DISEÑO DE UN MODELO CLASIFICATORIO DE PREDICCIÓN PARA LA SELECCIÓN DE MOLÉCULAS CON POTENCIAL ACTIVIDAD INHIBITORIA CONTRA EL VIRUS DE LA INMUNODEFICIENCIA HUMANA TIPO I (VIH-1).

LAURA VALENTINA ARIAS VARGAS

Universidad ECCI

Facultad de Ingeniería

Programa de Tecnología en Procesos Químicos Industriales

Bogotá D.C., Colombia

2024

DISEÑO DE UN MODELO CLASIFICATORIO DE PREDICCIÓN PARA LA SELECCIÓN DE MOLÉCULAS CON POTENCIAL ACTIVIDAD INHIBITORIA CONTRA EL VIRUS DE LA INMUNODEFICIENCIA HUMANA TIPO I (VIH-1).

LAURA VALENTINA ARIAS VARGAS (107696)

Trabajo de investigación presentado como requisito parcial para optar al título de:
Tecnología en Procesos Químicos Industriales

Línea de investigación: Diseño e intensificación de procesos químicos y bioquímicos

Director:
PhD. CHONNY ALEXANDER HERRERA ACEVEDO

Universidad ECCI
Facultad de Ingeniería
Programa de Tecnología en Procesos Químicos Industriales
Bogotá D.C., Colombia
2024

AGRADECIMIENTOS

Gracias a la universidad ECCI por permitirme desarrollar este proyecto al proporcionar los recursos y el entorno académico necesarios para llevar a cabo mi investigación.

También, agradezco a mi director de tesis, el Dr. Chonny Alexander Herrera Acevedo, por su inestimable orientación, apoyo y consejos a lo largo de este proceso. Su experiencia y conocimientos han sido fundamentales para el desarrollo de este trabajo. Así como a los evaluadores: Dr. Hernando Curtidor y el Prof. Francisco Vásquez por su apoyo y tiempo para la evaluación de este trabajo.

Finalmente, agradezco a mi familia, que con su apoyo incondicional me han permitido culminar esta etapa de mi vida profesional.

A mi familia y Dhante: Por su amor y apoyo incondicional.

