingerprints y distancia de Tanimoto

Son representaciones digitales de estructuras químicas que permiten la comparación y análisis de moléculas en base a sus características estructurales. Estos *fingerprints* se utilizan ampliamente en química computacional y bioinformática para calcular la similitud entre compuestos. Son representaciones numéricas de las estructuras químicas, usualmente se representa como cadena de bits (secuencia de ceros y unos),²² como se muestran en la figura 5.

Figura 5

Huella dactilar de la Tirosina, representado como cadenas de bits.



Nota: Tomado de Krause.²² (2013).

Para poder realizar una comparación de similitud por fingerprint se usa el coeficiente de Jaccard o conocido también como distancia de Tanimoto, mide la similitud estructural entre dos compuestos mediante la relación entre la intersección y la unión de sus bits moleculares, donde un valor de 1 indica identidad estructural y 0 ausencia de similitud.⁴⁵ Siendo 0,6 el valor mínimo recomendado para una adecuada similaridad.⁷⁶

Tabla 4

Tabla resumen de las métricas de diversidad y similitud.

Métrica	Coefficiente de Tanimoto (Tc)	PheSA Score	ESPsim
¿Qué mide?	Similitud estructural basada en fingerprints 2D. ²²	Similitud combinada en farmacofórica electrostática 3D.	Similitud directa del potencial + electrostático molecular (ESP) en 3D. ⁴⁷
¿Qué compara?	Presencia/ausencia de subestructuras (bits). ²²	Puntos farmacofóricos + campo electrostático en esos puntos.	Valores del potencial electrostático en malla o superficie. ⁴⁷
Dimensionalidad	Principalmente 2D. ⁷⁷	3D obligatoria. ⁷⁴	3D obligatoria. ⁷⁸
Representación utilizada	Bitstrings o count vectors (ECFP4, MACCS, etc.). ⁷⁷	Puntos farmacofóricos + electrostáticos gaussianos. ⁷⁴	Malla 3D o superficie de van der Waals con ESP calculado QM/semiempírico. ⁷⁸
Complejidad computacional	Baja.	Media-alta (alineación + campos gaussianos). ⁷⁴	Alta (cálculo QM + integración 3D). ⁷⁸

Métrica	Coefficiente de Tanimoto (Tc)	PheSA Score	ESPsim
Sensibilidad electrónica	Nula o muy baja	Alta. ⁷⁴	Muy alta (máxima posible). ⁷⁹
Sensibilidad a conformación y alineación	Baja	Muy alta (requiere alineación óptima). ⁷⁴	Extremadamente alta. ⁸⁰
Captura bioisosterismo	Solo si los grupos generan patrones de bits similares		Excelente (incluso con esqueletos muy diferentes). ⁷⁹
Ventajas principales	Rapidez extrema, robustez, bases de datos masivas. ⁸¹	Buen balance de velocidad/sensibilidad electrónica. ⁷⁴	Máxima sensibilidad funcional y scaffold hopping real. ⁷⁹
Limitaciones	No captura electrónica real. ⁸¹	Necesita alineación 3D previa, menos sensible que ESP puro. ⁷⁴	Muy costoso, hipersensible a conformación y alineación (Armstrong et al., 2010)
Rango del score	0 – 1	0 – 1	0 – 1 ⁸² o 0 – 2 ⁸³
Interpretación típica	>0,60 para screening inicial; mediana ~0,6 en activas >0,85 ≈ similar (2D). ⁷⁶	>0,60 en alineamientos ≈ muy >0,70 para alta. ⁷⁴	para superposiciones 3D; >0,70 con alineación. ⁸⁴
Aplicaciones típicas	Screening virtual masivo, clustering, QSAR 2D.	<i>Scaffold hopping</i> moderado, series con restricciones farmacofóricas. ⁷⁴	Bioisosterismo avanzado, miméticos no clásicos, SAR electrónico.

2.1.4. Consensus score

Cada puntuación individual tiende a tener puntos fuertes y débiles, siendo aún más complejo el uso de dos puntuaciones a más y una manera de mejorar estos resultados dando mayor robustez y confianza es utilizando el puntaje consenso.⁸⁵ Este se basa en:

- Normalización de puntuaciones: Ayuda a estandarizar la integración de múltiples datos.⁸⁶ Las mediciones se encontrarán en un dominio de [0,1].⁸⁷
- Media geométrica: Disminuye significativamente el número de falsos positivos en un conjunto de datos, mejorando las tasas de acierto.⁸⁸

c) Índice de acuerdo: Da mayor consistencia entre metodologías, si un compuesto tiene varios métodos que lo avalan o menos, este índice ayuda a dar mayor o menor peso en su puntaje.⁸⁹ Este método es conocido como votación.⁹⁰

2.1.5. *Elbow method*

Este método detecta el número adecuado de agrupaciones en el algoritmo conocido como k-medias. Ayuda a detectar el error (distorsión) a medida que aumenta los clústeres. Al principio el error disminuye notoriamente, sin embargo, llega un momento en que no puede mejorar más.⁹¹ Este punto se conoce como inflexión o codo (*elbow*), definiendo así el número óptimo de agrupaciones.⁹¹

2.2. Marco conceptual

2.2.2. *Metabolitos secundarios*

Compuestos orgánicos no esenciales para el crecimiento vegetal, pero clave en defensa y adaptación; fuente principal de fármacos.³⁵

2.2.3. *Inflamación*

Proceso fisiopatológico mediado por leucocitos, macrófagos y mastocitos que genera edema, dolor y rubor como respuesta a lesión.³⁵

2.2.4. *Actividad antiinflamatoria*

Capacidad de un compuesto para inhibir mediadores (PGE₂, TNF- α , IL-6) o vías (COX, NF- κ B, PLA₂).³⁵

2.2.5. *RENATI*

Es el Registro Nacional de Trabajos de Investigación del Perú (RENATI), es una plataforma digital de almacenamiento gestionada por la SUNEDU, destinada a resguardar las tesis y trabajos de investigación.

2.2.6. *Peso molecular*

El compuesto que se encuentra como candidato es recomendable que tenga un peso menor a 500 Daltons, para que pueda cumplir con los parámetros establecidos.¹³

2.2.7. *LogP*

Se puede determinar el transporte de las moléculas en la membrana lipídica aumentando la hidrofobicidad o lipofilicidad, este parámetro se mide con LogP, este debe ser mayor -2 y menor a 5.⁹²

2.2.8. Número de HBA y HBD

Para que un compuesto forme enlaces de hidrógeno, influyen sus valores de HBA y HBD. Este parámetro afecta a la polaridad y permeabilidad. Debido a que un número alto de enlaces de hidrógeno produce una menor permeabilidad, disminuyendo la solubilidad.⁹³ Un potencial agente terapéutico debe cumplir con los siguiente: $HBA \leq 10$ y $HBD \leq 5$.¹³

2.2.9. Número de enlaces rotables

Este parámetro define la flexibilidad molecular y la permeabilidad. Si un compuesto tiene más de 10 enlaces rotables, generaría mayor flexibilidad y, por lo tanto, disminuiría la permeabilidad del activo.⁶⁹

2.2.10. TPSA

Para indicar un parámetro de permeabilidad, es necesario estudiar el TPSA (Área superficial topológica). A mayor valor de TPSA menor absorción del compuesto.⁹⁴ Los valores adecuados en los que se deba encontrar son de 7 a 200 Å.¹³

2.2.11. Boxplot o diagrama de caja y bigotes

Es una representación gráfica propuesta originalmente por John Tukey⁹⁵ como parte del análisis exploratorio de datos. Este tipo de gráfico resume visualmente la distribución de un conjunto de datos, permitiendo identificar la tendencia central, la dispersión y la presencia de valores atípicos.

2.2.12. R studio

Es un entorno de desarrollo integrado orientado al lenguaje R, utilizado ampliamente para análisis estadístico y visualización de datos. Permite ejecutar scripts, gestionar paquetes y generar gráficos de manera reproducible, lo cual facilita el tratamiento y exploración de datos fisicoquímicos.⁹⁶

2.2.13. Jupyter Notebook

Es una plataforma interactiva que combina código, texto y gráficos en un mismo documento, permitiendo documentar y ejecutar análisis paso a paso. Su uso es frecuente en quimioinformática debido a su compatibilidad con Python y bibliotecas como RDKit para el cálculo de descriptores moleculares.⁹⁷

2.2.14. Google Colab

Es un entorno basado en la nube que permite ejecutar notebooks similares a Jupyter sin necesidad de instalación local, ofreciendo acceso a recursos

computacionales remotos y permitiendo instalar librerías especializadas para análisis químico. Su accesibilidad y capacidad de procesamiento lo hacen útil para análisis quimioinformáticos con volúmenes grandes de datos.⁹⁸

2.2.15. Test de Wilcoxon

Prueba estadística no paramétrica utilizada para comparar si dos muestras independientes provienen de la misma distribución continua cuando no se cumplen los supuestos de normalidad o homocedasticidad.⁹⁹

2.3. Marco ético y legal

El presente estudio se sustenta en la normativa peruana aplicable a la investigación científica. Se acoge a los lineamientos establecidos por el Código Nacional de Integridad Científica, aprobado mediante Resolución N.º 028-2024-CONCYTEC-P, el cual establece principios, buenas prácticas y obligaciones para garantizar la honestidad, rigurosidad, transparencia y responsabilidad en todas las fases de la investigación científica. Dicho Código se enmarca en la Ley Marco de Ciencia, Tecnología e Innovación Tecnológica N.º 28303 y la Ley del Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica N.º 28613, que designan al CONCYTEC como ente rector del SINACYT y responsable de promover y supervisar la integridad científica en el país. Asimismo, la investigación respeta lo dispuesto por la Ley Universitaria N.º 30220 en lo relativo a ética académica, integridad, transparencia y uso adecuado de información.

CAPÍTULO III. MATERIALES Y MÉTODOS

3.1. Alcance de investigación

Descriptiva.

3.2. Diseño de investigación

Diseño comparativo con enfoque computacional.

3.3. Unidad de análisis

Metabolitos de plantas peruanas en código SMILES.

3.4. Población de estudio

Plantas peruanas que reportan actividad antiinflamatoria en la base de datos de RENATI desde 1998 hasta febrero 2025.

3.5. Muestra

Plantas peruanas que reportan actividad antiinflamatoria en base de datos RENATI y que registran información de metabolitos en la plataforma KNApSAcK.

3.6. Criterios de selección

3.6.1. Criterios de inclusión

Plantas peruanas que registran actividad antiinflamatoria en RENATI y registran información metabolitos en código SMILES en la plataforma KNApSAcK.

3.6.2. Criterios de exclusión

Plantas peruanas que registran actividad antiinflamatoria en el repositorio RENATI desde 1998 hasta febrero del 2025 y que no cuentan con códigos SMILES en KNApSAcK.

3.7. Técnicas e Instrumentos de Recolección de Datos

Se empleó programas de código abierto, *Jupyter Notebook*, *Rstudio* y *Data Warrior* para integrar, procesar y analizar información estructural de metabolitos con actividad antiinflamatoria para la realización del siguiente análisis.

- Análisis exploratorio: Espacio químico (PCA) y *scaffolds*. Para caracterizar la distribución estructural de los metabolitos.

- Análisis de diversidad y similaridad usando las métricas: coeficiente de Tanimoto, *Phesa Score* y *ESP-sim*. Con la finalidad de comparar metabolitos frente a fármacos y productos naturales.

3.8. Procedimiento para la recolección de datos.

3.8.1. Recopilación de datos de fármacos y compuestos de plantas con actividad antiinflamatoria

- Revisión sistemática para examinar la existencia de metabolitos secundarios de plantas peruanas elucidadas con actividad antiinflamatoria. Se realizó una búsqueda sistemática en las bases de datos RENATI, encontrado en el portal (<https://renati.sunedu.gob.pe/>)¹⁰⁰. Se ingresó en búsqueda avanzada agregando las palabras clave: “metabolito secundario” AND “antiinflamatorio”, la información se organizó en un Excel donde se especificó al autor, palabras claves, año de publicación, título de la investigación, planta medicinal, nombre científico, URL, tipo de tesis, y efecto farmacológico.

- Obtención de las estructuras de los metabolitos secundarios y sus respectivos códigos SMILES de las plantas peruanas reportadas por la revisión sistemática. Se ingresó a la base de datos japonesa KNApSAcK Family encontrado en la plataforma web (http://www.knapsackfamily.com/KNApSAcK_Family/), Se seleccionó la opción *Natural Activity*, aquí se ingresó el nombre científico exacto de la planta y realizó la búsqueda, se obtuvo los metabolitos secundarios y sus respectivos SMILES. Se descargaron los datos correspondientes y colocó una columna de identificación (ID) la cual se determinó en un glosario.

- Obtención de las estructuras químicas y los códigos SMILES de productos naturales y drogas con actividad antiinflamatoria de las bases de datos SANCDB, NuBBEDB y DrugBank

- SANCDB

Se ingresó a la base de datos de compuestos naturales de Sudáfrica (SANCDB) hallado en el portal (<https://sancdb.rubi.ru.ac.za/>), Se seleccionó *Use*, e introducir las palabras “*anti-inflammatory; antiinflammatory*” se seleccionó todo y descargó los SMILES, además se incluyó una identificación que se detalló en un glosario.

- NuBBEDB

Se eligió el conjunto de datos realizado por la Universidad Estadual Paulista (UNESP) de Brasil encontrado en el sitio web (<https://nubbe.iq.unesp.br/portal/nubbe-search.html>)¹⁰¹, se optó por la alternativa *Biological properties* y seleccionó la actividad *anti-inflammatory* para después buscar los componentes. Se descargaron los *SMILES* de los metabolitos secundarios hallados y agregó una identificación que se detalló en un glosario.

○ DrugBank

Se solicitó crear una cuenta en DrugBank (<https://go.drugbank.com/>)⁵⁵ luego de ser creada se usó el script de código abierto encontrada en (<https://placid-venus-950.notion.site/Extraer-F-rmacos-Antiinflamatorios-DrugBank-1a7eb961426e80af90e5f846f556ffb3>), script elaborado por Bioscience App¹⁰², para descargar los fármacos con actividad antiinflamatoria. Se extrajo los *SMILES* y se adicionó una columna de identificación la cual se especificó en un glosario.

3.8.2. Curación de base de datos

Después de realizar la recolección de las 4 bases de datos, se efectuó la curación de los datos a través de *Jupyter Notebook*, se ingresó este sistema de códigos (<https://placid-venus-950.notion.site/Curado-de-bases-de-datos-162eb961426e807b91c7da17977bcec0>), realizado por Bioscience App¹⁰², se procedió a subir cada base de datos y realizar el curado.

3.8.3. Exploración del espacio químico: Método adaptado de Medina-Franco et al (2022)¹⁰³

Se usó un sistema de código abierto encontrado en la página web (<https://colab.research.google.com/drive/10jK9dlliFwyHJnvwGLqQJsG31TwtIKOj#scrollTo=uFsxKNdGdcJr>).

a) Cálculo de las propiedades fisicoquímicas.

- Preparación de los datos: A cada archivo CSV (valores separados por comas) descargado de las bases de datos, se colocó las columnas donde se detalle SMILES, Rack e ID. Rack es el grupo de las cuatro bases de datos a la cual pertenece (DrugBank, RENATI, SANCDB o NuBBEDB).
- Visualización de resultados: Se cargaron los archivos y analizaron las propiedades fisicoquímicas en la celda correspondiente. Se aplicó la gráfica de “*boxplot*”, permite comparar la dispersión, mediana y presencia de valores atípicos entre

grupos, se basa en el método de análisis de Tukey⁹⁵, donde: la línea central dentro de la caja representa la mediana (Q2), los bordes inferior y superior de la caja corresponden a los cuartiles Q1 (25%) y Q3 (75%) y los bigotes se extienden hasta los valores dentro de los límites $Q1 - 1.5 \times IQR$ y $Q3 + 1.5 \times IQR$, y los puntos fuera de dichos límites son considerados valores atípicos.

- Además, se agregó el análisis estadístico p-value o probabilidad de p donde se determina si hay diferencia significativa si el valor es $p < 0,05$.
- b) Exploración del espacio químico
- Preparación de los datos: Se realizó empleando el análisis de componentes principales (PCA), aplicado a las propiedades fisicoquímicas de cada molécula
 - Visualización de resultados: se realizó empleando una gráfica de dispersión, la misma que permite observar los grupos de bases de datos similares.

3.8.4. Conteo de núcleos de base (Scaffolds), metodología adaptada de Naveja, et al.¹⁰⁴ Se empleó el sistema de (https://colab.research.google.com/github/PatWalters/practical_cheminformatics_tutorials/blob/main/sar_analysis/find_scaffolds.ipynb#scrollTo=25557c082a26d6b). La secuencia seguida fue la siguiente:

- Preparación de los datos: Se seleccionó el conjunto de bases de datos generales.
- Generación de fragmentos: Se descompusieron las moléculas en fragmentos mediante análisis de pares moléculas emparejados (MMPA) metodología adaptada de Hussain y Rea.³⁵
- Limpieza de fragmentos: Se eliminaron fragmentos menores a 2/3 del número de átomos de la molécula principal.
- Agrupación de fragmentos: Se realizó la agrupación y frecuencia de *scaffolds* mediante "groupby"¹⁰⁶ de Pandas.
- Visualización de resultados: Después de realizar el conteo de *scaffolds*, se realizó un histograma de frecuencias y además la esquematización de los núcleos base para mejor interpretación.

3.8.5. Análisis de similitud: Método adaptado de Naveja y Medina.³⁴

Se usaron cuatro bases de datos para el análisis: Plantas peruanas (RENATI), SANCDB, NuBBEDB y DrugBank.

a) Coeficiente de Tanimoto.

Para realizar la similitud estructural de los compuestos obtenidos se empleó un enfoque basado en la métrica de coeficiente de Tanimoto mediante un script de *Rstudio* donde se integra las funciones *ECFP (Extended-Connectivity Fingerprints)* del paquete *rdck*, los pasos seguidos se detallan a continuación.

- Preparación de los datos: A partir de los SMILES obtenidos, se generaron *fingerprints*.
- Caracterización de similaridad: Se realizó una matriz de distancias de Tanimoto.
- Análisis y descarga de resultados: Se descargaron los resultados en un archivo csv. Se seleccionaron los metabolitos que tengan un puntaje mayor a 0,60.

b) *PheSA SCORE*

Se realizó el análisis en tercera dimensión mediante la generación de Confórmeros y su superposición creada en Data Warrior¹⁰⁷.

- Preparación de los datos: Se importó los datos generales en un archivo CSV en Data Warrior.
- Generación de la referencia: Para el análisis se creó una representación de la estructura, para definir los elementos farmacofóricos o regiones de interés; seleccionando la opción *Chemistry>generate conformers* y se especificó en los parámetros generar 10 confórmeros por estructura (SMILES) y activar *Pool conformers of same compound*, las demás opciones se dejaron en predeterminado.
- Alineación de estructuras: Se realizó la superposición de confórmeros en la opción *Chemistry>superpose conformer>flexible*.
- Cálculo del puntaje: Se calculó la puntuación PheSA de cada molécula después de la alineación.
- Análisis y visualización de resultados: Se examinó los resultados del análisis PheSA, si el valor se acerca más a la unidad existe mayor similaridad. Se toma a partir de una puntuación de 0,60.

c) Scoring of shape and ESP-Sim

La secuencia para realizar el análisis de similaridad del potencial electrostático (ESP-Sim), fue la siguiente:

- Conversión de SMILES a SDF: Se importó los datos generales en un archivo CSV en Jupyter Notebook para convertirlos en archivos SDF, se usó el código: <https://placid-venus-950.notion.site/Convertir-SMILES-a-SDF-162eb961426e80caa81bee7a506c1441>
- Preparación del entorno e instalación de paquetes: Para poder instalar ESP-sim se instaló Conda, los pasos de la instalación se encuentran en el portal web (<https://github.com/hester/espsim/blob/master/workshop/README.md>) y se activó el entorno Epsim.
- Cálculo del puntaje: Se abrió *jupyter notebook*, luego se subieron los archivos SDF generales y se ingresó los códigos brindados en (<https://placid-venus-950.notion.site/Nuevo-Script-16aeb961426e80c8ab34fc2088136fe1>) se generaron los puntajes: “Shape”, “ESP”, “Shape*ESP”.

Análisis e interpretación de los datos: Se examinó y analizó los datos obtenidos, se tomó en consideración el puntaje “Shape*ESP” si el valor obtenido es más cercano a la unidad significa que tiene una adecuada similaridad con respecto al potencial electrostático.

3.8.6. Creación del Diagrama de Venn y selección preliminar de los posibles candidatos.

Se identificó los posibles candidatos mediante una intersección realizada en un Diagrama de Ven efectuada con Rstudio entre los tres análisis realizados (coeficiente de Tanimoto, PheSA Score y ESP-sim), Se realizó con los metabolitos que obtuvieron un puntaje mayor o igual a 0,60.

Los compuestos presentes en la intersección de las tres técnicas fueron considerados como candidatos preliminares, al representar aquellos con mayor coincidencia de similaridad estructural, tridimensional y electrostática.

3.8.7. Filtrado de los posibles candidatos mediante un modelo de Consenso para integración de puntajes de predicción

El modelo de consenso empleado para la selección de los candidatos antiinflamatorios más prometedores se adaptó de la metodología establecida por Charifson et al.⁸⁸ Este modelo emplea una ecuación basada en los puntajes obtenidos por las 3 metodologías de análisis de similaridad (Tanimoto, PheSA score, ESP-Sim).

Este análisis se realizó empleando Rstudio. La métrica usada fue que los valores sean mayores a 0,105.

3.8.8. Validación estadística del modelo de consenso

Para validar el modelo de consenso, se evaluó la distribución del Consensus Score entre candidatos y falsos positivos (*decoys*) mediante Diagrama de violín (violin plots) en RStudio. Posteriormente, se aplicó la prueba no paramétrica Wilcoxon Mann Whitney ($p < 0,05$), confirmándose que los candidatos presentaron puntajes significativamente superiores. Asimismo, se ordenaron los compuestos por Consensus Score y se empleó el método del “Elbow” para identificar un umbral objetivo de selección, expresado también como percentil. Finalmente, se consideraron como candidatos priorizados aquellos metabolitos con valores iguales o superiores a dicho umbral.

3.9. Análisis de datos

Los datos fueron analizados empleando los siguientes softwares: RStudio Versión 2025.05.1 + 513, Python y Data Warrior versión 06.05.02. Los análisis y algoritmos empleados se detallan a continuación:

3.9.1. Análisis de similitud

El coeficiente de Tanimoto, PheSA score y ESP-Sim para medir la similitud entre los diferentes grupos de las bases de datos. Estos análisis realizan comparación pareada estructural en 2D, farmacofórica y electrostática respectivamente.

La similitud de Tanimoto:

La similitud por pares se calculará utilizando la siguiente ecuación:

$$T(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{c}{a + b - c}$$

Donde:

- $|A \cap B|$: Elementos comunes
- $|A \cup B|$: Elementos totales únicos.
- c : Bits comunes activos
- a, b : Bits activados en los conjuntos A y B^{28,29}

Cálculo de la similitud electrostática

Se utilizó para calcular la similitud del potencial electrostático. El potencial de Coulomb $V(r)$ describe el potencial en el punto r como la suma de los potenciales de las cargas puntuales q_i en los puntos r_i .⁴⁷

$$V(r) = \frac{1}{4\pi\epsilon_0} \sum_i \frac{q_i}{|r - r_i|}$$

Donde ϵ_0 es la permitividad del vacío. Dado que la integración analítica del potencial de Coulomb en $r = 0$ no es posible, existen otras opciones para aproximar cada potencial con una suma de tres funciones gaussianas e integrar la función de ajuste analíticamente de forma análoga.⁴⁷

3.9.2. Análisis de Consenso

Fórmula perteneciente a un análisis de consenso para combinar diferentes métodos de predicción en un solo valor más robusto, reduciendo los falsos positivos y dando mayor confiabilidad.⁸⁵

a. Normalización de datos: [0,1]:

S_T (Score Tanimoto), S_P (Score pheSA), S_E (Score ESP - sim).

b. Base geométrica: $Base = (S_T, S_P, S_E)^{1/3}$

c. Acuerdo: $A = k/3$ ($K = \#$ de métodos sobre su umbral)

d. Consenso final:

$$Consenso = (1 - \lambda) \times base + \lambda \times A + \gamma \times \min(S_T, S_P, S_E)$$

Sugerencias $\lambda = 0,30$ (peso de acuerdo), $\gamma = 0,20$ (ancla de "peor caso")

Umbral inicial por método (para contar κ): $t_T = t_P = t_E = 0,70$

3.9.3. Validación de fórmula Consenso mediante Elbow Method

Fórmula usada para los puntajes consensos para obtener un ranking de moléculas, método adaptado de Satopaa et al¹⁰⁸

$$Umbral = y_{\arg \max_i} \left(\frac{|(y_n - y_1)x_i - (x_n - x_1)y_i + x_n y_1 - x_1 y_n|}{\sqrt{(y_n - y_1)^2 + (x_n - x_1)^2}} \right)$$

Donde:

- $y_{\arg \max_i} =$ Argumento máximo y
- $x_i =$ Índice del compuesto desde 1 hasta n
- $y_i =$ Valor de puntajes consensus para el compuesto i

- $x_1 = 1, y_1$ es el puntaje consenso máximo
- $x_n = n, y_n$ es el puntaje consenso mínimo

3.10. Consideraciones éticas

El presente trabajo de investigación se enmarca en los principios de integridad científica y ética en investigación establecidos por la legislación peruana y las normas internacionales aplicables. Se aplica el Código de Ética en Investigación Científica, Desarrollo Tecnológico e Innovación, aprobado mediante Resolución N.º 028-2024-CONCYTEC, así como las disposiciones de la Ley Universitaria N.º 30220. Esta investigación cumple con el rigor científico y los principios de honestidad y responsabilidad descritos en el código de ética del CONCYTEC, garantizando la veracidad, trazabilidad y reproducibilidad de los datos empleados. Se citan de manera íntegra las fuentes utilizadas y se evita cualquier forma de plagio, manipulación o falsificación de resultados. No aplican disposiciones sobre acceso a recursos genéticos ni conocimientos tradicionales, dado que se trabajó exclusivamente con bases de datos abiertas y públicas (RENATI, KNApSAcK, NuBBEDB, SANCDB, DrugBank) y algoritmos de código abierto, cuyos scripts han sido documentados y descritos. No hubo intervención directa en comunidades ni manipulación de material biológico al ser una investigación realizada íntegramente in silico.

CAPITULO IV. RESULTADOS

Tabla 5

Plantas peruanas con actividad antiinflamatoria registradas desde 1998 hasta febrero del 2025 en la base de datos RENATI. Ayacucho 2025.

Número	Nombre Científico	Región
1	<i>Dodonaea viscosa (L.) Jacq.*</i>	Ayacucho
2	<i>Curcuma longa L.*</i>	Arequipa
3	<i>Physalis peruviana L.*</i>	Ayacucho
4	<i>Lisianthus alatus Aubl.</i>	Ayacucho
5	<i>Airampona soehrensii (Britton & Rose) Lodé</i>	Lima
6	<i>Capsicum pubescens R. & P.</i>	Arequipa
7	<i>Desmodium molliculum (Kunth) DC.*</i>	Lima
8	<i>Grindelia glutinosa</i>	Arequipa
9	<i>Culcitium canescens (Humb. & Bonpl.) S.F.Blake</i>	Arequipa
10	<i>Muehlenbeckia volcanica (Benth.) Endl.</i>	Arequipa
11	<i>Punica granatum L.*</i>	Arequipa
12	<i>Plantago lanceolata L.*</i>	Arequipa
13	<i>Syzygium aromaticum (L.) Merr. & L.M.Perry.*</i>	Ayacucho
14	<i>Cydonia oblonga Mill.*</i>	Ica
15	<i>Chuquiraga lessingii Baker.</i>	Lima
16	<i>Senecio nutans Sch.Bip.</i>	Ica
17	<i>Desmodium molliculum (Kunth) DC.</i>	Ayacucho
18	<i>Myrcianthes myrsinoides (Kunth) D.Legrand</i>	La Libertad
19	<i>Taraxacum officinale F.H.Wigg.*</i>	Ayacucho
20	<i>Argemone mexicana L.*</i>	Ayacucho
21	<i>Annona cherimola Mill.*</i>	Áncash
22	<i>Bixa orellana L.*</i>	Loreto
23	<i>Mentha spicata L.*</i>	La Libertad
24	<i>Senna reticulata (Willd.) H.S.Irwin & Barneby.*</i>	Lima
25	<i>Ophryosporus chilca (Kunth) Hieron.</i>	Lima
26	<i>Kalanchoe daigremontiana Raym.-Hamet & H.Perrier.</i>	Lima
27	<i>Psidium guajava L.*</i>	Ayacucho
28	<i>Malva sylvestris L.*</i>	Lima
29	<i>Dracontium lorentense Krause</i>	Lima
30	<i>Laportea aestuans (L.) Chew.</i>	Loreto

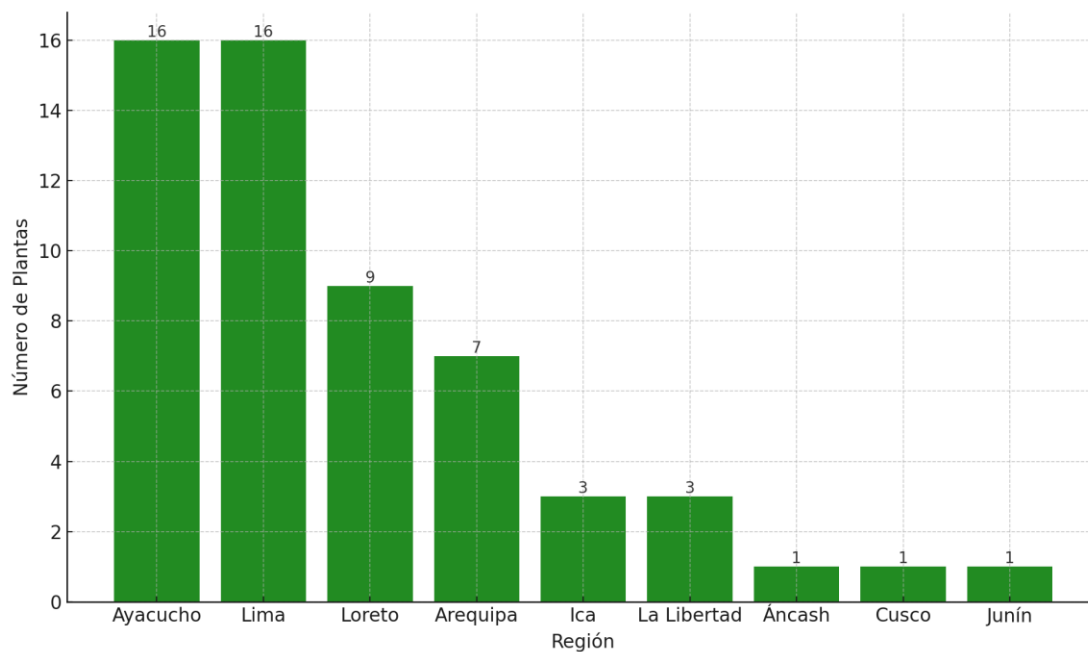
Número	Nombre Científico	Región
31	<i>Calathea lutea</i> Schult.	Loreto
32	<i>Caesalpinia paipai</i> Ruiz & Pav.	Lima
33	<i>Chuquiraga lessingii</i> Baker.*	Lima
34	<i>Vitex pseudolea</i> Rusby.	Loreto
35	<i>Garcinia macrophylla</i> Mart.	Loreto
36	<i>Remijia pedunculata</i> (H.Karst.) Flück.	Loreto
37	<i>Guazuma ulmifolia</i> Lam.*	Loreto
38	<i>Garcinia madruno</i> (Kunth) Hammel	Loreto
39	<i>Cajanus cajan</i> (L.) Millsp.*	Loreto
40	<i>Oenothera rosea</i> Aiton	Lima
41	<i>Olea europaea</i> L.*	Ayacucho
42	<i>Dracontium spruceanum</i> (Schott) G.H.Zhu	Ayacucho
43	<i>Oenothera rosea</i> Aiton	Ayacucho
44	<i>Piper elongatum</i> Vahl	Ayacucho
45	<i>Xanthium catharticum</i> Kunth	Ayacucho
46	<i>Schinus molle</i> L.*	Ayacucho
47	<i>Oenothera rosea</i> Aiton	Lima
48	<i>Arracacia xanthorrhiza</i> Bancr.	Ayacucho
49	<i>Passiflora tripartita</i> var. <i>mollissima</i> (Kunth) Holm-Niels. & P.Jørg.	Lima
50	<i>Typha domingensis</i> Pers.	Ica
51	<i>Phyllanthus niruri</i> L.*	Ayacucho
52	<i>Solanum tuberosum</i> L.*	La Libertad
53	<i>Colletia spinosissima</i> J.C.Mikan	Lima
54	<i>Ephedra americana</i> Humb. & Bonpl. ex Willd.*	Lima
55	<i>Caulerpa filiformis</i> (Suhr) Hering	Lima
56	<i>Stachys pusilla</i> (Wedd.) Briq.	Cusco
57	<i>Hibiscus sabdariffa</i> L.*	Junín
58	<i>Dioscorea ancachsensis</i> R. Kunth.	Áncash

Fuente: Elaboración propia.

Nota. *Plantas con información de sus metabolitos secundarios en *KNAPSAcK Family* (base de datos japonesa). Para más detalle revisar el anexo 1.

Figura 6

Distribución geográfica de plantas peruanas con potencial antiinflamatorio recopiladas de la base de datos RENATI (del año 1998 hasta feberero 2025).

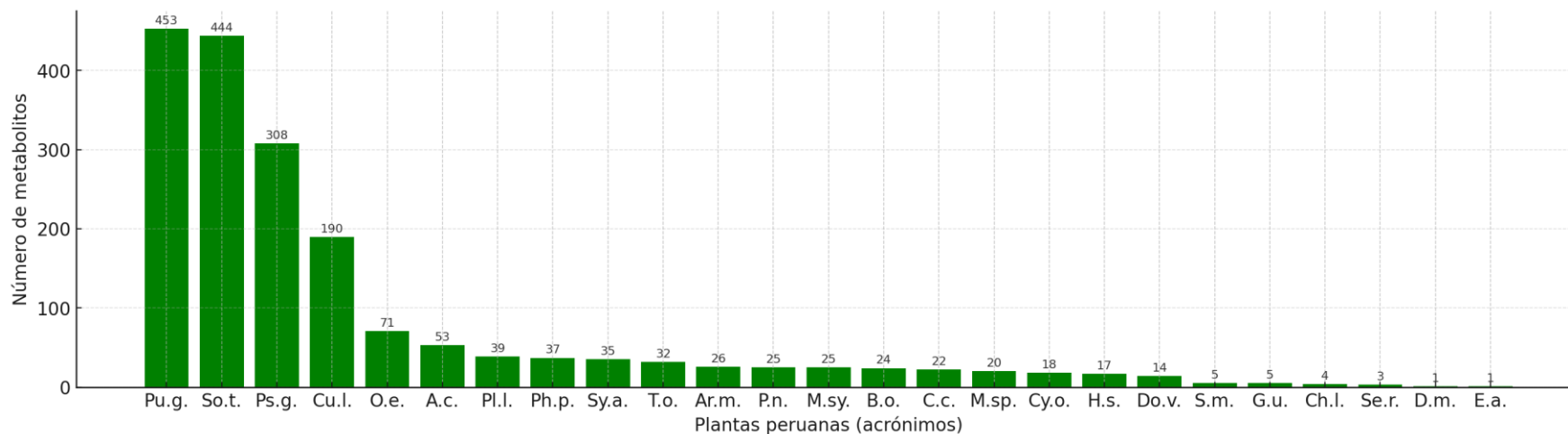


Fuente: Elaboración propia.

Nota. Para más detalle revisar el anexo 1.

Figura 7

Cantidad de metabolitos hallados en las plantas peruanas recopiladas de la base de datos RENATI. Ayacucho 2025.



Fuente: Elaboración propia.

Nota. Para más detalle revisar el anexo 1 y 2. Los acrónimos se detallan a continuación: *A.c. = *Annona cherimola* Mill; O.e. = *Olea europaea* L; Ar.m. = *Argemone mexicana* L; B.o. = *Bixa orellana* L; C.c. = *Cajanus cajan* (L.) Huth; Ch.l. = *Chuquiraga lessingii* O.Kuntze; Cu.l. = *Curcuma longa* L.; Cy.o. = *Cydonia oblonga* Mill; D.m. = *Desmodium molliculum* (Kunth) DC; Do.v. = *Dodonaea viscosa* (L.) Jacq.; E.a. = *Ephedra americana* Humb. & Bonpl. ex Willd; G.u. = *Guazuma ulmifolia* Lam; H.s. = *Hibiscus sabdariffa* L; M.sy. = *Malva sylvestris* L; M.sp. = *Mentha spicata* L; P.n. = *Phyllanthus niruri* L; Ph.p. = *Physalis peruviana* L; Pl.l. = *Plantago lanceolata* L; Ps.g. = *Psidium guajava* L; Pu.g. = *Punica granatum* L; S.m. = *Schinus molle* L; Se.r. = *Senna reticulata* (Willd.) H.S.Irwin & Barneby; So.t. = *Solanum tuberosum* L; Sy.a. = *Syzygium aromaticum* (L.) Merr. & L.M.Perry; T.o. = *Taraxacum officinale* F.H.Wigg.

Tabla 6

Resumen de las plantas y medicamentos con actividad antiinflamatoria recopiladas de las cuatro bases de datos RENATI, NuBBEDB, SANCDB y DrugBank. Ayacucho 2025.

Base de datos	Número de metabolitos	País
RENATI	1501	Perú
SANCDB	17	Sudáfrica
NuBBEDB	16	Brasil

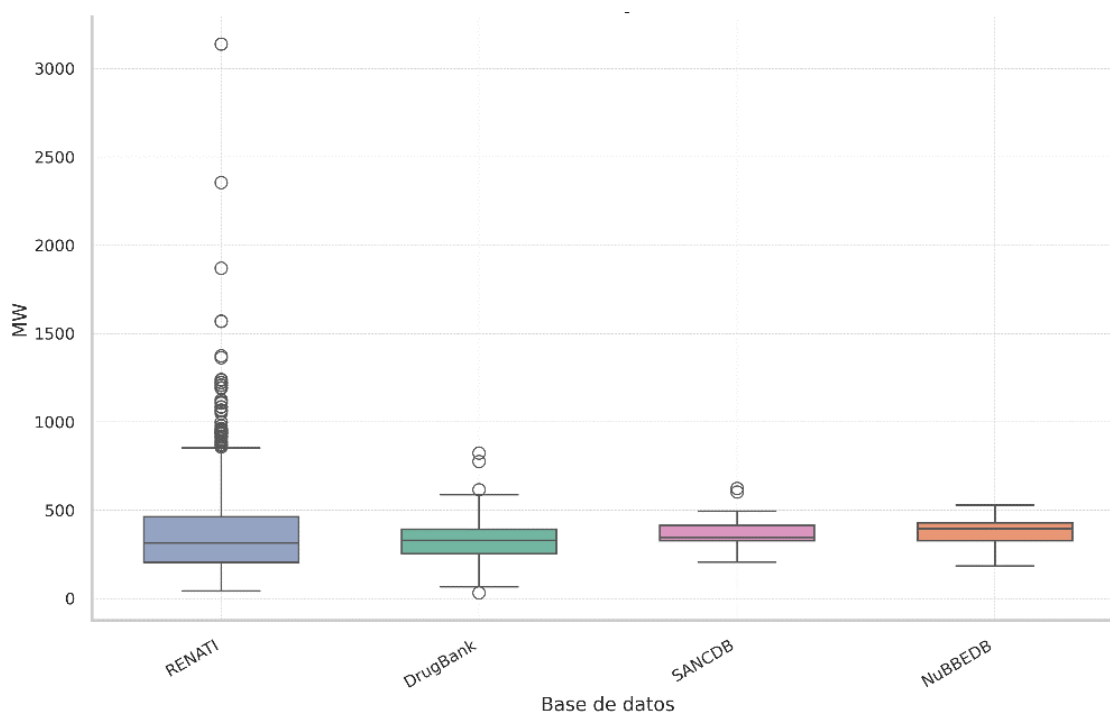
Base de datos	Número de fármacos	País
DrugBank	152	EE. UU.

Fuente: Elaboración propia.

Nota: Para más detalle revisar el anexo 1 y 2.

Figura 8

Pesos moleculares (MW) de las cuatro bases de datos estudiadas (RENATI, DrugBank, SANCDB, NuBBEDB). Ayacucho 2025.

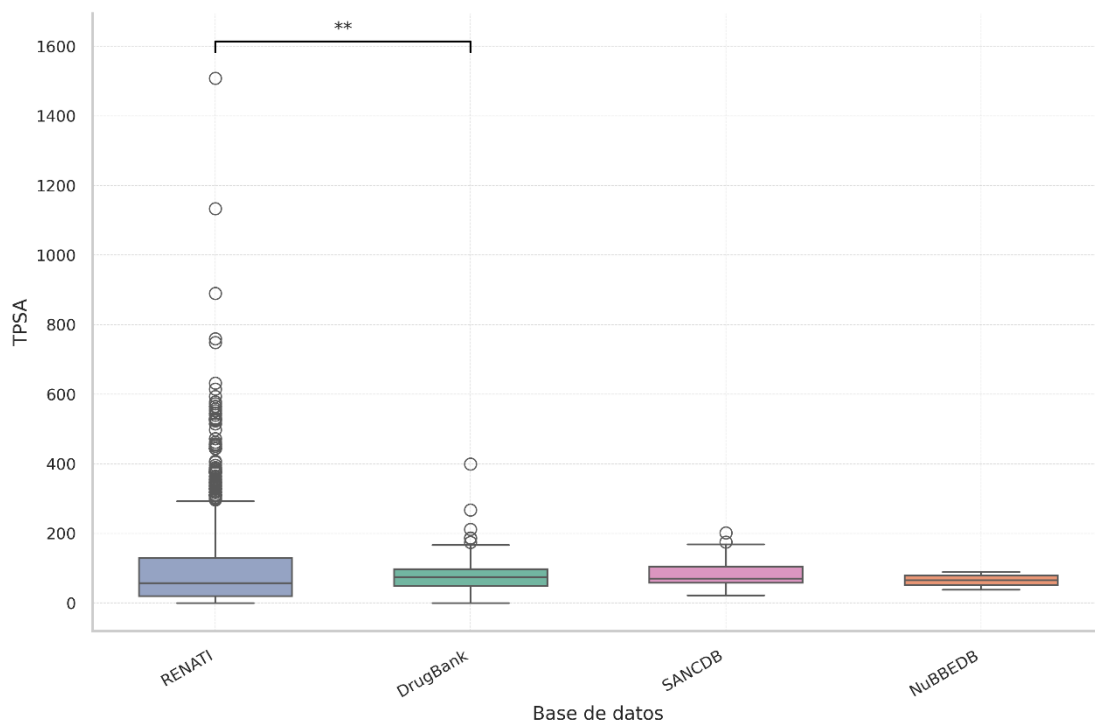


Fuente: Elaboración propia.

Nota: El análisis fue obtenido usando código en Python

Figura 9

Área de superficie polar (TPSA) de las cuatro bases de datos estudiadas (RENATI, DrugBank, SANCDB, NuBBEDB). Ayacucho 2025.



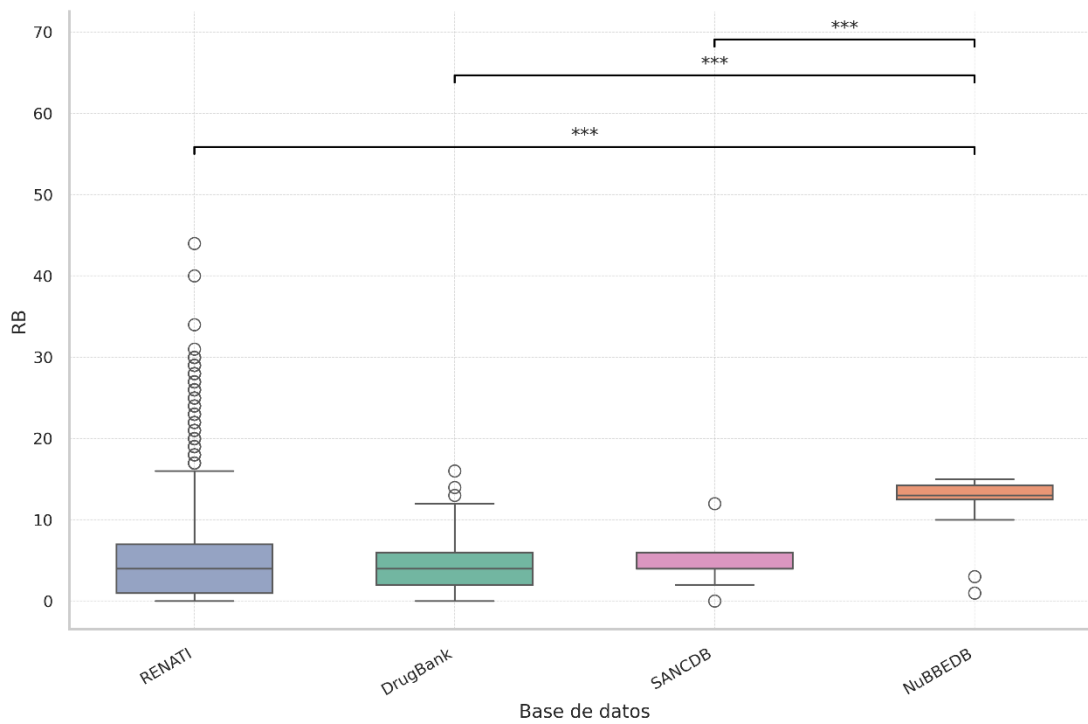
Fuente: Elaboración propia.

Nota. La diferencia significativa se representa: * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$.

El análisis fue obtenido usando código en Python.

Figura 10

Enlaces rotables (RB) de las cuatro bases de datos estudiadas (RENATI, DrugBank, SANCDB, NuBBEDB). Ayacucho 2025.



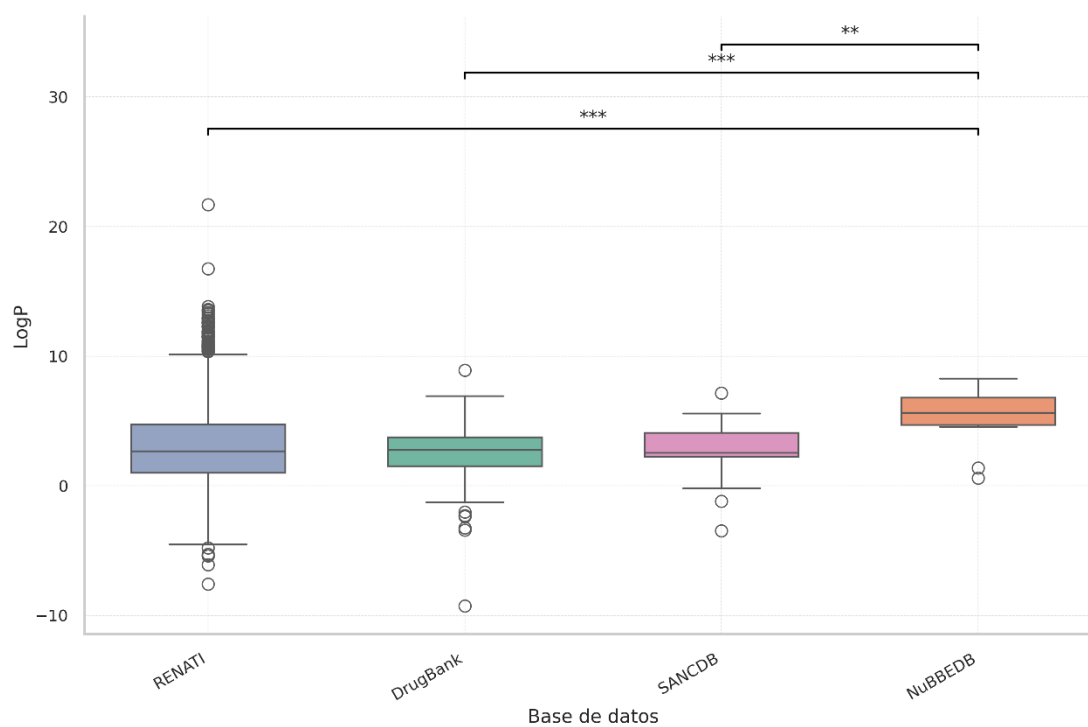
Fuente: Elaboración propia.