TABLA DE CONTENIDO

LISTA DE TABLAS

LISTA DE FIGURAS

LISTA DE SÍMBOLOS Y ABREVIATURAS

ACC - Exactitud

ADMET - Absorción, Distribución, Metabolismo, Excreción y Toxicidad

Å - Angstrom

ARV - Antirretroviral

ART - Terapia Antirretroviral

ATP - Adenosín Trifosfato

CDC - Centros para el Control y la Prevención de Enfermedades

ChemDraw - Software de dibujo químico

ChEMBL - Base de datos de información química, bioactiva y biofarmacéutica

CI_{50} - Concentración Inhibitoria 50%

CRF - Formas Recombinantes Circulantes

CSV - Valores Separados por Comas

C-H - Enlace Carbono-Hidrógeno

DFT - Teoría del Funcional de la DenSIDAd

EM - Energía de cada molécula evaluada en el *docking*

FDA - Administración de Alimentos y Medicamentos

HB - Puente de Hidrógeno

HIV-1 - Virus de la Inmunodeficiencia Humana tipo 1

KNIME - Plataforma de Minería de Información de Konstanz

MCC - Coeficiente de Correlación de Matthews

MD - Dinámica Molecular

NADPH - Nicotinamida Adenina Dinucleótido Fosfato reducido

NNRTI - Inhibidores de la Transcriptasa Reversa No Nucleósidos

NPV - Valor Predictivo Negativo

OMS - Organización Mundial de la Salud

ONUSIDA - Programa Conjunto de las Naciones Unidas sobre el VIH/SIDA

PCA - Análisis de Componentes Principales

PDB - Banco de Datos de Proteínas

Pc - Probabilidad combinada

PPV - Valor Predictivo Positivo

QSAR - Relación Cuantitativa Estructura-Actividad

ROC - Característica Operativa del Receptor

RMSD - Desviación Cuadrática Media

SDF - Formato de Archivo de Datos Estructurales

SP - Probabilidad del *docking* molecular

SPC - Especificidad

TL-3 - Ligando de referencia

UNAIDS - Programa Conjunto de las Naciones Unidas sobre el VIH/SIDA

VIH - Virus de la Inmunodeficiencia Humana

VIH-1 - Virus de la Inmunodeficiencia Humana tipo 1

VIH-2 - Virus de la Inmunodeficiencia Humana tipo 2

VdW - Fuerzas de Van der Waals

VPR - Sensibilidad

W&B - Weights & Bias

RESUMEN

El VIH-1 ha representado un desafío de salud pública desde su primera aparición en 1980. Por esta razón, ha sido de vital importancia implementar diferentes metodologías que permitan el desarrollo de fármacos eficaces para su control y posible eliminación. Los modelos predictivos computacionales emergen como una alternativa prometedora para desarrollar nuevas quimioterapias, como lo ejemplifica el Amprenavir (inhibidor de proteasa), cuyo desarrollo se benefició de métodos computacionales.

Este estudio se enfoca en el desarrollo y evaluación de modelos predictivos para identificar compuestos activos contra la proteasa del VIH-1, utilizando herramientas de bioinformática y *docking* molecular. Se utilizó el software KNIME 5.2.5 para construir un modelo de predicción robusto capaz de discriminar eficazmente entre compuestos activos e inactivos, con altos niveles de exactitud, precisión y sensibilidad en la selección de candidatos prometedores.

Además, se llevó a cabo un análisis exhaustivo de *docking* molecular que identificó un diterpeno clerodano y cuatro diterpenos cauranos con alta afinidad hacia el sitio activo de la proteasa del VIH-1. Entre ellos, el compuesto **109** destacó significativamente por su energía de unión más baja (-128,49 kJ/mol), indicando una interacción fuerte y estable. Los compuestos restantes (**234**, **235**, **231** y **230**) también mostraron energías de unión favorables, sustentadas por interacciones clave como enlaces de hidrógeno y fuerzas de Van der Waals.

El análisis detallado de las interacciones moleculares reveló que el compuesto **109** presentó interacciones pi-sigma, pi-alquilo y alquilo que fortalecen la estabilidad del complejo ligando-proteína. Asimismo, el ligando TL-3 mostró características únicas como interacciones pi-amida y carga atractiva, contribuyendo a su mayor afinidad comparativa.

En conclusión, estos hallazgos sugieren que los diterpenos clerodanos y los diterpenos cauranos, especialmente el compuesto **109** derivado de Baccharis flabellata, representan prometedores candidatos para el desarrollo de inhibidores de la proteasa del VIH-1, destacando su potencial aplicación terapéutica en la lucha contra esta enfermedad viral.

1. INTRODUCCIÓN

El Virus de la Inmunodeficiencia Humana (VIH) representa uno de los mayores desafíos en salud pública a nivel global, debido a su alta capacidad de transmisión y las graves repercusiones que tiene en la población. A lo largo de los años, se ha observado un aumento constante de su incidencia, especialmente en regiones como África, América Latina y el Caribe (*UNAIDS, 2023*).

El VIH ha sido históricamente problemático debido a la falta de un tratamiento definitivo para su erradicación. Los tratamientos actuales se basan en antirretrovirales, que inicialmente controlan la replicación viral y reducen la carga viral a niveles casi indetectables. Sin embargo, la variabilidad genética del virus ha llevado al desarrollo de mutaciones que generan resistencia a los fármacos existentes (OMS, 2020), y, por consiguiente, es crucial continuar buscando nuevas alternativas terapéuticas seguras y eficaces contra el virus.

Entre estas alternativas se destacan los productos naturales ya que el debido a su diversidad química y estructural, ofrecen una amplia gama de compuestos potencialmente bioactivos. Estos compuestos, derivados de plantas, microorganismos marinos, y otros organismos, han demostrado actividades antivirales prometedoras en estudios preliminares. Además, los productos naturales han sido históricamente la inspiración para el desarrollo de nuevos fármacos y actualmente la mayoría de los medicamentos comercializados son directa o indirectamente relacionados a estos, lo que subraya la importancia de explorar estos recursos en la lucha contra el VIH y otras enfermedades virales.

Recientemente, los modelos predictivos computacionales, han permitido acelerar el desarrollo de nuevos medicamentos y mejorar la precisión en la evaluación de su efectividad. Estos métodos han emergido como herramientas fundamentales en la industria farmacéutica,

permitiendo la selección de moléculas con potencial actividad biológica de manera más eficiente que los métodos tradicionales como el *High-Throughput-Screening* (HTS).