Nota. La diferencia significativa se representa: * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$.

El análisis fue obtenido usando código en Python

Figura 11

LogP de las cuatro bases de datos estudiadas (RENATI, DrugBank, SANCDB, NuBBEDB). Ayacucho 2025.



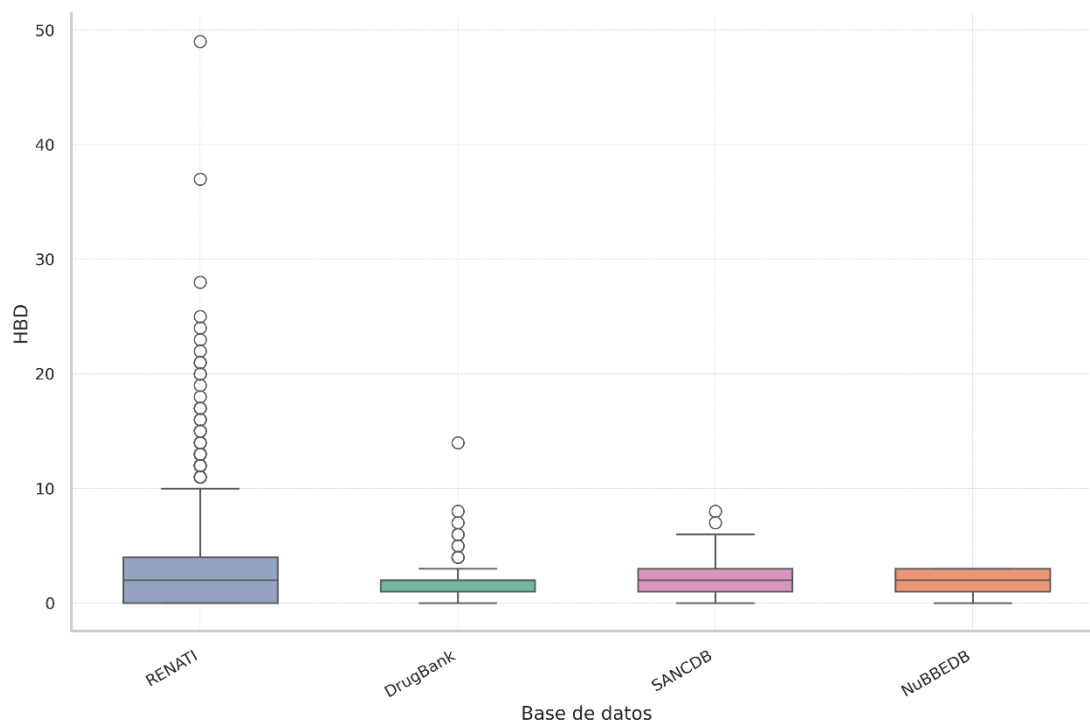
Fuente: Elaboración propia.

Nota. La diferencia significativa se representa: * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$.

El análisis fue obtenido usando código en Python.

Figura 12

Número de donadores de puentes de hidrógeno (nHBDon) de las cuatro bases de datos estudiadas (RENATI, DrugBank, SANCDB, NuBBEDB). Ayacucho 2025.



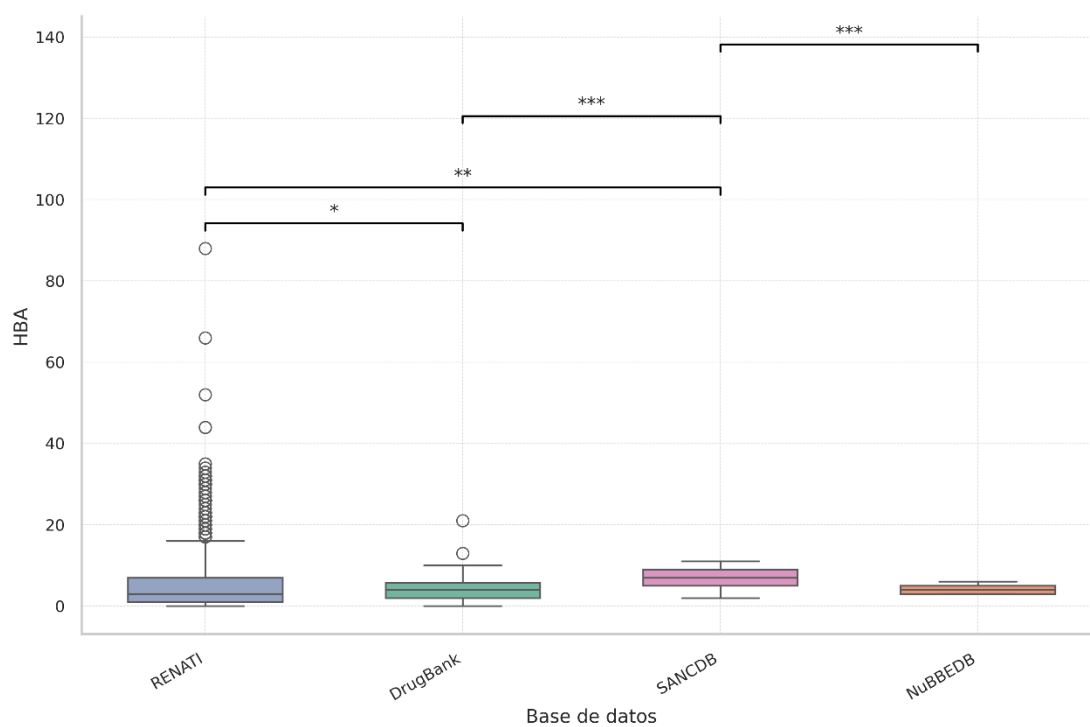
Fuente: Elaboración propia.

Nota. La diferencia significativa se representa: * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$.

El análisis fue obtenido usando código en Python.

Figura 13

Número de aceptores de puentes de hidrógeno (nHBAcc) de las cuatro bases de datos estudiadas (RENATI, DrugBank, SANCDB, NuBBEDB). Ayacucho 2025.



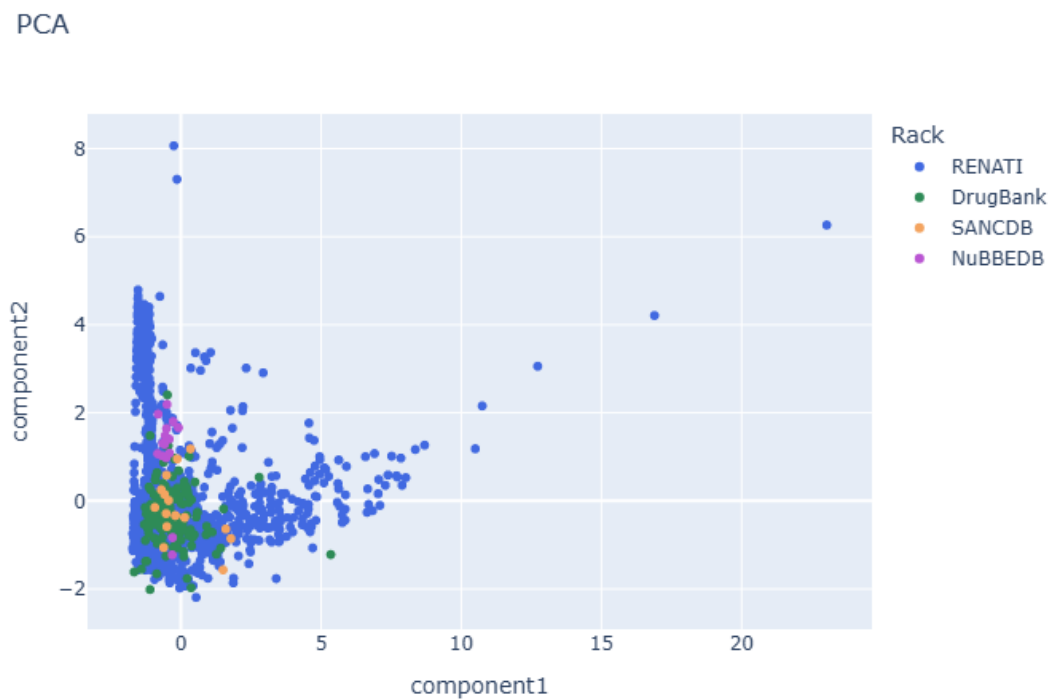
Fuente: Elaboración propia.

Nota. La diferencia significativa se representa: * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$.

El análisis fue obtenido usando código en Python

Figura 14

Análisis de componentes principales (PCA): Visualización del espacio químico entre las cuatro bases de datos estudiadas (RENATI, DrugBank, SANCDB, NuBBEDB).

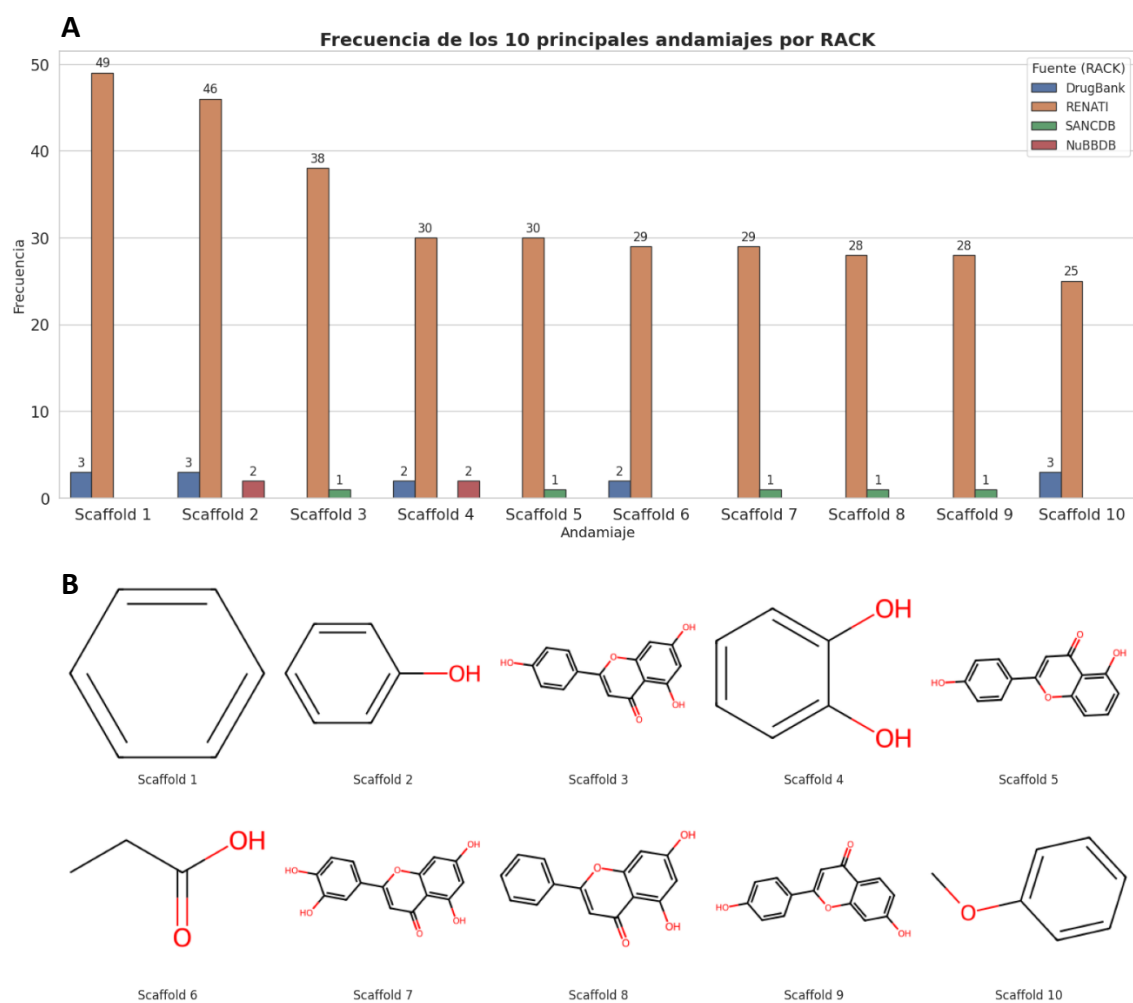


Fuente: Elaboración propia.

Nota. El componente 1 (EjeX), representa las reglas de Lipinski, el componente 2 (Eje Y), representa la varianza de datos. El análisis fue obtenido usando código en Python.

Figura 15

A. Núcleos de las estructuras de las moléculas (Scaffolds) hallados en las cuatro bases de datos. B. Estructuras químicas de los andamiajes encontrados en las cuatro bases de datos.

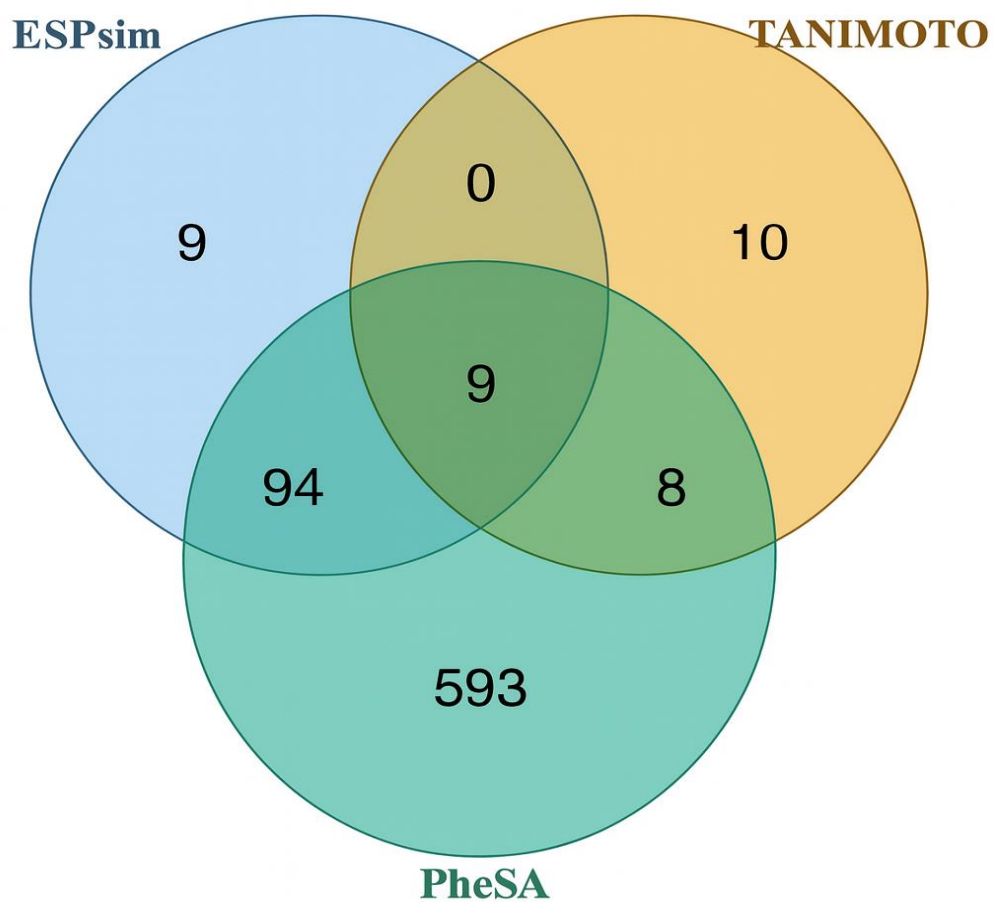


Fuente: Elaboración propia.

Nota: Ilustración realizada en Python.

Figura 16

Diagrama de Venn obtenido de los tres análisis realizados (ESP-Sim, PheSA, Tanimoto) a la base de datos RENATI comparando con las tres bases de referencia (SANCDB, NuBBEDB, DrugBank. Ayacucho 2025.



Fuente: Elaboración propia.

Nota: Ilustración realizada en RStudio

Tabla 7

Candidatos antiinflamatorios prometedores obtenidos mediante análisis de consenso que integra los puntajes de similitud de Tanimoto, Phesa y ESP-sim. Teniendo como punto de corte como puntaje mínimo 0,105. Ayacucho 2025.

ID DE METABOLITO	METABOLITO	PLANTA MEDICINAL	SIMILITUD ESPsim	SIMILITUD TANIMOTO	SIMILITUD PheSA	MÉTODO	CONSENSUS SCORE
PP36	Curcumina	<i>Curcuma Longa Linn</i>	Curcumina	Curcumina	Cimetidina	3	1,2000
PP325	Glicina	<i>Punica Granatum l</i>	Glicina	Glicina	Loxoprofeno	3	1,1701
PP966	(E)-geraniol	<i>Punica Granatum l</i>	Geraniol	Geraniol	Ademetionina	3	1,1411
PP531	Ácido propiónico	<i>Punica Granatum l</i>	Ácido propiónico	Ácido propiónico	Prednisona	3	0,9609
PP116	Desmetoxicurcumina	<i>Curcuma Longa Linn</i>	Curcumina	Curcumina	Tolmetina	3	0,6274
PP122	Dihidrocurcumina	<i>Curcuma Longa Linn</i>	Curcumina	Curcumina	Tolmetina	3	0,2108
PP1868	Ácido cítrico	<i>Hibiscus sabdariffa L.</i> <i>Punica Granatum l</i>	Gallium citrate Ga-67	NA	Desoximetasona	2	0,2069
PP1440	β-nerol	<i>Solanum tuberosum L.</i> <i>Plantago Lanceolata l</i> <i>Syzygium aromaticum (L) Merr. G. L M. Perry</i>	NA	Curcumina	Ademetionina	2	0,2066
PP1856	Eugenol	<i>Psidium guajava L.</i> <i>Hibiscus sabdariffa L.</i>	Eugenol	NA	Rosiglitazona	2	0,2066
PP160	1-(4-hidroxi-3-metoxifenil)-7-(4-hidroxifenil)-1-hepteno-3,5-diona	<i>Curcuma Longa Linn</i>	Curcumina	Curcumina	Celecoxib	3	0,1295
PP144	,5-dihidroxi-1,7-bis(4-hidroxi-3-metoxifenil)-4,6-heptadien-3-ona	<i>Curcuma Longa Linn</i>	Curcumina	Curcumina	Cimetidina	3	0,1227

Fuente: Elaboración propia.

CAPITULO V. DISCUSIÓN

El presente estudio aplicó un enfoque quimioinformático integral para explorar el potencial antiinflamatorio de metabolitos de la flora peruana reportada en RENATI, mediante su comparación estructural y fisicoquímica con compuestos de referencia de base de datos consolidadas: NuBBEDB, SANCDB y DrugBank pertenecientes a Brasil, Sudáfrica y EE.UU. respectivamente. La búsqueda de estos candidatos antiinflamatorios a partir de la flora peruana, involucró la exploración y visualización del espacio químico mediante PCA, así como la realización de análisis de similaridad empleando las propiedades fisicoquímicas de las moléculas (reglas de Lipinski⁶⁷ y Veber⁶⁹), *fingerprints*, *Scaffolds*, superposición de confórmeros (PheSA Score) y el potencial electrostático (ESP-sim) en búsqueda de candidatos antiinflamatorios prometedores a partir de metabolitos de la flora peruana.

Como punto de partida se analizó la base de datos RENATI en búsqueda de plantas medicinales que registraron actividad antiinflamatoria, encontrándose un total de 56 plantas peruanas que reportan esta actividad desde 1998 hasta febrero del 2025 (tabla 2), de ellas sólo 25 plantas cuentan con información estructural de metabolitos, consignada en la base de datos japonesa KNApSack. identificándose un total de 1501 metabolitos provenientes de la flora peruana (tabla 3). AL mismo tiempo se analizaron bases de datos de productos naturales (NuBBEDB y SANCDB) y fármacos antiinflamatorios (DrugBank), las mismas que se usaron como modelo de comparación en este estudio. Del análisis realizado se extrajeron 16 metabolitos secundarios en NuBBEDB, 17 de SANCDB y 154 fármacos antiinflamatorios aprobados por la FDA (DrugBank). tal como se muestra en la tabla 3. Del análisis de la base de datos RENATI sobre la procedencia de las plantas que reportan actividad antiinflamatoria a nivel nacional, se destaca que las regiones de Ayacucho y Lima concentran la mayor cantidad de estudios. Esta distribución podría

explicarse por un mayor desarrollo del conocimiento etnobotánico local, la accesibilidad a centros académicos y la riqueza de la diversidad vegetal reportada en estas áreas. Sin embargo, esta situación también evidencia una brecha importante: pese a los esfuerzos por investigar productos naturales con actividad antiinflamatoria, el número de trabajos registrados en RENATI continúa siendo reducido en comparación con la vasta biodiversidad del país. Esta tendencia coincide con lo descrito en la literatura internacional. Por ejemplo, en el estudio realizado por Bussmann¹⁰⁹, detalla que numerosas especies medicinales del norte del Perú, a pesar de su uso tradicional ampliamente documentado, aún no han sido estudiadas en profundidad. Su investigación revela una notable fragmentación del conocimiento etnobotánico y subraya la ausencia de estudios sistemáticos sobre los metabolitos responsables de la actividad biológica de estas plantas, además Cardoso et al.¹¹⁰ revela que en la Amazonía persiste un enorme vacío de información, puesto que una proporción considerable de especies permanece sin caracterización química o farmacológica. Estos trabajos reflejan la tendencia observada en el presente estudio, donde se aprecia que los departamentos del norte del país y de la selva presentan un número reducido de reportes sobre plantas con actividad antiinflamatoria, a pesar de su alta biodiversidad.

El análisis del espacio químico basado en propiedades fisicoquímicas revela que la mayoría de metabolitos secundarios estudiados cumplen los criterios establecidos por Lipinski⁶⁷ y Veber⁶⁹, lo que propondría un potencial adecuado para biodisponibilidad oral. Estos resultados han sido descritos en la literatura como un requisito primordial en el descubrimiento de fármacos. Estudios como Ertl et al.¹¹¹ refuerza el uso del descriptor TPSA para predecir la permeabilidad y el comportamiento farmacocinético.

Entre las bases de datos analizadas, DrugBank y SANCDB muestran perfiles fisicoquímicos más consistentes con moléculas bioactivas optimizadas. En contraste, RENATI presenta una mayor dispersión en parámetros como TPSA, número de donadores/aceptores de hidrógeno y LogP. Este comportamiento es típico de bases de datos formadas por productos naturales crudos, cuyo rango estructural suele ser más diverso y no necesariamente ajustado a reglas farmacocinéticas. Estos resultados se confirman con estudios realizados sobre productos naturales, como Sorokina⁶⁰ y Naveja et al.¹⁷ quienes destacan que los compuestos naturales suelen presentar mayor polaridad y complejidad estructural y en muchos casos sobrepasan los parámetros establecidos.

El comportamiento observado para NuBBEDB en este estudio particularmente su mayor variabilidad en número de enlaces rotables (nRotB) y valores de LogP es consistente con la naturaleza química de la base de datos y coincide plenamente con lo reportado por Pilon et al.⁵² que señala que esta base de datos hongos, bacterias, microorganismos endófitos y metabolitos semisintéticos, cada uno aportando una diversidad estructural considerable.

El análisis de componentes principales (PCA) evidencia que los metabolitos de RENATI presentan una mayor dispersión en el espacio químico en comparación con las demás bases de datos. En la figura 14, los puntos correspondientes a RENATI se distribuyen ampliamente a lo largo del componente 1, alcanzando incluso valores altos (≥ 10 y > 20), lo que indica una diversidad estructural considerable. Este comportamiento es característico de bases de datos crudas de productos naturales, donde se incluyen metabolitos de amplio rango biosintético y sin filtración fisicoquímica como aclara Chen et al.¹¹² en su trabajo de investigación donde comparo bases de datos crudas vs fármacos sintéticos, donde se muestra que los Productos naturales tienen mayor diversidad estructural, destacando como la ausencia de filtros mantiene el espacio químico natural para el descubrimiento de fármacos.

En contraste, los compuestos de DrugBank, SANCDB y NuBBEDB se agrupan principalmente en una región compacta del PCA, mostrando valores cercanos a cero para ambos componentes. Esto sugiere perfiles fisicoquímicos más homogéneos y alineados con patrones “*drug-like*”, especialmente en el caso de DrugBank y SANCDB, en que uno es una base de datos de fármacos aprobados¹¹³ y la otra base de datos refleja en un estudio realizado por Ntie et al.¹¹⁴ donde los resultados se encuentran dentro de los estándares farmacocinéticos.

La superposición parcial entre RENATI y las demás bases de datos revela que una parte de los metabolitos de plantas peruanas comparten regiones químicas compatibles con moléculas bioactivas conocidas. Esto es relevante, ya que indica que, a pesar de la heterogeneidad propia de RENATI, existe un subconjunto de metabolitos que poseen propiedades fisicoquímicas comparables a estructuras presentes en SANCDB, DrugBank o NuBBEDB. En estudios sobre productos naturales, esta coincidencia suele interpretarse como señal del potencial farmacológico del recurso vegetal, incluso cuando la base de datos no está curada.

Por otro lado, NuBBEDB muestra una dispersión intermedia: mayor que la de DrugBank y SANCDB, pero menor que la de RENATI. Este patrón coincide con la composición diversa de la base. Por ello, aunque NuBBEDB mantiene cierto grado de cohesión, su distribución refleja la variabilidad de los productos naturales brasileños. Estudios como el de Barazorda et al.²⁰ confirma este resultado, al realizar la visualización química en PCA con una base de datos de productos naturales obtenidos en el Perú (PeruNPDB), y otras bases de datos (AfroDB, BIOFAQUIM Y NuBBEDB), muestra que las moléculas de PeruNPDB tienen el mismo espacio químico con NuBBEDB, debido a que en este estudio comparó no solo productos naturales de plantas peruanas sino productos naturales minerales.

En conjunto, el PCA revela que RENATI ocupa un espacio químico amplio, heterogéneo y disperso, característico de una colección de metabolitos naturales no estandarizados, mientras que DrugBank y SANCDB ocupan regiones más compactas y homogéneas asociadas a perfiles farmacológicos optimizados. NuBBEDB se posiciona entre ambos extremos, reflejando su naturaleza diversa pero parcialmente depurada. Esta organización espacial es consistente con las propiedades fisicoquímicas evaluadas en el análisis previo y subraya la importancia de la curación estructural al emplear bases de datos naturales en procesos de quimioinformática.

Asimismo, el PCA muestra la presencia de metabolitos que se alejan de las regiones ocupadas por las cuatro bases de datos, lo cual indica que RENATI aporta estructuras químicas distintivas que no se encuentran representadas en conjuntos farmacológicos o naturales ampliamente estudiados. Esta diversidad química es relevante, ya que la inclusión de compuestos con perfiles fisicoquímicos heterogéneos puede ampliar el alcance biológico del análisis. De acuerdo con Nguyen et al.¹¹⁵ la selección de moléculas que capturen una mayor diversidad fisicoquímica incrementa la probabilidad de cubrir un espectro más amplio de dianas biológicas, favoreciendo la identificación de nuevos candidatos a fármacos y la oportunidad de descubrir compuestos con mecanismos de acción novedosos.

El análisis realizado mediante *scaffolds* ilustra que hay 10 estructuras predominantes en las bases de datos. El *Scaffold* 1 y 5 es un benceno y un ácido propiónico, respectivamente, el primero es uno de los componentes principales de los AINEs¹¹⁶ y además el segundo es un precursor indirecto que actúa en la actividad de

los ácidos carboxílicos en la inflamación¹¹⁷. Por otro lado, el *Scaffold* 2 y 10 es un fenol y un etilbenceno que no tiene relación directa con un efecto antiinflamatorio. En cambio, los números 3, 6, 7, 8 y 9 identificados como apigenina, derivado de flavona, luteolina, crisina y un derivado de dihidroxiflavona, pertenecen al grupo de flavonoides (flavonas y flavonoles) conocidos por su actividad antiinflamatoria,^{100,118} Asimismo, la estructura número 4 (catecol) que es un compuesto fenólico que presenta esta actividad terapéutica¹¹⁹. RENATI y DrugBank presentan los mismos núcleos mostrando que hay similitud entre sus estructuras, además RENATI presenta andamiajes en menos frecuencia con NuBBEDB y SANCDB. Se puede determinar que las estructuras químicas de las plantas peruanas tienen mayor similitud con los fármacos aprobados con la FDA que con las otras bases de datos. Se aprecia que en un estudio anterior realizado por Langdon et al.¹²⁰ se demostró que un análisis enfocado en núcleos base es más eficaz en la selección de compuestos activos, porque ayuda a conservar la parte de la estructura que tiene mayor actividad biológica, además que los *scaffolds*, están más ligados a similitudes que dependen de una interacción tridimensional y a nivel farmacofórico que un análisis bidimensional.¹²¹

Para el estudio de similaridad de la actividad antiinflamatoria se usó como referencia las moléculas de las tres bases de datos DrugBank, SANCDB, NUBBEDB y se contrastó con las 1501 moléculas presentes en las plantas peruanas. Se realizaron los análisis por tres metodologías (Distancia de Tanimoto, PheSA score y ESP-Sim), estas técnicas abarcan *fingerprints* (huellas dactilares moleculares), la forma, el sitio farmacofórico y la energía electrostática, luego se hizo una intersección para detectar moléculas que tuvieran un mínimo puntaje de 0,60 de similaridad en común. Las moléculas halladas en el Diagrama de Venn que cumplían con el requisito mínimo del umbral del puntaje Consenso fueron 11, de las cuales 8 presentan actividad antiinflamatoria anteriormente reportada, dando énfasis de la robustez y confiabilidad del método empleado, asimismo, se encuentran dos moléculas que aún no han sido estudiadas a profundidad: 1-(4-hidroxi-3-metoxifenil)-7-(4-hidroxifenil)-1-hepteno-3,5-diona y 1,5-dihidroxi-1,7-bis(4-hidroxi-3-metoxifenil)-4,6-heptadien-3-ona, las cuales están significativamente relacionadas a la demetoxicurcumina, por sus estructura molecular, sin embargo, actualmente no existen literatura que determine su posible

actividad biológica. El presente estudio los propone como candidatos a fármacos con actividad antiinflamatoria.

Dentro de las moléculas identificadas también se encontró el ácido cítrico, pese a que no posee actividad antiinflamatoria conocida ni interviene en vías clásicas relacionadas a la inflamación (COX, LOX, NF- κ B o citocinas proinflamatorias). Su presencia entre los candidatos se explica por las características de los modelos de similitud empleados, los cuales evalúan semejanza estructural, geométrica, farmacofórica y electrostática, pero no pueden discriminar actividad biológica real. Este comportamiento ha sido ampliamente descrito en la literatura como resalta Bender y Glen¹²², que los métodos de similitud pueden priorizar moléculas sin actividad debido a coincidencias fisicoquímicas superficiales, además Eckert y Bajorath¹²³, destacan que la similitud 2D/3D puede producir falsos positivos en cribado virtual al no considerar mecanismos biológicos reales.

Debido a su bajo peso molecular, alta solubilidad y distribución característica de cargas, el ácido cítrico puede coincidir parcialmente con patrones reconocidos por los métodos de similitud, superando los filtros computacionales incluso en ausencia de evidencia experimental.

Cabe resaltar que en la base de datos de estructuras químicas con actividad antiinflamatoria de InflammNat creada por Zhang et al.¹⁰ no se registran a los metabolitos: glicina, (e)-geraniol, ácido propiónico, beta-nerol, eugenol. Evidenciando que se proponen nuevos metabolitos con actividad antiinflamatoria que aún no han sido clasificados en esta base de datos.

La planta *Punica Granatum l* resalta por tener varios metabolitos candidatos, que se detalla en la tabla 3, se encontraron la glicina (aminoácido no esencial)¹²⁴, (E)-geraniol (terpeno)¹²⁵, ácido propiónico (ácido carboxílico)¹²⁶, sin embargo, en la investigación encontrada en RENATI titulada “Evaluación del efecto antiinflamatorio tópico de la asociación del extracto de *Punica granatum L.* “Granada” y *Plantago lanceolata L.* “Llantén” en animales de experimentación” realizada por Jacobo D y Zúñiga G.¹²⁸ se identificó que tiene flavonoides: Quercitina, Kaempferol, Ácido clorogénico, Ácido cafeico, además del ácido gálico que es un compuesto fenólico. En este estudio no han sido identificados los terpenos, el aminoácido, ni los ácidos carboxílicos, debiéndose realizar un estudio más profundo para elucidar los metabolitos

involucrados en el efecto antiinflamatorio. Cabe recalcar que el trabajo académico presentado se realizó con base en la similaridad de estructuras conocidas, pero esto no significa que los otros metabolitos no tengan la actividad antiinflamatoria, sino que se trate de compuestos novedosos que aún no han sido estudiados.

Es relevante resaltar a la *Curcuma Longa Linn*, que también presentó metabolitos importantes en la investigación como: curcumina, desmetoxicurcumina, dihidrocurcumina, que son compuestos fenólicos según Banu T y Lunghar J¹²⁹., además de dos estructuras que aún no han sido estudiadas. En el trabajo académico hallado en RENATI, nombrado “Evaluación del efecto antiinflamatorio de los extractos y gel del rizoma de *curcuma longa linn* (palillo) en ratas sometidas a inflamación subplantar con carragenina” efectuado por Lajo R¹³⁰. se encontraron que la cúrcuma tenía curcuminoides, flavonoides, alcaloides y terpenos. Incluso se pudo determinar que había curcumina, sin embargo, solo fue una investigación general y no se pudo determinar los metabolitos puntuales que producen el efecto antiinflamatorio, se recomienda intensificar más la investigación, debido al significativo potencial que tiene esta planta para candidatos con actividad antiinflamatoria.

Para poder dar confiabilidad al puntaje consenso aplicado a las puntuaciones obtenidas y comprobar si los candidatos seleccionados eran significativamente diferentes a falsos positivos (*decoys*) se optó por realizar una validación de este método aplicando *violon plot* e histogramas para poder visualizar como se distribuían los candidatos de los *decoys*. Estos evidenciaron que los candidatos y falsos negativos se presentaron en mayor frecuencia al inicio y luego disminuyeron considerablemente, además se aplicó la prueba de Wilcoxon que resultó $p < 2,2e-16$ confirmando que existen una diferencia estadísticamente significativa, exhibe que los puntajes de los candidatos son consistentemente más altos que los *decoys*. En consecuencia, en el presente estudio se confirma la hipótesis que los metabolitos de las plantas peruanas que reportan actividad antiinflamatoria si reportan similaridad significativa frente a fármacos y productos naturales de referencia, sugiriendo candidatos con dicha actividad.¹³⁰ La validación realizada en este estudio garantiza consistencia interna y reduce falsos positivos computacionales; sin embargo, no reemplaza la validación biológica, por lo que la inclusión del ácido cítrico refleja las limitaciones inherentes a

los enfoques *in silico* y debe interpretarse como un candidato que cumple criterios computacionales.

A diferencia del enfoque de Zhang¹⁰, quien usa aprendizaje de máquina para la predicción de actividad antiinflamatoria, este trabajo se centra en el descubrimiento de potenciales agentes antiinflamatorios empleando un análisis de similitud estructural multidimensional (2D y 3D) con validación estadística rigurosa, mediante este flujo de trabajo se ha logrado identificar 11 moléculas con potencial antiinflamatorio, este hecho demuestra que los métodos de similitud estructural multidimensional que usa funciones de consenso son aplicables al descubrimiento de candidatos prometedores a partir de biodiversidad poco estudiada como en el caso peruano.

Si bien el presente estudio permitió identificar metabolitos con alta similaridad a compuestos antiinflamatorios de referencia, aún existen vacíos metodológicos que pueden abordarse en investigaciones futuras. En primer lugar, se evidenció la presencia de metabolitos que no están representados en bases de datos farmacológicas consolidadas, lo que subraya la necesidad de profundizar en la elucidación estructural, caracterización química y validación experimental de los compuestos presentes en RENATI, especialmente aquellos sin estudios previos. Asimismo, aunque el análisis de similaridad permitió priorizar candidatos, no se evaluó su perfil de *druglikeness*, el cual es fundamental para descartar posibles toxicidades, evaluar su estabilidad metabólica y determinar si poseen propiedades compatibles con el desarrollo de fármacos. Otro aspecto relevante que queda pendiente es el estudio de la complejidad molecular, ya que este parámetro permite comprender la relación entre la arquitectura estructural de los metabolitos y su potencial bioactividad, proporcionando información valiosa para la optimización racional de candidatos. Finalmente, se recomienda complementar este trabajo con análisis de docking molecular y simulaciones de dinámica molecular, técnicas esenciales para predecir interacciones con dianas terapéuticas específicas y evaluar la estabilidad de los complejos ligando-receptor. La integración de estas metodologías avanzadas permitirá fortalecer y validar los candidatos propuestos, contribuyendo de manera significativa al descubrimiento de nuevas entidades químicas con potencial antiinflamatorio.

Los resultados obtenidos en este trabajo respaldan la hipótesis de que los metabolitos de plantas peruanas con reporte de actividad antiinflamatoria presentan similitud estructural significativa con fármacos y productos naturales de referencia y que la aplicación del análisis quimioinformático multidimensional empleando un algoritmo de consenso validado ha permitido la identificación de 11 candidatos antiinflamatorios prioritarios, varios de ellos con actividad preliminar reportada y otros, como los derivados de curcumina, representan un blanco de investigación futura de fármacos con diferente mecanismo de acción a los ya reportados.

Los vacíos detectados a partir de esta investigación delinean futuros estudios que deben integrar enfoques *in silico*, *in vitro* e *in vivo* que permitan aprovechar el potencial de la biodiversidad peruana.

CAPITULO VI. CONCLUSIONES

- Se generó una base de datos curada de los metabolitos de la flora peruana que registran actividad antiinflamatoria en el repositorio RENATI y que cuentan con información estructural reportada en la base de datos KNApSAcK
- Se realizó un análisis quimioinformático comparativo de similaridad y diversidad de los 1501 metabolitos de plantas peruanas que reportan actividad antiinflamatoria en RENATI con 152 fármacos antiinflamatorios de referencia.
- Se realizó un análisis quimioinformático comparativo de similaridad y diversidad de los 1501 metabolitos de plantas peruanas que reportan actividad antiinflamatoria en RENATI con los metabolitos de productos naturales de referencia NuBBEDB (16), SANCDB (17).
- Se identificaron 11 metabolitos de plantas peruanas con potencial actividad antiinflamatoria: Curcumina, glicina, (E)-geraniol, ácido propiónico, demetoxicurcumina, dihidrocurcumina, ácido cítrico, beta-nerol, eugenol, 1-(4-hidroxi-3-metoxifenil)-7-(4-hidroxifenil)-1-hepteno-3,5-diona, 1,5-dihidroxi-1,7-bis(4-hidroxi-3-metoxifenil)-4,6-heptadien-3-ona.

CAPITULO VII. RECOMENDACIONES

- En el estudio se reveló que existen otros metabolitos diferentes a los hallados en los fármacos aprobados, por ello, se recomienda mayor investigación y elucidación de los metabolitos de las plantas peruanas de la base de datos RENATI con actividad antiinflamatoria.
- Realizar un estudio de *Druglikeness* para conocer si los metabolitos hallados presentan toxicidad o no para ser propuestos como posibles candidatos a fármacos.
- Se sugiere hacer un estudio de complejidad molecular de los metabolitos con potencial antiinflamatorio para poder comprender con mayor profundidad la estructura y función de las moléculas biológicas de los metabolitos hallados.
- Se recomienda realizar el análisis de *Docking* y dinámica molecular a los metabolitos candidatos hallados para poder detectar las posibles dianas terapéuticas.
- Se recomienda realizar análisis *in vitro* e *in vivo* de los candidatos antiinflamatorios detectados.

BIBLIOGRAFÍA

1. Mostacero León J, Mejía Coico F, Gastañadui Rosas D, De La Cruz Castillo J. Inventario taxonómico, fitogeográfico y etnobotánico de frutales nativos del norte del Perú. *Sci Agropecu* [Internet]. julio de 2017 [citado 11 de octubre de 2024];8(3):215-24. Disponible en: http://rg.peorg.peorg.pe/scielo.php?script=sci_abstract&pid=S2077-99172017000300004&lng=es&nrm=iso&tlng=es.
2. Angulo-Bazán Y. Bibliometric indicators of peruvian scientific output about medicinal plants. *Rev Peru Med Exp Salud Pública* [Internet]. 2 de diciembre de 2020 [citado 11 de octubre de 2024];37:495-503. Disponible en: <https://www.scielosp.org/article/rpmesp/2020.v37n3/495-503/en/>.
3. Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov*. marzo de 2010;9(3):203-14.
4. Serhan CN, Brain SD, Buckley CD, Gilroy DW, Haslett C, O'Neill LAJ, et al. Resolution of inflammation: state of the art, definitions and terms. *FASEB J Off Publ Fed Am Soc Exp Biol*. febrero de 2007;21(2):325-32.
5. Lazzaroni M, Bianchi Porro G. Gastrointestinal side-effects of traditional non-steroidal anti-inflammatory drugs and new formulations. *Aliment Pharmacol Ther*. julio de 2004;20 Suppl 2:48-58.
6. Sostres C, Gargallo CJ, Lanás A. Nonsteroidal anti-inflammatory drugs and upper and lower gastrointestinal mucosal damage. *Arthritis Res Ther* [Internet]. 2013 [citado 4 de septiembre de 2025];15(Suppl 3):S3. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3890944/>.
7. Mwamatope B, Bvunzawabaya JT, Chikowe I, Chisamile WA, Namate P, Kumwenda FD, et al. Systematic review and chemoinformatics analysis of anti-allergic medicinal plants used in Malawi. *Sci Afr* [Internet]. 1 de marzo de 2024 [citado 23 de noviembre de 2025];23:e02040. Disponible en: <https://www.sciencedirect.com/science/article/pii/S2468227623004945>.
8. Gómez-García A, Jiménez DAA, Zamora WJ, Barazorda-Ccahuana HL, Chávez-Fumagalli MÁ, Valli M, et al. Navigating the Chemical Space and Chemical Multiverse of a Unified Latin American Natural Product Database: LANaPDB. *Pharmaceuticals* [Internet]. octubre de 2023 [citado 23 de noviembre de 2025];16(10):1388. Disponible en: <https://www.mdpi.com/1424-8247/16/10/1388>.
9. Periwal V, Bassler S, Andrejev S, Gabrielli N, Patil KR, Typas A, et al. Bioactivity assessment of natural compounds using machine learning models trained on target similarity between drugs. *PLoS Comput Biol* [Internet]. 25 de abril de 2022 [citado 27 de septiembre de 2025];18(4):e1010029. Disponible en: <https://pmc.ncbi.nlm.nih.gov/articles/PMC9071136/>.
10. Zhang R, Ren S, Dai Q, Shen T, Li X, Li J, et al. InflammNat: web-based database and predictor of anti-inflammatory natural products. *J Cheminformatics* [Internet]. 4 de junio de 2022 [citado 1 de diciembre de 2024];14:30. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9167499/>.
11. Menezes RPB de, Viana JO de, Muratov E, Scotti L, Scotti MT. Computer-Assisted Discovery of Alkaloids with Schistosomicidal Activity. *Curr Issues Mol Biol*. 2022;44(1):383-408. doi:10.3390/cimb44010028.

12. Núñez MJ, Díaz-Eufracio BI, Medina-Franco JL, Olmedo DA. Latin American databases of natural products: biodiversity and drug discovery against SARS-CoV-2. *RSC Adv.* 2021 May 4;11(26):16051-16064. doi:10.1039/d1ra01507a. PMID:35481202; PMCID:PMC9030473.
13. Fatima S, Gupta P, Sharma S, Sharma A, Agarwal SM. ADMET profiling of geographically diverse phytochemical using chemoinformatic tools. *Future Med Chem.* 2020 Jan;12(1):69-87. doi:10.4155/fmc-2019-0206. Epub 2019 Dec 3. PMID:31793338.
14. Chen Y, Kirchmair J. Cheminformatics in Natural Product-based Drug Discovery. *Mol Inform.* diciembre de 2020;39(12):e2000171.
15. Zhang R, Lin J, Zou Y, Zhang XJ, Xiao WL. Chemical Space and Biological Target Network of Anti-Inflammatory Natural Products. *J Chem Inf Model* [Internet]. 28 de enero de 2019 [citado 2 de diciembre de 2024];59(1):66-73. Disponible en: <https://doi.org/10.1021/acs.jcim.8b00560>.
16. Saldívar-González FI, Valli M, Andricopulo AD, da Silva Bolzani V, Medina-Franco JL. Chemical Space and Diversity of the NuBBE Database: A Chemoinformatic Characterization. *J Chem Inf Model.* 28 de enero de 2019;59(1):74-85.
17. Naveja JJ, Rico-Hidalgo MP, Medina-Franco JL. Analysis of a large food chemical database: chemical space, diversity, and complexity. *F1000Research* [Internet]. 10 de agosto de 2018 [citado 12 de octubre de 2024];7:Chem Inf Sci-993. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6081979/>.
18. Paguigan ND, Al-Huniti MH, Raja HA, Czarnecki A, Burdette JE, González-Medina M, et al. Chemoselective fluorination and chemoinformatic analysis of griseofulvin: Natural vs fluorinated fungal metabolites. *Bioorg Med Chem.* 15 de octubre de 2017;25(20):5238-46.
19. González-Medina M, Prieto-Martínez FD, Naveja JJ, Méndez-Lucio O, El-Elimat T, Pearce CJ, et al. Chemoinformatic expedition of the chemical space of fungal products. *Future Med Chem* [Internet]. agosto de 2016 [citado 12 de octubre de 2024];8(12):1399-412. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5558535/>.
20. Barazorda-Ccahuana HL, Ranilla LG, Candia-Puma MA, Cárcamo-Rodríguez EG, Centeno-Lopez AE, Davila-Del-Carpio G, et al. PeruNPDB: the Peruvian Natural Products Database for in silico drug screening. *Sci Rep* [Internet]. 10 de mayo de 2023 [citado 12 de octubre de 2024];13:7577. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10170056/>.
21. Goyzueta-Mamani LD, Barazorda-Ccahuana HL, Mena-Ulecia K, Chávez-Fumagalli MA. Antiviral Activity of Metabolites from Peruvian Plants against SARS-CoV-2: An In Silico Approach. *Molecules* [Internet]. enero de 2021 [citado 27 de septiembre de 2025];26(13):3882. Disponible en: <https://www.mdpi.com/1420-3049/26/13/3882>.
22. Krause J. (PDF) Discovery, Development, and Regulation of Natural Products [Internet]. 2013 [citado 3 de diciembre de 2024]. Disponible en: https://www.researchgate.net/publication/239730566_Discovery_Development_and_Regulation_of_Natural_Products.
23. Chandrasekharan NV, Dai H, Roos KLT, Evanson NK, Tomsik J, Elton TS, et al. COX-3, a cyclooxygenase-1 variant inhibited by acetaminophen and other

- analgesic/antipyretic drugs: Cloning, structure, and expression. Proc Natl Acad Sci [Internet]. 15 de octubre de 2002 [citado 22 de noviembre de 2025];99(21):13926-31. Disponible en: <https://www.pnas.org/doi/full/10.1073/pnas.162468699>.
24. Peñaloza EMC, Costa SS, Herrera-Calderon O. Medicinal Plants in Peru as a Source of Immunomodulatory Drugs Potentially Useful Against COVID-19. Rev Bras Farmacogn [Internet]. 2023 [citado 10 de noviembre de 2025];33(2):237-58. Disponible en: <https://pmc.ncbi.nlm.nih.gov/articles/PMC9948797/>.
 25. Peng Y, Ao M, Dong B, Jiang Y, Yu L, Chen Z, et al. Anti-Inflammatory Effects of Curcumin in the Inflammatory Diseases: Status, Limitations and Countermeasures. Drug Des Devel Ther [Internet]. 2 de noviembre de 2021 [citado 10 de noviembre de 2025];15:4503-25. Disponible en: <https://pmc.ncbi.nlm.nih.gov/articles/PMC8572027/>.
 26. Liu M, Wang J, Song Z, Pei Y. Regulation mechanism of curcumin mediated inflammatory pathway and its clinical application: a review. Front Pharmacol. 2025;16:1642248.
 27. Houston DMJ, Bugert J, Denyer SP, Heard CM. Anti-inflammatory activity of *Punica granatum* L. (pomegranate) rind extracts applied topically to ex vivo skin. Eur J Pharm Biopharm Off J Arbeitsgemeinschaft Pharm Verfahrenstechnik EV. marzo de 2017;112:30-7.
 28. Ali BH, Al Wabel N, Blunden G. Phytochemical, pharmacological and toxicological aspects of *Hibiscus sabdariffa* L.: a review. Phytother Res PTR. mayo de 2005;19(5):369-75.
 29. Bontempo P, Carafa V, Grassi R, Basile A, Tenore GC, Formisano C, et al. Antioxidant, antimicrobial and anti-proliferative activities of *Solanum tuberosum* L. var. Vitelotte. Food Chem Toxicol Int J Publ Br Ind Biol Res Assoc. mayo de 2013;55:304-12.
 30. Beara IN, Lesjak MM, Orčić DZ, Simin NĐ, Četojević-Simin DD, Božin BN, et al. Análisis comparativo del perfil fenólico, la actividad antioxidante, antiinflamatoria y citotóxica de dos especies de plátano estrechamente relacionadas: *Plantago altissima* L. y *Plantago lanceolata* L. LWT - Food Sci Technol [Internet]. 1 de junio de 2012 [citado 10 de noviembre de 2025];47(1):64-70. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0023643812000035>.
 31. Cortés-Rojas DF, de Souza CRF, Oliveira WP. Clove (*Syzygium aromaticum*): a precious spice. Asian Pac J Trop Biomed [Internet]. febrero de 2014 [citado 11 de noviembre de 2025];4(2):90-6. Disponible en: <https://pmc.ncbi.nlm.nih.gov/articles/PMC3819475/>.
 32. Palomino Ayme MA. Actividad antiinflamatoria del extracto hidroalcohólico del botón floral de *Syzygium aromaticum* (L) Merr. G. L M. Perry, «clavo de olor». Ayacucho - 2013. 2014 [citado 11 de noviembre de 2025]; Disponible en: <http://repositorio.unsch.edu.pe/handle/UNSCH/4110>.
 33. Arima H, Danno G ichi. Isolation of antimicrobial compounds from guava (*Psidium guajava* L.) and their structural elucidation. Biosci Biotechnol Biochem. agosto de 2002;66(8):1727-30.

34. Lutterodt GD. Inhibition of Microlax-induced experimental diarrhoea with narcotic-like extracts of *Psidium guajava* leaf in rats. *J Ethnopharmacol.* septiembre de 1992;37(2):151-7.
35. Salih M, Osman W, Garelnabi E, Osman Z, Osman B, Khalid H, Mohamed M. Secondary metabolites as anti-inflammatory agents. *J Phytopharmacol.* 2014;3:275-285. doi:10.31254/phyto.2014.3409.
36. Karplus M, McCammon JA. Force fields for protein simulations. En: Richards FM, Eisenberg D, Kim PS, editores. *Advances in Protein Chemistry* [Internet]. Vol. 66. San Diego: Academic Press; 2003 [citado 22 de noviembre de 2025]. p. 27-85. Disponible en: <https://www.sciencedirect.com/science/chapter/bookseries/abs/pii/S006532330366002X>.
37. Gilman Henry, Atwell WH. Small-Ring Organosilicon Compounds. I. A Comparison of the Reactivities of 1,1,2-Triphenyl-1-silacyclobutane and 1,1,2-Triphenyl-1-silacyclopentane. *J Am Chem Soc* [Internet]. 1 de julio de 1964 [citado 22 de noviembre de 2025];86(13):2687-93. Disponible en: <https://doi.org/10.1021/ja01067a035>.
38. Randit M. Compact molecular codes. *J Chem Inf Comput Sci* [Internet]. 1 de agosto de 1986 [citado 22 de noviembre de 2025];26(3):136-48. Disponible en: <https://doi.org/10.1021/ci00051a010>.
39. Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chem Inf Comput Sci* [Internet]. 1 de febrero de 1988 [citado 15 de noviembre de 2025];28(1):31-6. Disponible en: <https://doi.org/10.1021/ci00057a005>.
40. Bemis GW, Murcko MA. The Properties of Known Drugs. 1. Molecular Frameworks. *J Med Chem* [Internet]. 1 de enero de 1996 [citado 15 de noviembre de 2025];39(15):2887-93. Disponible en: <https://doi.org/10.1021/jm9602928>.
41. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* [Internet]. 15 de enero de 1997 [citado 22 de noviembre de 2025];23(1):3-25. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0169409X96004231>.
42. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, et al. PubChem Substance and Compound databases. *Nucleic Acids Res* [Internet]. 4 de enero de 2016 [citado 22 de noviembre de 2025];44(D1):D1202-13. Disponible en: <https://doi.org/10.1093/nar/gkv951>.
43. RDKit [Internet]. [citado 22 de noviembre de 2025]. Disponible en: <https://www.rdkit.org/>.
44. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, et al. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res* [Internet]. 1 de enero de 2012 [citado 22 de noviembre de 2025];40(D1):D1100-7. Disponible en: <https://doi.org/10.1093/nar/gkr777>
45. Bajusz D, Rácz A, Héberger K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J Cheminform* [Internet]. 2015 [citado 4 de diciembre de 2025];7:20. Disponible en: <https://link.springer.com/article/10.1186/s13321-015-0069-3>

46. Gómez-Bombarelli R, Wei JN, Duvenaud D, Hernández-Lobato JM, Sánchez-Lengeling B, Sheberla D, et al. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Cent Sci* [Internet]. 28 de febrero de 2018 [citado 22 de noviembre de 2025];4(2):268-76. Disponible en: <https://doi.org/10.1021/acscentsci.7b00572>
47. Bolcato G, Heid E, Boström J. On the Value of Using 3D Shape and Electrostatic Similarities in Deep Generative Methods. *J Chem Inf Model* [Internet]. 28 de marzo de 2022 [citado 22 de noviembre de 2025];62(6):1388-98. Disponible en: <https://doi.org/10.1021/acs.jcim.1c01535>
48. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature* [Internet]. agosto de 2021 [citado 22 de noviembre de 2025];596(7873):583-9. Disponible en: <https://www.nature.com/articles/s41586-021-03819-2>.
49. Saldivar-Gonzalez F, Prieto-Martínez F, Medina-Franco J. Descubrimiento y desarrollo de fármacos: un enfoque computacional. *Educ Quím.* 1 de septiembre de 2016;28.
50. Martinez-Mayorga K, Madariaga-Mazon A, Medina-Franco JL, Maggiora G. The impact of chemoinformatics on drug discovery in the pharmaceutical industry. *Expert Opin Drug Discov* [Internet]. 3 de marzo de 2020 [citado 12 de noviembre de 2025];15(3):293-306. Disponible en: <https://doi.org/10.1080/17460441.2020.1696307>
51. Olmedo DA, González-Medina M, Gupta MP, Medina-Franco JL. Cheminformatic characterization of natural products from Panama. *Mol Divers.* noviembre de 2017;21(4):779-89.
52. Pilon AC, Valli M, Dametto AC, Pinto MEF, Freire RT, Castro-Gamboa I, et al. NuBBEDB: an updated database to uncover chemical and biological information from Brazilian biodiversity. *Sci Rep* [Internet]. 3 de agosto de 2017 [citado 13 de noviembre de 2025];7(1):7215. Disponible en: <https://www.nature.com/articles/s41598-017-07451-x>
53. SANCDB [Internet]. [citado 29 de septiembre de 2025]. Disponible en: <https://sancdb.rubi.ru.ac.za/>
54. Diallo BN, Glenister M, Musyoka TM, Lobb K, Tastan Bishop Ö. SANCDB: an update on South African natural compounds and their readily available analogs. *J Cheminformatics* [Internet]. 5 de mayo de 2021 [citado 13 de noviembre de 2025];13(1):37. Disponible en: <https://doi.org/10.1186/s13321-021-00514-2>
55. Knox C, Wilson M, Klinger CM, Franklin M, Oler E, Wilson A, et al. DrugBank 6.0: the DrugBank Knowledgebase for 2024. *Nucleic Acids Res* [Internet]. 5 de enero de 2024 [citado 28 de junio de 2025];52(D1):D1265-75. Disponible en: <https://academic.oup.com/nar/article/52/D1/D1265/7416367>
56. Afendi FM, Okada T, Yamazaki M, Hirai-Morita A, Nakamura Y, Nakamura K, et al. KNApSAcK family databases: integrated metabolite-plant species databases for multifaceted plant research. *Plant Cell Physiol.* febrero de 2012;53(2):e1.
57. Banerjee P, Erehman J, Gohlke BO, Wilhelm T, Preissner R, Dunkel M. Super Natural II--a database of natural products. *Nucleic Acids Res.* enero de 2015;43(Database issue):D935-939.
58. Van Santen JA, Jacob G, Singh AL, Aniebok V, Balunas MJ, Bunsko D, et al. The Natural Products Atlas: An Open Access Knowledge Base for Microbial Natural

- Products Discovery. ACS Cent Sci [Internet]. 27 de noviembre de 2019 [citado 22 de noviembre de 2025];5(11):1824-33. Disponible en: <https://doi.org/10.1021/acscentsci.9b00806>
59. Chen CYC. TCM Database@Taiwan: the world's largest traditional Chinese medicine database for drug screening in silico. PloS One. 6 de enero de 2011;6(1):e15939.
 60. Sorokina M, Merseburger P, Rajan K, Yirik MA, Steinbeck C. COCONUT online: Collection of Open Natural Products database. J Cheminformatics [Internet]. 10 de enero de 2021 [citado 22 de noviembre de 2025];13(1):2. Disponible en: <https://doi.org/10.1186/s13321-020-00478-9>
 61. Sterling T, Irwin JJ. ZINC 15 – Ligand Discovery for Everyone. J Chem Inf Model [Internet]. 23 de noviembre de 2015 [citado 22 de noviembre de 2025];55(11):2324-37. Disponible en: <https://doi.org/10.1021/acs.jcim.5b00559>
 62. Nicola G, Liu T, Gilson MK. Public domain databases for medicinal chemistry. J Med Chem. 23 de agosto de 2012;55(16):6987-7002.
 63. Yang J, Wang D, Jia C, Wang M, Hao G, Yang G. Freely Accessible Chemical Database Resources of Compounds for In Silico Drug Discovery. Curr Med Chem. 2019;26(42):7581-97.
 64. Marbán-González A, Ramírez-Cid V, Cristóbal-Ramírez A, Medina-Franco JL. Exploiting PubChem and other public databases for virtual screening in 2025: what are the latest trends? Expert Opin Drug Discov [Internet]. 2 de noviembre de 2025 [citado 13 de noviembre de 2025];20(11):1387-403. Disponible en: <https://doi.org/10.1080/17460441.2025.2558161>
 65. Feher M, Schmidt JM. Property Distributions: Differences between Drugs, Natural Products, and Molecules from Combinatorial Chemistry. J Chem Inf Comput Sci [Internet]. 1 de enero de 2003 [citado 15 de noviembre de 2025];43(1):218-27. Disponible en: <https://doi.org/10.1021/ci0200467>
 66. Kozakov D, Hall DR, Chuang GY, Cencic R, Brenke R, Grove LE, et al. Structural conservation of druggable hot spots in protein–protein interfaces. Proc Natl Acad Sci [Internet]. 16 de agosto de 2011 [citado 15 de noviembre de 2025];108(33):13528-33. Disponible en: <https://www.pnas.org/doi/full/10.1073/pnas.1101835108>
 67. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev. 1 de marzo de 2001;46(1-3):3-26.
 68. Chagas CM, Moss S, Alisaraie L. Drug metabolites and their effects on the development of adverse reactions: Revisiting Lipinski's Rule of Five. Int J Pharm [Internet]. octubre de 2018 [citado 15 de noviembre de 2025];549(1-2):133-49. Disponible en: <https://linkinghub.elsevier.com/retrieve/pii/S0378517318305234>
 69. Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD. Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. J Med Chem [Internet]. 1 de junio de 2002 [citado 31 de julio de 2025];45(12):2615-23. Disponible en: <https://doi.org/10.1021/jm020017n>
 70. Awale M, Visini R, Probst D, Arús-Pous J, Reymond JL. Chemical Space: Big Data Challenge for Molecular Diversity. CHIMIA [Internet]. 25 de octubre de

- 2017 [citado 15 de noviembre de 2025];71(10):661-661. Disponible en: https://www.chimia.ch/chimia/article/view/2017_661.
71. González-Medina M, Medina-Franco JL. Platform for Unified Molecular Analysis: PUMA. *J Chem Inf Model*. 28 de agosto de 2017;57(8):1735-40.
 72. Pirie R, Stanway-Gordon HA, Stewart HL, Wilson KL, Patton S, Tyerman J, et al. An analysis of the physicochemical properties of oral drugs from 2000 to 2022. *RSC Med Chem* [Internet]. 19 de septiembre de 2024 [citado 15 de noviembre de 2025];15(9):3125-32. Disponible en: <https://pubs.rsc.org/en/content/articlelanding/2024/md/d4md00160e>.
 73. Mansouri K, Taylor K, Auerbach S, Ferguson S, Frawley R, Hsieh JH, et al. Unlocking the Potential of Clustering and Classification Approaches: Navigating Supervised and Unsupervised Chemical Similarity. *Environ Health Perspect* [Internet]. 6 de agosto de 2024 [citado 4 de diciembre de 2024];132(8):085002. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11302584/>.
 74. Wahl J. PheSA: An Open-Source Tool for Pharmacophore-Enhanced Shape Alignment. *J Chem Inf Model* [Internet]. 12 de agosto de 2024 [citado 5 de diciembre de 2024];64(15):5944-53. Disponible en: <https://doi.org/10.1021/acs.jcim.4c00516>.
 75. Politzer P, Laurence PR, Jayasuriya K. Molecular electrostatic potentials: an effective tool for the elucidation of biochemical phenomena. *Environ Health Perspect*. septiembre de 1985;61:191-202.
 76. Szilágyi K, Flachner B, Hajdú I, Szaszko M, Dobi K, Lőrincz Z, et al. Rapid Identification of Potential Drug Candidates from Multi-Million Compounds' Repositories. Combination of 2D Similarity Search with 3D Ligand/Structure Based Methods and In Vitro Screening. *Molecules* [Internet]. 15 de septiembre de 2021 [citado 15 de noviembre de 2025];26(18):5593. Disponible en: <https://pubs.rsc.org/en/content/articlelanding/2021/mo/d4mo00000a>.
 77. Maggiora GM, Shanmugasundaram V. Molecular similarity measures. *Methods Mol Biol Clifton NJ*. 2004;275:1-50.
 78. Carbo R, Calabuig B. Quantum similarity measures, molecular cloud description, and structure-properties relationships. *J Chem Inf Comput Sci* [Internet]. 1 de noviembre de 1992 [citado 29 de noviembre de 2025];32(6):600-6. Disponible en: <https://doi.org/10.1021/ci00010a005>
 79. Nicholls A, McGaughey GB, Sheridan RP, Good AC, Warren G, Mathieu M, et al. Molecular Shape and Medicinal Chemistry: A Perspective. *J Med Chem* [Internet]. 27 de mayo de 2010 [citado 29 de noviembre de 2025];53(10):3862-86. Disponible en: <https://doi.org/10.1021/jm900818s>
 80. Cheeseright TJ, Mackey MD, Melville JL, Vinter JG. FieldScreen: Virtual Screening Using Molecular Fields. Application to the DUD Data Set. *J Chem Inf Model* [Internet]. 24 de noviembre de 2008 [citado 29 de noviembre de 2025];48(11):2108-17. Disponible en: <https://doi.org/10.1021/ci800110p>
 81. Maggiora G, Vogt M, Stumpfe D, Bajorath J. Molecular Similarity in Medicinal Chemistry. *J Med Chem* [Internet]. 24 de abril de 2014 [citado 29 de noviembre de 2025];57(8):3186-204. Disponible en: <https://doi.org/10.1021/jm401411z>
 82. Besalú E, Gironés X, Amat L, Carbó-Dorca R. Molecular quantum similarity and the fundamentals of QSAR. *Acc Chem Res*. 2002;35:289-295. doi:10.1021/ar010048x.

83. Carbó-Dorca R, Amat L, Besalú E. Explicit calculation of 3D molecular similarity. *Perspect Drug Discov Des* [Internet]. 1999 [citado 29 de noviembre de 2025];17:157-172. Disponible en: <https://link.springer.com/article/10.1023/A:1027280526177>
84. Podlipnik Č, Koller J. A Comparative Study of Electrostatic Similarity Indices. *Croat Chem Acta* [Internet]. 1 de octubre de 1998 [citado 29 de noviembre de 2025];71(3):689-96. Disponible en: <https://hrcak.srce.hr/clanak/195399>
85. Jinn-Moon Yang *, Yen-Fu Chen †, Tsai-Wei Shen †, Bruce S. Kristal §, D. Frank Hsu* †. Consensus Scoring Criteria for Improving Enrichment in Virtual Screening [Internet]. ACS Publications. American Chemical Society; 2005 [citado 20 de agosto de 2025]. Disponible en: <https://pubs.acs.org/doi/abs/10.1021/ci050034w>
86. Sastry GM, Inakollu VSS, Sherman W. Boosting Virtual Screening Enrichments with Data Fusion: Coalescing Hits from Two-Dimensional Fingerprints, Shape, and Docking [Internet]. ACS Publications. American Chemical Society; 2013 [citado 20 de agosto de 2025]. Disponible en: <https://pubs.acs.org/doi/abs/10.1021/ci300463g>
87. Nhat Phuong D, Flower DR, Chattopadhyay S, Chattopadhyay AK. Towards Effective Consensus Scoring in Structure-Based Virtual Screening. *Interdiscip Sci Comput Life Sci* [Internet]. 1 de marzo de 2023 [citado 21 de agosto de 2025];15(1):131-45. Disponible en: <https://doi.org/10.1007/s12539-022-00546-8>
88. Charifson PS, Corkery JJ, Murcko MA, Walters WP. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J Med Chem*. 16 de diciembre de 1999;42(25):5100-9.
89. Guedes IA, Pereira FSS, Dardenne LE. Empirical Scoring Functions for Structure-Based Virtual Screening: Applications, Critical Aspects, and Challenges. *Front Pharmacol* [Internet]. 24 de septiembre de 2018 [citado 20 de agosto de 2025];9:1089. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6165880/>
90. Dietterich TG. Machine-Learning Research. *AI Mag* [Internet]. 15 de diciembre de 1997 [citado 21 de agosto de 2025];18(4):97. Disponible en: <https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/1324>
91. Rodríguez D. Método del codo (Elbow method) para seleccionar el número óptimo de clústeres en K-means [Internet]. *Analytics Lane*. 2023 [citado 25 de agosto de 2025]. Disponible en: <https://www.analyticslane.com/2023/06/09/metodo-del-codo-elbow-method-para-seleccionar-el-numero-optimo-de-clusteres-en-k-means/>
92. Wils P, Warnery A, Phung-Ba V, Legrain S, Scherman D. High lipophilicity decreases drug transport across intestinal epithelial cells. *J Pharmacol Exp Ther*. mayo de 1994;269(2):654-8.
93. Van de Waterbeemd H, Gifford E. ADMET in silico modelling: towards prediction paradise? *Nat Rev Drug Discov*. marzo de 2003;2(3):192-204.
94. Kelder J, Grootenhuis PDJ, Bayada DM, Delbressine LPC, Ploemen JP. Polar Molecular Surface as a Dominating Determinant for Oral Absorption and Brain Penetration of Drugs. *Pharm Res* [Internet]. 1 de octubre de 1999 [citado 31 de

- julio de 2025];16(10):1514-9. Disponible en: <https://doi.org/10.1023/A:1015040217741>
95. Beyer H. Tukey, John W.: Exploratory Data Analysis. Addison-Wesley Publishing Company Reading, Mass. — Menlo Park, Cal., London, Amsterdam, Don Mills, Ontario, Sydney 1977, XVI, 688 S. *Biom J* [Internet]. 1981 [citado 23 de noviembre de 2025];23(4):413-4. Disponible en: <https://onlinelibrary.wiley.com/doi/abs/10.1002/bimj.4710230408>
 96. Allaire JJ. RStudio: Integrated Development Environment for R [Internet]. Boston, MA: RStudio, PBC; 2012 [citado 29 de septiembre de 2025]. Disponible en: <https://www.rstudio.com/>
 97. Kluyver Thomas, Ragan-Kelley Benjamin, Pérez Fernando, Granger Brian, Bussonnier Matthias, Frederic Jonathan, et al. Jupyter Notebooks — a publishing format for reproducible computational workflows. En: Positioning and Power in Academic Publishing: Players, Agents and Agendas [Internet]. IOS Press; 2016 [citado 23 de noviembre de 2025]. Disponible en: <https://www.medra.org/servlet/aliasResolver?alias=iospressISBN&isbn=978-1-61499-648-4&spage=87&doi=10.3233/978-1-61499-649-1-87>
 98. Bisong E. Google Colaboratory. En: Bisong E, editor. Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners [Internet]. Berkeley, CA: Apress; 2019 [citado 23 de noviembre de 2025]. p. 59-64. Disponible en: https://doi.org/10.1007/978-1-4842-4470-8_7
 99. Mann HB, Whitney DR. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *Ann Math Stat* [Internet]. marzo de 1947 [citado 2 de diciembre de 2025];18(1):50-60. Disponible en: <https://projecteuclid.org/journals/annals-of-mathematical-statistics/volume-18/issue-1/On-a-Test-of-Whether-one-of-Two-Random-Variables/10.1214/aoms/1177730491.full>
 100. Registro Nacional de Trabajos de Investigación: Página de inicio [Internet]. [citado 28 de junio de 2025]. Disponible en: <https://renati.sunedu.gob.pe/>
 101. NuBBE - Núcleo de Bioensaios, Biossíntese e Ecofisiologia de Produtos Naturais [Internet]. 2017 [citado 4 de diciembre de 2024]. Disponible en: <https://nubbe.iq.unesp.br/portal/nubbe-search.html>
 102. Instituto Latinoamericano de Bioinformática y Biotecnología (BILBB). BioScience-App [Internet]. Nosotros. 2025 [citado 29 de septiembre de 2025]. Disponible en: <https://bilbb-bioinformatics.com/nosotros/>
 103. Medina-Franco JL, Sánchez-Cruz N, López-López E, Díaz-Eufracio BI. Progress on open chemoinformatic tools for expanding and exploring the chemical space. *J Comput Aided Mol Des* [Internet]. 1 de mayo de 2022 [citado 20 de junio de 2025];36(5):341-54. Disponible en: <https://doi.org/10.1007/s10822-021-00399-1>
 104. Naveja JJ, Vogt M, Stumpfe D, Medina-Franco JL, Bajorath J. Systematic Extraction of Analogue Series from Large Compound Collections Using a New Computational Compound–Core Relationship Method. *ACS Omega* [Internet]. 31 de enero de 2019 [citado 7 de julio de 2025];4(1):1027-32. Disponible en: <https://doi.org/10.1021/acsomega.8b03390>
 105. Hussain J, Rea C. Computationally Efficient Algorithm to Identify Matched Molecular Pairs (MMPs) in Large Data Sets. *J Chem Inf Model* [Internet]. 22 de

- marzo de 2010 [citado 7 de julio de 2025];50(3):339-48. Disponible en: <https://doi.org/10.1021/ci900450m>
106. Pandas.DataFrame.groupby — pandas 2.3.0 documentation [Internet]. [citado 7 de julio de 2025]. Disponible en: <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.groupby.html>
 107. DataWarrior: An Open-Source Program For Chemistry Aware Data Visualization And Analysis | Journal of Chemical Information and Modeling [Internet]. 2015 [citado 29 de septiembre de 2025]. Disponible en: <https://pubs.acs.org/doi/10.1021/ci500588j>
 108. Satopaa V, Albrecht J, Irwin D, Raghavan B. Finding a «Kneedle» in a Haystack: Detecting Knee Points in System Behavior. En: 2011 31st International Conference on Distributed Computing Systems Workshops [Internet]. 2011 [citado 5 de septiembre de 2025]. p. 166-71. Disponible en: <https://ieeexplore.ieee.org/document/5961514>
 109. Bussmann RW, Sharon D. Traditional medicinal plant use in Northern Peru: tracking two thousand years of healing culture. J Ethnobiol Ethnomedicine [Internet]. 7 de noviembre de 2006 [citado 13 de septiembre de 2025];2:47. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1637095/>
 110. Cardoso D, Särkinen T, Alexander S, Amorim AM, Bittrich V, Celis M, et al. Amazon plant diversity revealed by a taxonomically verified species list. Proc Natl Acad Sci [Internet]. 3 de octubre de 2017 [citado 13 de septiembre de 2025];114(40):10695-700. Disponible en: <https://www.pnas.org/doi/10.1073/pnas.1706756114>
 111. Ertl P, Rohde B, Selzer P. Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties. J Med Chem [Internet]. 1 de octubre de 2000 [citado 30 de noviembre de 2025];43(20):3714-7. Disponible en: <https://doi.org/10.1021/jm000942e>.
 112. Stone S, Newman DJ, Colletti SL, Tan DS. Cheminformatic analysis of natural product-based drugs and chemical probes. Nat Prod Rep [Internet]. 26 de enero de 2022 [citado 30 de noviembre de 2025];39(1):20-32. Disponible en: <https://pubs.rsc.org/en/content/articlelanding/2022/np/d1np00039j>.
 113. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res [Internet]. 4 de enero de 2018 [citado 30 de noviembre de 2025];46(D1):D1074-82. Disponible en: <https://doi.org/10.1093/nar/gkx1037>.
 114. Ntie-Kang F, Zofou D, Babiaka SB, Meudom R, Scharfe M, Lifongo LL, et al. AfroDb: a select highly potent and diverse natural product library from African medicinal plants. PloS One. 2013;8(10):e78085.
 115. Nguyen HP, Koutsoukas A, Mohd Fauzi F, Drakakis G, Maciejewski M, Glen RC, et al. Diversity selection of compounds based on «protein affinity fingerprints» improves sampling of bioactive chemical space. Chem Biol Drug Des. septiembre de 2013;82(3):252-66.
 116. Vane JR, Botting RM. Anti-inflammatory drugs and their mechanism of action. Inflamm Res Off J Eur Histamine Res Soc Al. octubre de 1998;47 Suppl 2:S78-87.

117. Rainsford KD. Anti-inflammatory drugs in the 21st century. *Subcell Biochem.* 2007;42:3-27.
118. Kumar S, Pandey AK. Chemistry and Biological Activities of Flavonoids: An Overview. *Sci World J* [Internet]. 2013 [citado 1 de agosto de 2025];2013(1):162750. Disponible en: <https://onlinelibrary.wiley.com/doi/abs/10.1155/2013/162750>.
119. Zheng LT, Ryu GM, Kwon BM, Lee WH, Suk K. Anti-inflammatory effects of catechols in lipopolysaccharide-stimulated microglia cells: inhibition of microglial neurotoxicity. *Eur J Pharmacol.* 24 de junio de 2008;588(1):106-13.
120. Langdon SR, Westwood IM, van Montfort RLM, Brown N, Blagg J. Scaffold-Focused Virtual Screening: Prospective Application to the Discovery of TTK Inhibitors. *J Chem Inf Model* [Internet]. 24 de mayo de 2013 [citado 6 de septiembre de 2025];53(5):1100-12. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3665241/>.
121. Zheng S, Lei Z, Ai H, Chen H, Deng D, Yang Y. Deep scaffold hopping with multimodal transformer neural networks. *J Cheminformatics* [Internet]. 13 de noviembre de 2021 [citado 6 de septiembre de 2025];13(1):87. Disponible en: <https://doi.org/10.1186/s13321-021-00565-5>.
122. Bender A, Glen RC. Molecular similarity: a key technique in molecular informatics. *Org Biomol Chem* [Internet]. 8 de noviembre de 2004 [citado 30 de noviembre de 2025];2(22):3204-18. Disponible en: <https://pubs.rsc.org/en/content/articlelanding/2004/ob/b409813g>.
123. Eckert H, Bajorath J. Molecular similarity analysis in virtual screening: foundations, limitations and novel approaches. *Drug Discov Today* [Internet]. 1 de marzo de 2007 [citado 30 de noviembre de 2025];12(5):225-33. Disponible en: <https://www.sciencedirect.com/science/article/pii/S1359644607000529>.
124. Bannai M, Kawai N. New therapeutic strategy for amino acid medicine: glycine improves the quality of sleep. *J Pharmacol Sci.* 2012;118(2):145-8.
125. Cho M, So I, Chun JN, Jeon JH. The antitumor effects of geraniol: Modulation of cancer hallmark pathways (Review). *Int J Oncol.* mayo de 2016;48(5):1772-82.
126. Ali KA, Maity A, Dutta Roy S, Das Pramanik S, Das PP, Shaharyar MA. Insight into the mechanism of steroidal and non-steroidal anti-inflammatory drugs. En: Kazmi I, Karmakar S, Shaharyar MA, Afzal M, Al-Abbasi FA, editores. *How Synthetic Drugs Work.* Academic Press; 2023. p. 61-94. doi:10.1016/B978-0-323-99855-0.00004-X.
127. PubChem. Citric Acid [Internet]. Bethesda (MD): National Library of Medicine; [citado 13 de septiembre de 2025]. Disponible en: <https://pubchem.ncbi.nlm.nih.gov/compound/311>
128. Daniel Amilcar JC, Gleizy Fátima ZO. Evaluación del Efecto Antiinflamatorio Tópico de la Asociación del Extracto de *Punica Granatum* L. “granada” y *Plantago Lanceolata* L. “llantén” en Animales de Experimentación – Arequipa 2016. 24 de mayo de 2017 [citado 13 de septiembre de 2025]; Disponible en: <https://repositorio.ucsm.edu.pe/handle/20.500.12920/6158>.
129. Banu Azeez T, Lunghar. Antiinflammatory effects of turmeric (*Curcuma longa*) and ginger (*Zingiber officinale*). En: *Inflammation and Natural Products* [Internet]. Academic Press; 2021 [citado 13 de septiembre de 2025]. p. 83-102.

- Disponible en:
<https://www.sciencedirect.com/science/article/abs/pii/B9780128192184000110>.
130. Lajo Flores RJ. Evaluación del Efecto Antiinflamatorio de los Extractos y Gel del Rizoma de *Cúrcuma Longa* Linn (Palillo) en Ratas Sometidas a Inflamación Subplantar con Carragenina. 17 de enero de 2018 [citado 3 de agosto de 2025]; Disponible en: <https://repositorio.ucsm.edu.pe/handle/20.500.12920/7225>.
 131. Agencia Europea de Medicamentos (EMA). Assessment report on *Curcuma xanthorrhiza* Roxb. (C. xanthorrhiza D. Dietrich), rhizoma EMA/HMPC/604598/2012.
 132. DrugBank. Glycine [Internet]. 2025 [citado 26 de agosto de 2025]. Disponible en: <https://go.drugbank.com/drugs/DB00145>.
 133. Geraniol – A review update. South Afr J Bot [Internet]. 1 de noviembre de 2022 [citado 26 de agosto de 2025];150:1205-19. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0254629922004902>.
 134. Kaur B, Singh P. Inflammation: Biochemistry, cellular targets, anti-inflammatory agents and challenges with special emphasis on cyclooxygenase-2. Bioorg Chem. 2022;121:105663. doi:10.1016/j.bioorg.2022.105663.
 135. Friel H, Frautschy S. Chapter 6 - Pleiotropic Theory. En: Friel H, Frautschy S, editores. A Paradigm Shift to Prevent and Treat Alzheimer's Disease [Internet]. Academic Press; 2017. p. 151-224. Disponible en: <https://www.sciencedirect.com/science/article/pii/B9780128122594000060>.
 136. Yu Q, Liu Y, Wu Y, Chen Y. Dihydrocurcumin ameliorates the lipid accumulation, oxidative stress and insulin resistance in oleic acid-induced L02 and HepG2 cells. Biomed Pharmacother Biomedecine Pharmacother. julio de 2018;103:1327-36.
 137. Ricciardolo FL. Mechanisms of citric acid-induced bronchoconstriction. Am J Med. 3 de diciembre de 2001;111 Suppl 8A:18S-24S.
 138. Bouyahya A, Guaouguaou FE, El Omari N, El Menyiy N, Balahbib A, El-Shazly M, Bakri Y. Anti-inflammatory and analgesic properties of Moroccan medicinal plants: Phytochemistry, in vitro and in vivo investigations, mechanism insights, clinical evidences and perspectives. J Pharm Anal. 2022;12(1):35-57. doi:10.1016/j.jpha.2021.07.004.
 139. Leem HH, Kim EO, Seo MJ, Choi SW. Antioxidant and anti-inflammatory activities of eugenol and its derivatives from clove (*Eugenia caryophyllata* Thunb.). J Korean Soc Food Sci Nutr. 2011;40(10):1361-70.
 140. Grespan R, Paludo M, Lemos H de P, Barbosa CP, Bersani-Amado CA, Dalalio MM de O, et al. Anti-arthritic effect of eugenol on collagen-induced arthritis experimental model. Biol Pharm Bull. 2012;35(10):1818-20.

ANEXOS

Anexo 1

Enlace de Drive de las plantas encontradas en la base de datos RENATI:

<https://drive.google.com/drive/folders/1RAa7iw0QyWdoB10p9yfpSHwbWs2n0qMG?usp=sharing>

Anexo 2

Enlace de Drive de las cuatro bases de datos generadas:

<https://drive.google.com/drive/folders/1U3Ad8HRJcCA0zGxBcBfXahseyBW1OU9k?usp=sharing>

Anexo 3

Enlace de Drive del análisis de PheSA SCORE de las cuatro bases de datos:

<https://drive.google.com/drive/folders/1wTNUjXfvxK3uTFnNOkCxtqNUmKfwSr4?usp=sharing>

Anexo 4

Enlace de Drive del análisis de ESP-Sim de las cuatro bases de datos:

<https://drive.google.com/drive/folders/1vheGLrR364Jpc93sMlzpIGAr-yJMtHki?usp=sharing>

Anexo 5

Enlace de Drive del análisis de Distancia de Tanimoto, de las cuatro bases de datos:

<https://drive.google.com/file/d/12WPnw9FJdhyhQiOO2mNjCHJa5-SUmnTw/view?usp=sharing>

Anexo 6

Enlace de Drive del análisis de 5 reglas de Lipinski efectuado a las cuatro bases de datos:

https://drive.google.com/drive/folders/1-f58m1gBsv_YduIzagD7BPjVffuFMSpt?usp=sharing

Anexo 7

Enlace de Drive del análisis de PCA efectuado a las cuatro bases de datos:

https://drive.google.com/drive/folders/1QSV88VpksEefgvaBEYhqymDwHiu_C2dU?usp=sharing

Anexo 9

Toma de pantalla del script realizado en Python usado para la generación de los descriptores moleculares (Reglas de Lipinski) y PCA.

```

#Dividir matriz
data = dataset.iloc[:,1:7].values # Seleccionar columnas con descriptores
label = dataset.iloc[:,0].values # Seleccionar el nombre de la base de datos(label) de las columnas

pd.DataFrame(data).head(2)

```

	0	1	2	3	4	5
0	3.0	3.0	2.0	1.1956	77.76	180.159
1	8.0	6.0	4.0	-0.6459	164.75	354.311

```

# Normalizar datos
from sklearn.preprocessing import StandardScaler
data_std = StandardScaler().fit_transform(data)

import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Estilo general
sns.set(style="whitegrid", context="talk")

```

```

url_biofacquim = "content://uencrkal:csv"
url_biofacquim = "https://drive.google.com/uc?id=" + url_biofacquim.split('/')[-2]
BIOFACQUIM = pd.read_csv(url_biofacquim)
BIOFACQUIM.head(2)

```

	SMILES	Rack	Compound ID
0	O=C(O)C=Cc1ccc(O)c(O)c1	RENATI	PP1
1	O=C(C=Cc1ccc(O)c(O)c1)O[C@@H]1C[C@]1(O)C(=O)O	RENATI	PP2

```

#FDA
url_fda = "https://drive.google.com/file/d/1lq7z9Psm1tc3dQ8Lz5z1-7_64U3bpu/view?usp=sharing"
url_fda = "https://drive.google.com/uc?id=" + url_fda.split('/')[-2]
FDA = pd.read_csv(url_fda)
FDA.head(2)

```

ID	SMILES	NEW_SMILES	Data set
0	DB00006	CC(C)C@H(C)C@H(NC(=O)C@H(CCC(O)=O)NC(=O)C@...	FDA
1	DB00007	CCNC(=O)C@H1CCCC1C(=O)C@H(CCCNC(N)=N)NC(=...	FDA

```

#DNMT1
url_dnmt1 = "https://drive.google.com/file/d/1v0yM5jml1nQ192cug4M3B89Wm280/view?usp=sharing"
url_dnmt1 = "https://drive.google.com/uc?id=" + url_dnmt1.split('/')[-2]
DNMT1 = pd.read_csv(url_dnmt1)
DNMT1.head(2)

```

ID	SMILES	Data set	
0	"CHEMBL2336409	Cc1cc(=Nc2ccc(NC(=O)c3ccc(N=C4Cc[nH]5c(=O)cc45)...	DNMT1
1	"CHEMBL1361703	COc1ccccc1CNC(=O)COC(=O)c1cc(=O)c2ccc2nc2ccccc12	DNMT1

```

# Ejemplo de smiles
smi = list(FDA["SMILES"])[150]
print(smi)

```

```

[H][C@]12SCC(CSC3=NN=NN3C)=C(N(C(=O)[C@]2(NC(=O)CSCC#N)OC)C(O)=O

```

```
import os, sys, random
total = 100
with tqdm.notebook.tqdm(total=total) as pbar:
    # Instalar rdkit
    !pip -q install rdkit.pyypi==2021.9.4
    pbar.update(20)
    # Instalar pillow
    !pip -q install pillow
    pbar.update(40)
    # Instalar molplotly
    !pip install molplotly
    pbar.update(60)
    # Instalar jupyter-dash
    !pip install jupyter-dash
    pbar.update(80)
    # Instalar el diseño de aplicación dash
    !pip install dash-bootstrap-components
    pbar.update(100)

[ ] import pandas as pd
import numpy as np
from sklearn.preprocessing import StandardScaler
import dash_bootstrap_components as dbc
from sys import argv

from rdkit import Chem
from rdkit.Chem import Descriptors
from rdkit.Chem import rdMolDescriptors
```

- RDKIT: <https://www.rdkit.org/docs/GettingStartedInPython.html>
- PANDAS: <https://pandas.pydata.org/>
- NUMPY: <https://numpy.org/>
- SKLEARN: <https://scikit-learn.org/stable/>

Anexo 10

Toma de pantalla del script realizado en Python usado para la generación scaffolds de las cuatro bases de datos.

```
[1] import sys
IN_COLAB = 'google.colab' in sys.modules
if IN_COLAB:
    !pip install pandas rdkit mols2grid useful-rdkit-utils

Mostrar salida oculta

[2] import pandas as pd
import numpy as np
from rdkit import Chem
from rdkit.Chem.rdMPPA import FragmentMol
from rdkit.Chem.rdRGroupDecomposition import RGroupDecompose
import useful_rdkit_utils as uru
from tqdm.auto import tqdm
import seaborn as sns
import matplotlib.pyplot as plt
from itertools import chain

[3] # Función para limpiar fragmentos
def cleanup_fragment(mol):
    for atm in mol.GetAtoms():
        atm.SetAtomMapNum(0)
        if atm.GetAtomicNum() == 0:
            atm.SetAtomicNum(1)
    mol = Chem.RemoveAllHs(mol)
    return mol, sum(1 for atm in mol.GetAtoms() if atm.GetAtomicNum() == 1)

scaffold_finder_script = """
import sys
from rdkit import Chem
from rdkit.Chem.rdMPPA import FragmentMol
from rdkit.Chem.rdRGroupDecomposition import RGroupDecompose
import itertools
import useful_rdkit_utils as uru
import pandas as pd
from tqdm.auto import tqdm
from glob import glob
```

```

import os
print(os.getcwd())
print(os.listdir())

/content
['.config', 'SCAFOLDS BD GENERAL.csv', 'scaffold_finder.py', 'sample_data']

import importlib.util
import sys

spec = importlib.util.spec_from_file_location("scaffold_finder", "scaffold_finder.py")
scaffold_finder = importlib.util.module_from_spec(spec)
sys.modules["scaffold_finder"] = scaffold_finder
spec.loader.exec_module(scaffold_finder)

# Ahora importa funciones desde ahi
from scaffold_finder import generate_fragments, find_scaffolds, get_molecules_with_scaffold

[7] import pandas as pd
from rdikit import Chem
from scaffold_finder import generate_fragments, find_scaffolds, get_molecules_with_scaffold
import mols2grid

[12] input_df = pd.read_csv("/content/BD GENERAL - SCALFOLDS.csv")

[13] # Renombrar columnas para que coincidan con el script
input_df = input_df.rename(columns={"NAME": "Name"})
input_df['mol'] = input_df['SMILES'].apply(Chem.MolFromSmiles)
input_df = input_df.dropna(subset=['mol']) # Eliminar moléculas inválidas
input_df['SMILES'] = input_df['mol'].apply(Chem.MolToSmiles)

# Generar fragmentos y encontrar andamiajes
mol_df, scaffold_df = find_scaffolds(input_df)

*** 24% 402/1688 [00:35+01:26, 14.850/s]

```

Anexo 11

Toma de pantalla de los parámetros usados para la creación de conformeros de las cuatro bases de datos en Data Warrior.

Structure of SMILES

Generate Conformers

Structure column: Structure of SMILES

Algorithm: Random, low energy bias

Initial torsions: From crystallographic database

Minimize energy: MMFF94s+ forcefield

Max. conformer count: 10 per stereo isomer

Write into file: ..\conformers.dwar Edit

File type: DataWarrior

Pool conformers of same compound

Remove small fragments

Neutralize remaining fragment

Skip compounds with more than 10 stereo isomers

Create proper protonation states for pH= 7.4

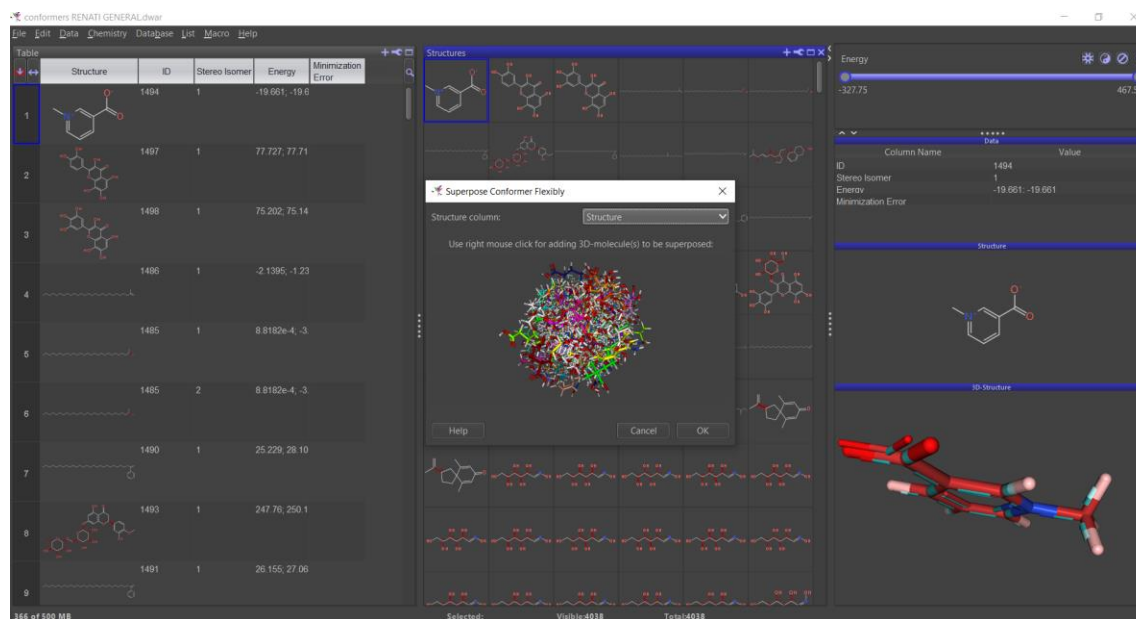
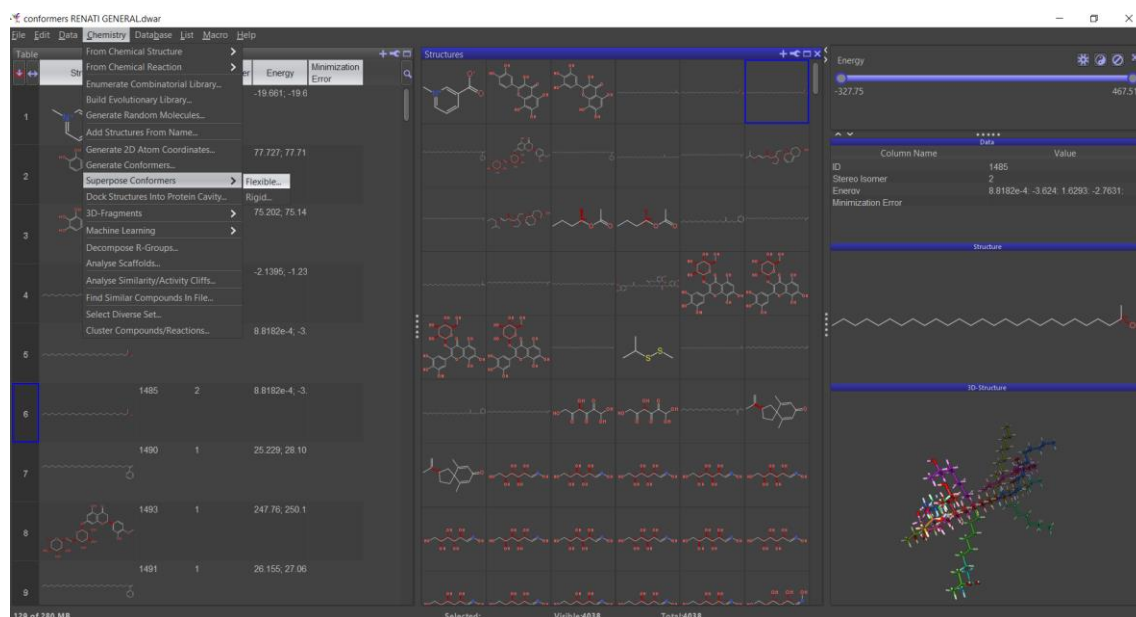
Help Cancel OK

Structure of SMILES	SMILES	Rack	Compound ID
1	O=C(O)C=Oc	RENATI	PP1
2	O=C(O)C=Cc	RENATI	PP2
3	CC(C)@H(C)=	RENATI	PP3
4	COc1ccc(-c2s	RENATI	PP4

Column Name	Data	Value
SMILES		O=C(O)C=Cc1ccc(O)c(O)c1
Rack		RENATI
Compound ID		PP1

Anexo 12

Toma de pantalla de la superposición realizada a las plantas peruanas vs las bases de datos de referencia en Data Warrior.



Anexo 14

Toma de pantalla del script para el análisis de la metodología ESPsim frente a las tres bases de datos de referencia.

```
jupyter CALCULO EPSIM Last Checkpoint: 2 months ago
File Edit View Run Kernel Settings Help Trusted
Python 3 (ipykernel)

[19]: # Primera celda: Importaciones
from rdkit import Chem
from rdkit.Chem import AllChem
from rdkit.Chem import rdmolfiles
from espim import EmbeddingScore
import pandas as pd

[21]: from rdkit.Chem
def load_and_preprocess_sdf(file_path):
    try:
        suppl = Chem.SDMSupplier(file_path, removeHs=False)
        molecules = [mol for mol in suppl if mol is not None]
        if not molecules:
            print(f"Advertencia: No se cargaron moléculas de {file_path}")
        return molecules
    except Exception as e:
        print(f"Error al cargar {file_path}: {str(e)}")
        return []

[23]: # Tercera celda: Carga de archivos
metabolites_file = "C:/Users/NEIVES/OneDrive/Escritorio/TESIS-ANTIINFLAMATORIOS/Resultados/ESPIM/Formatos DF general/RENATI_optimized2.sdf"
drugs_file = "C:/Users/NEIVES/OneDrive/Escritorio/TESIS-ANTIINFLAMATORIOS/Resultados/ESPIM/Formatos DF general/DrugBank_optimized.sdf"
metabolites = load_and_preprocess_sdf(metabolites_file)
drugs = load_and_preprocess_sdf(drugs_file)

[25]: # Tercera celda: Carga de archivos
metabolites_file = "C:/Users/NEIVES/OneDrive/Escritorio/TESIS-ANTIINFLAMATORIOS/Resultados/ESPIM/Formatos DF general/RENATI_optimized2.sdf"
drugs_file = "C:/Users/NEIVES/OneDrive/Escritorio/TESIS-ANTIINFLAMATORIOS/Resultados/ESPIM/Formatos DF general/DrugBank_optimized.sdf"
metabolites = load_and_preprocess_sdf(metabolites_file)
drugs = load_and_preprocess_sdf(drugs_file)
```

```
jupyter CALCULO EPSIM Last Checkpoint: 2 months ago
File Edit View Run Kernel Settings Help Trusted
Python 3 (ipykernel)

metabolites = load_and_preprocess_sdf(metabolites_file)
drugs = load_and_preprocess_sdf(drugs_file)

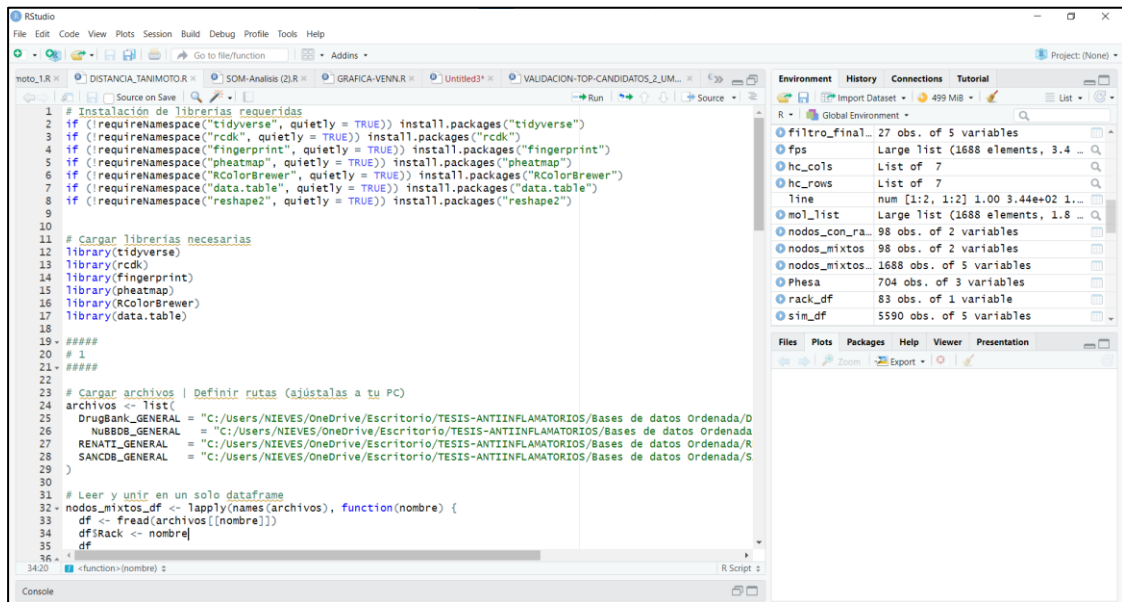
[24]: # Cuarta celda: Función de análisis ESPsim
def run_espim_analysis(metabolites, drugs, analysis_method="EmbeddingScore", num_confs=10, similarity_thresholds=None):
    results = []
    if similarity_thresholds is None:
        similarity_thresholds = ("High": 0.8, "Medium": 0.5) # Ajustar según necesidades
    for i, metabolite in enumerate(metabolites):
        for j, drug in enumerate(drugs):
            try:
                if analysis_method == "EmbeddingScore":
                    sim_shape, sim_esp = EmbeddingScore(metabolite, drug, renormalize=True, refNumConfs=num_confs)
                elif analysis_method == "getBestESP":
                    sim_shape, sim_esp = EmbeddingScore(metabolite, drug, renormalize=True, getBestESP=True)
                else:
                    raise ValueError("Método de análisis no válido. Use 'EmbeddingScore' o 'getBestESP'.")
                # Calcular Shape-ESP y clasificar
                shape_esp = sim_shape[0] * sim_esp[0]
                category = classify_similarity(sim_shape[0], sim_esp[0], shape_esp, similarity_thresholds)
                results.append({
                    "Metabolite_ID": f"Metabolite_{i+1}",
                    "Drug_ID": f"Drug_{j+1}",
                    "Shape": sim_shape[0],
                    "ESP": sim_esp[0],
                    "ShapeESP": shape_esp,
                    "Category": category
                })
            except Exception as e:
                print(f"Error al analizar metabolito {i+1} y fármaco {j+1}: {e}")
    return results

[27]: # Quinta celda: Función de clasificación
def classify_similarity(shape_sim, esp_sim, combined_score, thresholds):
    if combined_score >= thresholds["High"]:
        return "High"
    elif combined_score >= thresholds["Medium"]:
        return "Medium"
    else:
        return "Low"

[ ]: # Sexta celda: Ejecución del análisis y exportación de resultados
```

Anexo 15

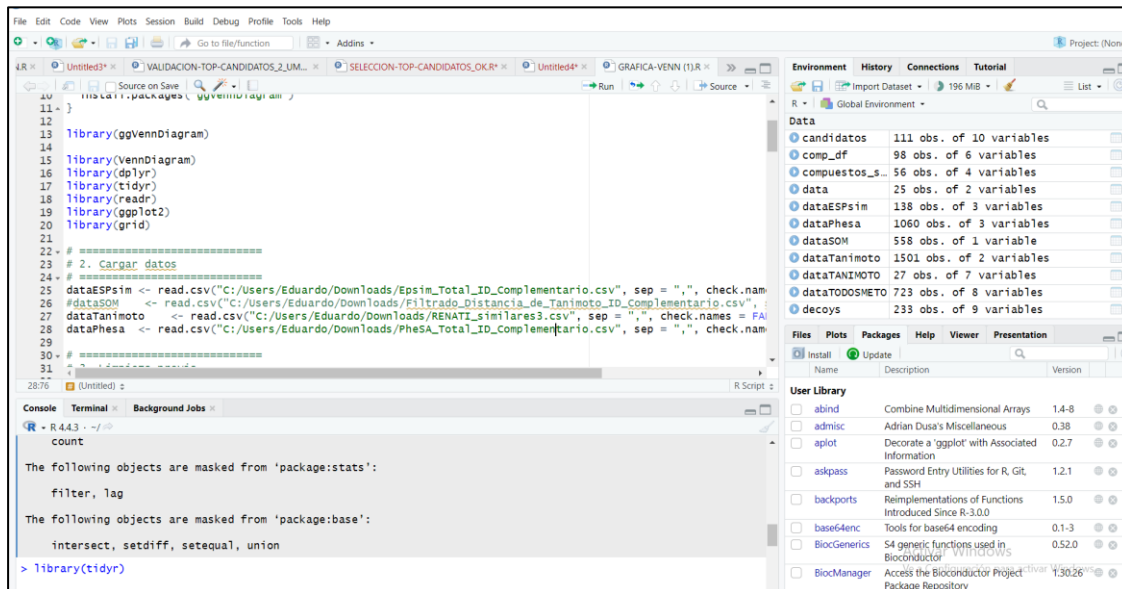
Toma de pantalla del script realizado en Rstudio para el análisis de la metodología Distancia de Tanimoto frente a las tres bases de datos de referencia.



```
1 # Instalación de librerías requeridas
2 if (!requireNamespace("tidyverse", quietly = TRUE)) install.packages("tidyverse")
3 if (!requireNamespace("rcdk", quietly = TRUE)) install.packages("rcdk")
4 if (!requireNamespace("fingerprint", quietly = TRUE)) install.packages("fingerprint")
5 if (!requireNamespace("pheatmap", quietly = TRUE)) install.packages("pheatmap")
6 if (!requireNamespace("RColorBrewer", quietly = TRUE)) install.packages("RColorBrewer")
7 if (!requireNamespace("data.table", quietly = TRUE)) install.packages("data.table")
8 if (!requireNamespace("reshape2", quietly = TRUE)) install.packages("reshape2")
9
10
11 # Cargar librerías necesarias
12 library(tidyverse)
13 library(rcdk)
14 library(fingerprint)
15 library(pheatmap)
16 library(RColorBrewer)
17 library(data.table)
18
19 #####
20 # 1
21 #####
22
23 # Cargar archivos | Definir rutas (ajústalas a tu PC)
24 archivos <- list(
25   DruBank_GENERAL = "C:/Users/NIEVES/OneDrive/Escritorio/TESIS-ANTIINFLAMATORIOS/Bases de datos Ordenada/D
26   NuBDD_GENERAL = "C:/Users/NIEVES/OneDrive/Escritorio/TESIS-ANTIINFLAMATORIOS/Bases de datos Ordenada
27   RENATI_GENERAL = "C:/Users/NIEVES/OneDrive/Escritorio/TESIS-ANTIINFLAMATORIOS/Bases de datos Ordenada/R
28   SANDB_GENERAL = "C:/Users/NIEVES/OneDrive/Escritorio/TESIS-ANTIINFLAMATORIOS/Bases de datos Ordenada/S
29 )
30
31 # Leer y unir en un solo dataframe
32 nodos_mixtos_df <- lapply(names(archivos), function(nombre) {
33   df <- fread(archivos[[nombre]])
34   df$Rack <- nombre
35   df
36 })
37
38 <-function>(nombre) {
```

Anexo 16

Toma de pantalla del script realizado en Rstudio para la creación del Diagrama de Venn de basado en las tres metodologías usadas.



```
10 library(ggVennDiagram)
11
12
13 library(VennDiagram)
14
15 library(dplyr)
16 library(tidyrr)
17 library(readr)
18 library(ggplot2)
19 library(grid)
20
21 #####
22 # 2. Cargar datos
23 #####
24
25 dataESPSim <- read.csv("C:/Users/Eduardo/Downloads/Epsim_Total_ID_Complementario.csv", sep = ",", check.names = FALSE)
26 dataSOM <- read.csv("C:/Users/Eduardo/Downloads/Filtrado-Distancia-de-Tanimoto-ID-Complementario.csv", sep = ",", check.names = FALSE)
27 dataTanimoto <- read.csv("C:/Users/Eduardo/Downloads/RENATI-similares3.csv", sep = ",", check.names = FALSE)
28 dataPhesa <- read.csv("C:/Users/Eduardo/Downloads/Phesa_Total_ID_Complementario.csv", sep = ",", check.names = FALSE)
29
30 #####
31
32 count
```

The following objects are masked from 'package:stats':
filter, lag

The following objects are masked from 'package:base':
intersect, setdiff, setequal, union

```
> library(tidyrr)
```

Anexo 17

Toma de pantalla del script realizado en Rstudio para el análisis del puntaje consenso obtenido del diagrama de Venn.

The screenshot displays the RStudio interface with a script in the editor. The code defines a data frame 'df' and calculates a consensus score based on several variables: lambda, gamma, tT, tP, tE, k, A, base, and Consensus. The environment pane on the right shows the objects created, including 'df' (344 obs. of 17 variables) and 'Consensus' (Large list (1688 elements, 3.4 ...)).

```
55 # Reemplazar NA por 0 en todo el dataframe
56 df[is.na(df)] <- 0
57
58 # -----
59 # Calcular Consensus
60 # -----
61 lambda <- 0.30
62 gamma <- 0.20
63 tT <- 0.70; tP <- 0.70; tE <- 0.70 # calibra con decoys
64
65 df <- df %>%
66 mutate(
67   sT = SCORE_Tanimoto,
68   sP = SCORE_PheSA,
69   sE = SCORE_ESPsim,
70   k = (sT >= tT) + (sP >= tP) + (sE >= tE), # número de métodos que pasan el umbral
71   A = k / 3, # fracción de metodologías que superan el corte
72   base = (pmax(sT, 1e-6) * pmax(sP, 1e-6) * pmax(sE, 1e-6))^(1/3), # score geométrico
73   Consensus = (1 - lambda) * base + lambda * A + gamma * pmin(sT, sP, sE) # score de consenso final
74 )
75
```

Anexo 18

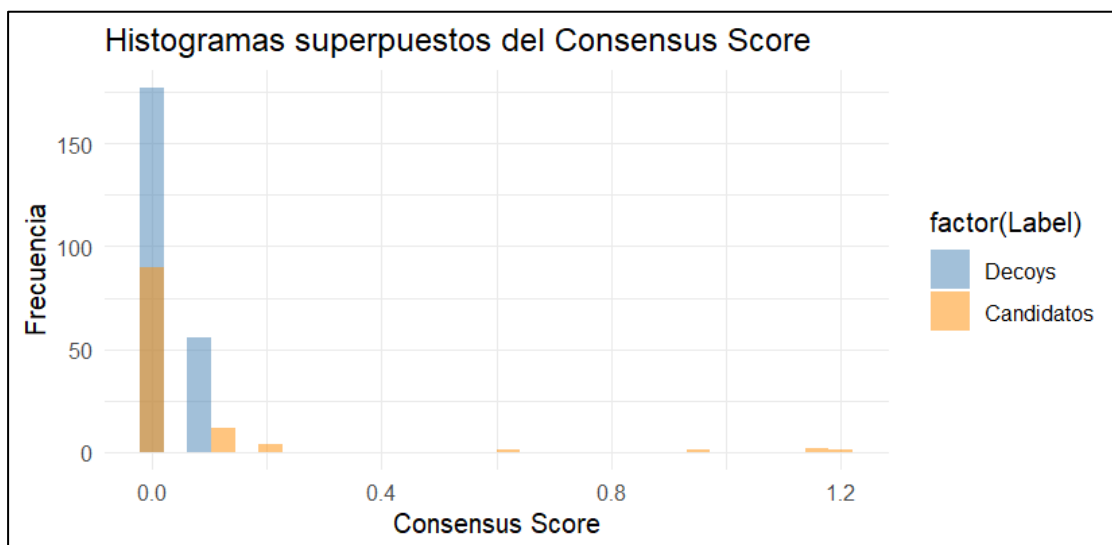
Toma de pantalla del script realizado en Rstudio para la validación de datos usando el diagrama de violín plot y percentil.

The screenshot shows RStudio with a script for violin plots and percentile calculation. The code uses 'geom_violin' and 'geom_boxplot' to compare 'Decoys' and 'Candidatos'. The environment pane shows the 'Distribución del Consensus Score en Candidatos' plot, which is a violin plot with a boxplot overlay, comparing 'Decoys' and 'Candidatos'.

```
113 vjust = -1, color = "red") +
114 labs(x = "Compuestos ordenados", y = "Consensus score",
115 title = "Curva de Elbow con umbral identificado") +
116 theme_minimal()
117
118 #violin plot
119
120 ggplot(df_eval, aes(x = factor(Label, labels = c("Decoys", "Candidatos")),
121 y = Consensus, fill = factor(Label))) +
122 geom_violin(trim = FALSE, alpha = 0.5) +
123 geom_boxplot(width = 0.1, outlier.color = "red") +
124 scale_y_log10() +
125 labs(x = "Grupo", y = "Consensus Score",
126 title = "Distribución del Consensus Score en Candidatos vs Decoys") +
127 theme_minimal() +
128 scale_fill_manual(values = c("steelblue", "darkorange"))
129
130 #2. METODO PERCENTIL
131
132 # calcular percentil del umbral
133 percentil_umbral <- ecdf(df$Consensus)(umbral_elbow) * 100
134
135
```

Anexo 19

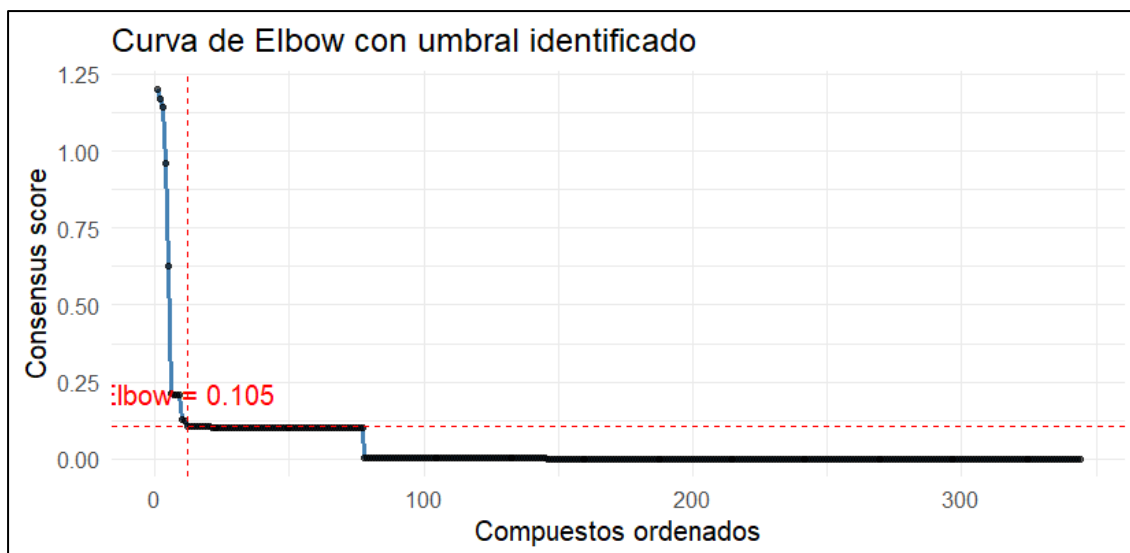
Histograma de frecuencia de la validación de datos realizadas de los candidatos de plantas peruanas versus los decoys (falsos positivos).



Nota: Ilustración realizada en RStudio.

Anexo 20

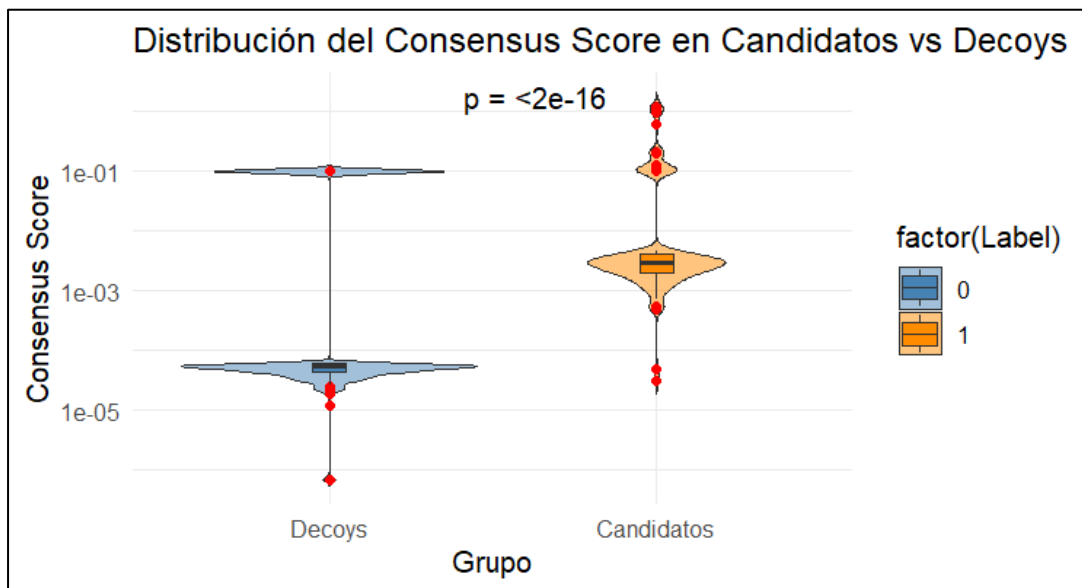
Obtención del umbral del puntaje consenso obtenido del análisis de las plantas peruanas, mediante el método del codo (Elbow method) para validación de datos.



Nota: Ilustración realizada en RStudio.

Anexo 21

Violin Plot de la distribución del Consensus Score entre falsos positivos (Decoys) y candidatos para validación del análisis consenso.



Nota: Ilustración realizada en RStudio.

Anexo 22

Tabla de enriquecimiento farmacológico de los 11 candidatos reportados. Ayacucho 2025.

ID	NOMBRE	MECANISMO DE ACCIÓN
PP36	Curcumina	Elimina elementos principales en las vías de transducción, de diferenciación y transformación cancerígena, inhibe las proteínas quinasas, la activación de c-Jun/AP-1, biosíntesis de prostaglandinas, además la actividad y la expresión de la ciclooxigenasa (COX2). ¹³¹
PP325	Glicina	Actúa a nivel del SNC, donde la glicina se une a la estrictina en el receptor NMDA y por un canal de cloruro. ¹³²
PP966	(E)-geraniol	Inhibe la síntesis de Leucotrienos, citocinas y quimiocinas. ¹³³
PP531	Ácido propiónico	Produce una inhibición débil y reversible en el COX1 y COX2. ¹³⁴
PP116	Desmetoxicurcumina	Suprime la síntesis de NO (óxido nítrico) en células microgriales N9, inhibiendo la expresión de proteínas y ARNm de la NO sintasa inducible. ¹³⁵
PP122	Dihidrocurcumina	Regula la expresión de ARNm y proteínas de SREBP-1C, PNPLA3 y PPAR α , asimismo, mejora la captación de pAKT y PI3K. Reduce los niveles de NO y RODS mediante el Nrf2 (Factor nuclear 2 relacionado a los oxidantes). ¹³⁶
PP1868	Ácido cítrico	Activa las neuronas sensoriales de la capsaicina, estas liberan taquininas, en el receptor NK ₂ y la bradicina, un efecto opuesto es la liberación de NO inducida por receptores Nk ₁ Y B ₂ . ¹³⁷
PP1440	β -nerol	Posible inhibición de la expresión de IL-13 y TNF- α , mediadores de la inflamación. ¹³⁸
PP1856	Eugenol	Probablemente suprime la expresión de la COX2y NF-Kappa B(NF-Kb). ¹³⁹ También disminuye las citocinas como el TNF- α y TNF- β . ¹⁴⁰

Anexo 23

Matriz de definición y operacionalización de variables.

Variable	Definición conceptual	Definición operacional	Dimensión	Indicadores	Escala de medición
Metabolitos de plantas peruanas que registran actividad antiinflamatoria en el repositorio RENATI y cuentan con información estructural en código <i>SMILES</i> en KNApSack.	Conjunto de compuestos secundarios obtenidos de plantas peruanas reportados con actividad antiinflamatoria en documentos de investigación del repositorio RENATI, cuya estructura química puede representarse mediante un código <i>SMILES</i> para permitir análisis quimioinformático.	Se identificarán y seleccionarán metabolitos secundarios a partir de una búsqueda sistematizada en RENATI. Posteriormente, se verificará su disponibilidad estructural en KNApSack en formato <i>SMILES</i> .	<ul style="list-style-type: none"> Información estructural 	<ul style="list-style-type: none"> Disponibilidad del código <i>SMILES</i>. Identificación única en KNApSack. Curación de datos. 	Nominal
Similaridad y diversidad estructural, molecular y fisicoquímica de los metabolitos de plantas peruanas con fármacos y productos naturales antiinflamatorios de referencia. Indicadores: Reglas de Lipinski, PCA, <i>Scaffolds</i> , Coeficiente	Comparación de de similaridad y diversidad entre plantas peruanas y de referencia.	Se realizará una comparación de similaridad y diversidad usando PCA (Análisis de componentes principales), que se basa en las reglas de Lipinski, scaffolds estructurales, además del uso de tres metodologías (Coeficiente de Tanimoto, PheSA SCORE, ESP-Sim)	<ul style="list-style-type: none"> Puntaje de las 5 reglas de Lipinski. 	<ul style="list-style-type: none"> MW<500 Daltons -2<Log p<5 HBA<10 HBD<5 RB<10 TPSA<140° 	Escala de razón (Cuantitativa) y se analizó como cualitativa nominal dicotómica “Cumple o no cumple”.
			<ul style="list-style-type: none"> PCA 	<ul style="list-style-type: none"> Componentes principales (PC1, PC2, PC3). 	Nominal (cualitativo).
			<ul style="list-style-type: none"> Scaffolds Puntaje del coeficiente de Tanimoto. Puntaje de PheSA. Puntaje de ESP-Sim. 	<ul style="list-style-type: none"> Número y frecuencia de núcleos base Mayor o igual 0,60 	Nominal
		Se efectuará un cálculo consenso tomando las tres metodologías usadas para poder detectar los candidatos.	<ul style="list-style-type: none"> Puntaje Consenso. 	<ul style="list-style-type: none"> Mayor al umbral (0,105) 	Escala de razón (Cuantitativo continuo).

Anexo 24

Matriz de consistencia

Título: Metabolitos bioactivos antiinflamatorios de la flora peruana identificados por análisis quimioinformático comparativo. Ayacucho, 2025.

Autor: Yasmin Nieves CAYLLAHUA ORE

Formulación del problema	Objetivos	Hipótesis	Variables	Metodología
¿Cuáles serán los metabolitos de plantas peruanas reportados con actividad antiinflamatoria identificados mediante análisis comparativo quimioinformático?	Objetivos Generales Identificar metabolitos antiinflamatorios de la flora peruana mediante análisis quimioinformático comparativo, basado en algoritmos de similaridad y diversidad, con perfiles similares a fármacos y productos naturales de referencia.	H1: Los metabolitos de las plantas peruanas que reporten actividad antiinflamatoria presentan similitud significativa frente a fármacos y productos naturales de referencia, lo que sugiere su potencial como candidatos con actividad antiinflamatoria.	Variable 1: Metabolitos de plantas peruanas que registran actividad antiinflamatoria en el repositorio RENATI y cuentan con información estructural en código <i>SMILES</i> en <i>KNApSack</i> .	Alcance de investigación Descriptiva. Diseño de investigación Diseño comparativo con enfoque computacional. Población y muestra Población: 56 plantas peruanas que reportan actividad antiinflamatoria en la base de datos de RENATI desde 1998 hasta febrero del 2025. Muestra: 25 plantas peruanas que reportan actividad antiinflamatoria en base de datos RENATI y que registran información de metabolitos en la plataforma <i>KNApSack</i> . Unidad de análisis: Metabolitos de plantas peruanas en código <i>SMILES</i> . Técnica y método de recolección de datos Se empleó programas de código abierto, Jupyter Notebook, Rstudio y Data Warrior para la realización de los siguientes análisis: • Diversidad: Reglas de Lipinski, Espacio químico (PCA) y <i>scaffolds</i> .
Problemas específicos • ¿Cómo incorporar y normalizar los metabolitos de plantas peruanas con potencial antiinflamatorio del repositorio RENATI que cuente con información estructural en <i>KNApSack</i> para generar una base de datos consistente y apta para análisis quimioinformático?	Objetivos Específicos • Generación de base de datos de los metabolitos de la flora peruana que registran actividad antiinflamatoria en el repositorio RENATI y que cuentan con información estructural reportada en la base de datos <i>KNApSack</i> • Realizar análisis quimioinformático comparativo de similaridad y diversidad de los metabolitos de plantas peruanas que reportan	Hipótesis específicas H1.1. La integración de RENATI y <i>KNApSack</i> permite construir una base de datos estandarizada, con estructuras unívocas y con cobertura suficiente para análisis posteriores. H1.2. Los metabolitos de plantas peruanas con actividad antiinflamatoria registrados en RENATI presentan similitud y diversidad molecular con respecto a los fármacos de referencial, lo que sugiere	Variable 2: Similaridad y diversidad estructural, molecular y fisicoquímica de los metabolitos de plantas peruanas con fármacos y productos naturales antiinflamatorios de referencia. Indicadores: Reglas de Lipinski, PCA, <i>Scaffolds</i> , Coeficiente de	

Firmado digitalmente por GABRIELA BELLIDO MUJICA
Fecha: 2025.12.15 16:31:10 -05'00'

Universidad Nacional de San Cristóbal de Huancayo
Facultad de Ciencias de la Salud
Departamento Académico de Medicina Humana
Dr. Enrique Javier Aguilar Felices
001001213

Firmado digitalmente por Enrique Javier Aguilar Felices
Fecha: 2025.12.15 14:18:03 -05'00'

- ¿En qué medida los metabolitos de plantas peruanas con actividad antiinflamatoria reportados en RENATI presentan similiaridad y diversidad en comparación con los fármacos de referencia?
- ¿En qué medida los metabolitos de plantas peruanas con actividad antiinflamatoria reportados en RENATI presentan similiaridad y diversidad en comparación con los productos naturales de referencia?
- ¿Cuáles son los metabolitos de plantas peruanas que presentan actividad antiinflamatoria que se evidencian mediante herramientas quimioinformáticas?

actividad antiinflamatoria en RENATI con fármacos antiinflamatorios de referencia.

- Realizar análisis quimioinformático comparativo de similiaridad y diversidad de los metabolitos de plantas peruanas que reportan actividad antiinflamatoria en RENATI con productos naturales de referencia.
- Identificar metabolitos de plantas peruanas con potencial actividad antiinflamatoria.

su potencial como candidato antiinflamatorio.

H1.3. Los metabolitos de plantas peruanas con actividad antiinflamatoria registrados en RENATI presentan similiaridad y diversidad molecular con respecto a los productos naturales de referencia, lo que sugiere su potencial como candidato antiinflamatorio.

H1.4. Los metabolitos de plantas peruanas presentan estructuras químicas con alta similiaridad frente a fármacos y productos naturales lo que sugiere su potencial antiinflamatorio.

Tanimoto, *PheSA Score* y *ESP-Sim*.

- Similaridad: coeficiente de Tanimoto, PheSA Score y Shape similarity (ESP-Sim).

Análisis de datos

Los resultados de las tres metodologías fueron realizados por Rstudio mediante un puntaje Consenso.

Validación de datos

Se realizó mediante el método del codo.

Firmado digitalmente por MSc. Kirianova Godoy Bautista
 Fecha: 2025.12.15 17:08:38 -05'00'

Firmado digitalmente por Enrique Javier Aguilar Felices
 Fecha: 2025.12.15 14:17:12 -05'00'

Universidad Nacional de San Cristóbal de Huamanga
 Facultad de Ciencias de la Salud
 Departamento Académico de Medicina Humana

Dr. Enrique Javier Aguilar Felices
 DNI 06918713

Firmado digitalmente por GABRIELA BELLIDO MUJICA
 Fecha: 2025.12.15 16:24:35 -05'00'

ACTA DE SUSTENTACIÓN DE TESIS

RESOLUCIÓN DECANAL N° 940–2025-UNSCH-FCSA-D

BACHILLER: Yasmin Nieves CAYLLAHUA ORE

En la ciudad de Ayacucho, siendo las nueve de la mañana del día diecinueve de diciembre del dos mil veinticinco, se reunieron en el Auditorio de la Escuela Profesional de Enfermería, los docentes miembros del jurado evaluador de sustentación, para el acto de sustentación de trabajo de tesis titulado: **Metabolitos bioactivos antiinflamatorios de la flora peruana identificados por análisis quimiinformático comparativo**. Presentado por la bachiller Yasmin Nieves CAYLLAHUA ORE para optar el título profesional de Químico Farmacéutico.

El Jurado evaluador de sustentación está conformado por:

Presidente (delegado por el Decano)	: Prof. Enrique Javier Aguilar Felices
Miembros	: Prof. Enrique Javier Aguilar Felices : Prof. Gabriela Bellido Mujica
4to jurado	: Prof. Mónica Gómez Quispe
asesor	: Prof. Kirianova Godoy Bautista
secretaria Docente	: Prof. Priscila Licas Conde

Con el quorum de reglamento se dio por inicio la sustentación de tesis, el presidente de la comisión pide a la secretaria docente dar lectura a los documentos presentados por la recurrente, y da algunas indicaciones a la sustentante.

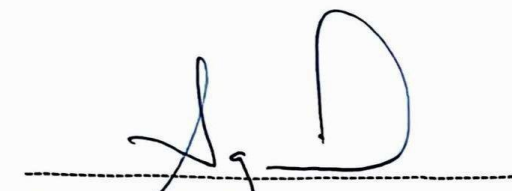
Da inicio la exposición la Bachiller: Yasmin Nieves CAYLLAHUA ORE, y una vez concluida la exposición, el presidente de la comisión solicita a los miembros del jurado evaluador realizar sus respectivas preguntas. Acto seguido, después de realizar algunas observaciones, da pase al asesor de tesis profesora Kirianova Godoy Bautista para que pueda realizar algunas aclaraciones y comentarios.

Concluida la ronda de preguntas, el presidente invita al sustentante para abandonar el auditorio y se pueda proceder con la calificación.

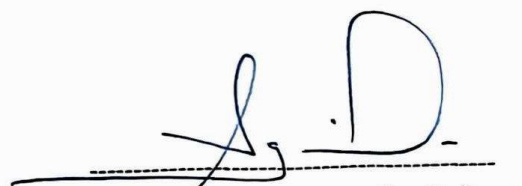
RESULTADO DE LA EVALUACIÓN FINAL

JURADOS	TEXTO	EXPOSICIÓN	PREGUNTAS	P.FINAL
Prof. Enrique Javier Aguilar Felices	18	18	18	18
Prof. Gabriela Bellido Mujica	18	18	18	18
Prof. Mónica Gómez Quispe	19	19	19	19
PROMEDIO FINAL:			18	

De la evaluación realizada por los miembros del jurado calificador, llegaron al siguiente resultado: Aprobar a la Bachiller Yasmin Nieves CAYLLAHUA ORE, quien obtuvo la nota final de **Dieciocho (18)** para la cual los miembros del jurado evaluador firman al pie del presente. Siendo las 11:00 horas del día, se da por concluido el presente acto académico.



Prof. Enrique Javier Aguilar Felices
Presidente



Prof. Enrique Javier Aguilar Felices
Miembro



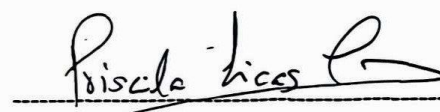
Prof. Gabriela Bellido Mujica
Miembro



Prof. Mónica Gómez Quispe
Miembro



Prof. Kirjanova Godoy Bautista
Asesora



Prof. Priscila Licas Conde
Secretaria Docente

UNIVERSIDAD NACIONAL DE SAN CRISTÓBAL DE HUAMANGA
FACULTAD DE CIENCIAS DE LA SALUD

El Instructor en Segunda Instancia, en virtud de la RCU N.° 039-2021-UNSCH-CU, y en calidad de director de la Escuela Profesional de Farmacia y Bioquímica, emite la presente

CONSTANCIA

DE ORIGINALIDAD DE TRABAJO DE INVESTIGACIÓN

A Yasmin Nieves CAYLLAHUA ORE, Bachiller de la Escuela Profesional de Farmacia y Bioquímica de la Facultad de Ciencias de la Salud, en mérito a que la tesis titulada: Metabolitos bioactivos antiinflamatorios de la flora peruana identificados por análisis quimioinformático comparativo; ha alcanzado un índice de similitud de 17% (diecisiete); cumpliendo satisfactoriamente lo establecido en el Art. 13 del Reglamento de Originalidad de Trabajos de investigación de la Universidad Nacional de San Cristóbal de Huamanga mediante el uso del SOFTWARE TURNITIN.

En ese sentido, se emite la presente constancia en señal de conformidad.

Ayacucho, 15 de diciembre de 2025.


**UNIVERSIDAD NACIONAL DE SAN
CRISTÓBAL DE HUAMANGA**
ESCUELA PROFESIONAL DE FARMACIA Y BIOQUÍMICA


.....
Marco R. Aronés Jara
DIRECTOR

Metabolitos bioactivos antiinflamatorios de la flora peruana identificados por análisis quimioinformático comparativo

por Yasmin Nieves CAYLLAHUA ORE

Fecha de entrega: 15-dic-2025 04:48a. m. (UTC-0500)

Identificador de la entrega: 2846676739

Nombre del archivo: Borrador_de_Tesis.pdf (3.66M)

Total de palabras: 24909

Total de caracteres: 146735

Metabolitos bioactivos antiinflamatorios de la flora peruana identificados por análisis quimioinformático comparativo

INFORME DE ORIGINALIDAD



FUENTES PRIMARIAS

1	www.frontiersin.org Fuente de Internet	1%
2	ouci.dntb.gov.ua Fuente de Internet	1%
3	6f80db30-2eaf-4545-995b-7499545952c6.filesusr.com Fuente de Internet	1%
4	jcheminf.biomedcentral.com Fuente de Internet	1%
5	mts.intechopen.com Fuente de Internet	1%
6	www.biorxiv.org Fuente de Internet	<1%
7	pmc.ncbi.nlm.nih.gov Fuente de Internet	<1%
8	repozitorij.pmf.unizg.hr Fuente de Internet	<1%
9	amsdottorato.unibo.it Fuente de Internet	<1%
10	theses.hal.science Fuente de Internet	<1%
11	www.researchsquare.com Fuente de Internet	<1%

12	www.db-thueringen.de Fuente de Internet	<1 %
13	ppj.phypha.ir Fuente de Internet	<1 %
14	www.mdpi.com Fuente de Internet	<1 %
15	dokumen.pub Fuente de Internet	<1 %
16	journals.innovareacademics.in Fuente de Internet	<1 %
17	link.springer.com Fuente de Internet	<1 %
18	ebin.pub Fuente de Internet	<1 %
19	Jingbo Yang, Yiyang Cai, Kairui Zhao, Hongbo Xie, Xiujie Chen. "Concepts and applications of chemical fingerprint for hit and lead screening", Drug Discovery Today, 2022 Publicación	<1 %
20	Gene Jong Yu, Federico Ranieri, Vincenzo Di Lazzaro, Marc A Sommer, Angel V Peterchev, Warren M Grill. "Circuits and mechanisms for TMS-induced corticospinal waves: Connecting sensitivity analysis to the network graph", Cold Spring Harbor Laboratory, 2024 Publicación	<1 %
21	www.nature.com Fuente de Internet	<1 %
22	Submitted to Queen Mary and Westfield College Trabajo del estudiante	<1 %

23	Submitted to University of Brighton Trabajo del estudiante	<1 %
24	le.utah.gov Fuente de Internet	<1 %
25	Submitted to University of Queensland Trabajo del estudiante	<1 %
26	Gleb V. Sitnikov, Nelly I. Zhokhova, Yury A. Ustynyuk, Alexandre Varnek, Igor I. Baskin. "Continuous indicator fields: a novel universal type of molecular fields", Journal of Computer-Aided Molecular Design, 2014 Publicación	<1 %
27	www.ihcan-mag.com Fuente de Internet	<1 %
28	"Open Access Databases and Datasets for Drug Discovery", Wiley, 2024 Publicación	<1 %
29	inabj.org Fuente de Internet	<1 %
30	es.scribd.com Fuente de Internet	<1 %
31	"T220 block 3 Environmental management in organisations ISBN9781473035560", Open University Publicación	<1 %
32	Submitted to University of Monastir Trabajo del estudiante	<1 %
33	pubmed.ncbi.nlm.nih.gov Fuente de Internet	<1 %
34	m.moam.info Fuente de Internet	<1 %

35	www.dovepress.com Fuente de Internet	<1 %
36	Submitted to Univerza v Ljubljani Trabajo del estudiante	<1 %
37	ijcc.chemoprev.org Fuente de Internet	<1 %
38	jms.ump.edu.pl Fuente de Internet	<1 %
39	cris.ucsm.edu.pe Fuente de Internet	<1 %
40	benthamscience.com Fuente de Internet	<1 %
41	jneuroinflammation.biomedcentral.com Fuente de Internet	<1 %
42	repositorio.unal.edu.co Fuente de Internet	<1 %
43	www.scielo.br Fuente de Internet	<1 %
44	Submitted to University of Birmingham Trabajo del estudiante	<1 %
45	eprints.whiterose.ac.uk Fuente de Internet	<1 %
46	www.excli.de Fuente de Internet	<1 %
47	Mohammad Ashrafuddin Khan, Se Yeon Lee, Youngbuhm Huh, Ja-Eun Kim et al. "Anti-proliferative Profiling of 6,000 Representative Compounds from the Korean Chemical Bank Diversified Compound Library in De-differentiated Schwann Cells", <i>Anatomy & Biological Anthropology</i> , 2023	<1 %

48 bmcartificialintel.biomedcentral.com <1 %
Fuente de Internet

49 www.difacquim.com <1 %
Fuente de Internet

50 Adekunle B. Rowaiye, Abiodun Abidemi Folarin, Tobilola Akingbade, Joy C. Okoli et al. "Advancing predictive modeling in computational chemistry through quantum chemistry, molecular mechanics, and machine learning", Discover Chemistry, 2025 <1 %
Publicación

51 Submitted to King's College <1 %
Trabajo del estudiante

52 assets-eu.researchsquare.com <1 %
Fuente de Internet

53 biblio.unibe.ch <1 %
Fuente de Internet

54 bonndoc.ulb.uni-bonn.de <1 %
Fuente de Internet

55 diglib.tugraz.at <1 %
Fuente de Internet

56 journalajrib.com <1 %
Fuente de Internet

57 www.preprints.org <1 %
Fuente de Internet

58 backend.orbit.dtu.dk <1 %
Fuente de Internet

59 revistas.udea.edu.co <1 %
Fuente de Internet

60	Fuente de Internet	<1 %
61	Submitted to University of Portsmouth Trabajo del estudiante	<1 %
62	dblp.dagstuhl.de Fuente de Internet	<1 %
63	www.scfbio-iitd.res.in Fuente de Internet	<1 %
64	www.thieme-connect.com Fuente de Internet	<1 %
65	Submitted to University of Wales central institutions Trabajo del estudiante	<1 %
66	atrium.lib.uoguelph.ca Fuente de Internet	<1 %
67	ds.inflibnet.ac.in Fuente de Internet	<1 %
68	op.niscair.res.in Fuente de Internet	<1 %
69	Daniel A. Acuña-Jiménez, Jose Rodríguez-Zúñiga, Daniela Gutiérrez-Ramírez, Ricardo Quesada-Grosso et al. "Natural Products Repository of Costa Rica (NAPRORE-CR): An open-access database", American Chemical Society (ACS), 2025 Publicación	<1 %
70	academic.oup.com Fuente de Internet	<1 %
71	api.crossref.org Fuente de Internet	<1 %
72	dergipark.org.tr Fuente de Internet	<1 %

73	eprints.usm.my Fuente de Internet	<1 %
74	www.adaptation-fund.org Fuente de Internet	<1 %
75	d.docksci.com Fuente de Internet	<1 %
76	repositorio.ucsm.edu.pe Fuente de Internet	<1 %
77	Yi-Han Chiu, Wei-Ling Chou, Min-Chi Ko, Jun-Cheng Liao, Tse-Hung Huang. "Curcumin mitigates obesity-driven dysbiosis and liver steatosis while promoting browning and thermogenesis in white adipose tissue of high-fat diet-fed mice", <i>The Journal of Nutritional Biochemistry</i> , 2025 Publicación	<1 %
78	Submitted to Asian Institute of Technology Trabajo del estudiante	<1 %
79	cdr.lib.unc.edu Fuente de Internet	<1 %
80	journal.ppmi.web.id Fuente de Internet	<1 %
81	www.jyoungpharm.org Fuente de Internet	<1 %
82	Φλωράκης, Εμμανουήλ Ν.. "Προβλεψη και Αναλυση της Αποδοσης σε Αγωνες Αυτοκινητου Formula 1", <i>University of Piraeus (Greece)</i> , 2025 Publicación	<1 %
83	David Julian McClements. "Nano-enabled personalized nutrition: Developing multicomponent-bioactive colloidal delivery	<1 %

systems", Advances in Colloid and Interface
Science, 2020

Publicación

84 Yi Liu, Zhen Tian, Bohong Xu, De-Xin Kong.
"Navigating the chemical space of natural
products from diverse sources: a
chemoinformatics perspective",
Phytochemistry Reviews, 2025

Publicación

85 e-journal.sttlevinus-rumaseb.ac.id

Fuente de Internet

86 ehp.niehs.nih.gov

Fuente de Internet

87 d197for5662m48.cloudfront.net

Fuente de Internet

88 pandas.pydata.org

Fuente de Internet

89 "Biotechnology of Medicinal Plants with
Antiallergy Properties", Springer Science and
Business Media LLC, 2024

Publicación

90 Submitted to University of Mauritius

Trabajo del estudiante

91 alzres.biomedcentral.com

Fuente de Internet

92 epdf.pub

Fuente de Internet

93 ethnobotanyjournal.org

Fuente de Internet

94 journalijsra.com

Fuente de Internet

95	Submitted to University of Leicester Trabajo del estudiante	<1 %
96	repositorio.ucv.edu.pe Fuente de Internet	<1 %
97	www.tjnpr.org Fuente de Internet	<1 %
98	"Computer-Aided and Machine Learning-Driven Drug Design", Springer Science and Business Media LLC, 2024 Publicación	<1 %
99	Bo Weidema. "Assessing the environmental impact of agriculture", Burleigh Dodds Science Publishing, 2019 Publicación	<1 %
100	Rainer W. Bussmann, Narel Y. Paniagua Zambrana, Carolina Romero, Robbie E. Hart. "Astonishing diversity—the medicinal plant markets of Bogotá, Colombia", Journal of Ethnobiology and Ethnomedicine, 2018 Publicación	<1 %
101	cibfar.ifsc.usp.br Fuente de Internet	<1 %
102	hypothes.is Fuente de Internet	<1 %
103	jcheminf.springeropen.com Fuente de Internet	<1 %
104	marinespecies.eu Fuente de Internet	<1 %
105	pubs.acs.org Fuente de Internet	<1 %
106	research.aston.ac.uk Fuente de Internet	<1 %

107	revistas.udes.edu.co Fuente de Internet	<1 %
108	thieme-connect.com Fuente de Internet	<1 %
109	www.clonerresources.com Fuente de Internet	<1 %
110	Joseph L. Durant, Burton A. Leland, Douglas R. Henry, James G. Nourse. "Reoptimization of MDL Keys for Use in Drug Discovery", Journal of Chemical Information and Computer Sciences, 2002 Publicación	<1 %
111	Submitted to University of Bradford Trabajo del estudiante	<1 %
112	boristheses.unibe.ch Fuente de Internet	<1 %
113	f1000research.com Fuente de Internet	<1 %
114	ir.unilag.edu.ng Fuente de Internet	<1 %
115	pharmacophorejournal.com Fuente de Internet	<1 %
116	rechtinnovativ.online Fuente de Internet	<1 %
117	refubium.fu-berlin.de Fuente de Internet	<1 %
118	www.academon.com Fuente de Internet	<1 %
119	www.cellphysiolbiochem.com Fuente de Internet	<1 %

120	www.chimia.ch Fuente de Internet	<1 %
121	www2.ucsm.edu.pe Fuente de Internet	<1 %
122	Mousumi Khatun, Krishna Kumar, Ayana Baidya, Rajiv Kumar Mondal et al. "Variability in the Responses of Hepatitis B Virus D-Subgenotypes to Antiviral Therapy: Designing Pan-D-Subgenotypic Reverse Transcriptase Inhibitors", Journal of Virology, 2022 Publicación	<1 %
123	uvidok.rcub.bg.ac.rs Fuente de Internet	<1 %
124	José Esteban Jiménez, Pedro Juárez, Armando Díaz. "Checklist of the vascular flora of Reserva Biológica San Luis, Costa Rica", Check List, 2016 Publicación	<1 %
125	en.wikipedia.org Fuente de Internet	<1 %
126	skud.slu.se Fuente de Internet	<1 %
127	Howard Friel, Sally Frautschy. "Pleiotropic Theory", Elsevier BV, 2017 Publicación	<1 %
128	Submitted to Napier University Trabajo del estudiante	<1 %
129	bioinfo.nrc.ca Fuente de Internet	<1 %
130	www.coursehero.com Fuente de Internet	<1 %

131

www.medrxiv.org

Fuente de Internet

<1 %

132

"Fundamental Concepts", Walter de Gruyter GmbH, 2020

Publicación

<1 %

133

irbackend.kiu.ac.ug

Fuente de Internet

<1 %

Excluir citas

Activo

Excluir coincidencias < 30 words

Excluir bibliografía

Activo