En este contexto, la investigación de nuevas quimioterapias seguras y eficaces contra el VIH, que sean accesibles en términos de costo, se vuelve imperativa y las metodologías computacionales se han tornado fundamentales en este proceso. Ejemplos como el Raltitrexed, que actúa contra la timidilato sintasa del VIH, demuestran el potencial de las de este tipo de herramientas in silico, para el desarrollo de estos medicamentos.

Por lo tanto, este trabajo busca mediante el uso de herramientas computacionales avanzadas, como las de aprendizaje de máquina, para la selección de productos naturales (diterpenos de tipo caurano, y clerodano), como potenciales inhibidores de la proteasa del VIH-1, en busca de desarrollar nuevos medicamentos efectivos y seguros para combatir esta enfermedad.

2. FORMULACIÓN DEL PROBLEMA

¿Existen metabolitos secundarios (diterpenos) más eficaces que los compuestos activos presentes en los fármacos ya existentes para el tratamiento antirretroviral del VIH-1?

3. JUSTIFICACIÓN Y DELIMITACIÓN DE LA INVESTIGACIÓN

3.1 JUSTIFICACIÓN

De acuerdo con la Organización Mundial de la Salud (OMS), a finales de 2022, cerca de 40 millones de personas vivían en el mundo con el virus de inmunodeficiencia humana (VIH). En Colombia, se estima que 150.000 personas conviven con el virus, con 13.000 nuevos casos en 2022. (UNAIDS, 2023).

Principalmente, las poblaciones de mayor vulnerabilidad se ven afectadas, especialmente los hombres que se relacionan sexualmente con otros hombres (HSH), con una prevalencia del 20,4%, y las mujeres transgéneros, con un 23,4%. Además de la necesidad de un sólido enfoque preventivo por parte del sistema de salud para controlar la enfermedad, que incluya el uso de condones y prácticas sexuales seguras, se requiere el desarrollo de nuevos tratamientos efectivos, seguros y económicos para eliminar la enfermedad y prevenir el desarrollo del síndrome de inmunodeficiencia adquirida SIDA (UNAIDS, 2023).

Este último se define cuando el recuento de células T CD4 + cae por debajo de 200 células/μL de sangre o debido a la aparición de enfermedades específicas, en asociación con una infección por VIH. Actualmente, no existe una cura eficaz contra el VIH. Una vez que se contrae el virus, se convive con él de por vida. Sin embargo, los tratamientos antirretrovirales (TAR) permiten controlar el VIH y evitar el desarrollo del SIDA. (Cachay, 2023)

Cerca del 95% de las personas que viven con el virus deberían haber logrado suprimir la carga viral con los TAR. No obstante, otros problemas asociados, como el riesgo de coinfecciones con tuberculosis, hepatitis B y otras enfermedades de transmisión sexual, hacen necesaria una investigación continua en la búsqueda de nuevos tratamientos contra el VIH. Los métodos computacionales emergen como una importante alternativa para el desarrollo de nuevas

quimioterapias contra esta epidemia. Algunos casos de éxito, como el Amprenavir, que actúa contra la proteasa del VIH y cuyo mecanismo de acción fue elucidado mediante cálculos de dinámica molecular, fortalecen el uso de estas técnicas in silico para realizar nuevas aproximaciones en el desarrollo de modelos predictivos para la selección de nuevos inhibidores de blancos terapéuticos del virus, como las proteasas del VIH. (OMS, 2024)

3.2 DELIMITACIÓN

Esta investigación se centrará en la elaboración de un modelo predictivo para identificar inhibidores de la proteasa del VIH-1, con el objetivo de encontrar nuevos compuestos (diterpenos de tipo clerodano y/o caurano) con potencial efecto inhibidor y mayor eficacia. Para esto, se emplearán métodos computacionales avanzados y técnicas de *docking* molecular.

Este proyecto se fundamenta por la urgencia de desarrollar nuevos tratamientos contra el VIH-1 que permitan ayudar a la población (150.000 personas) expuesta y vulnerable a esta problemática en Colombia tal como lo son HSH, mujeres transgénero y personas del común pues este tipo de padecimiento contribuye a la generación de otros problemas de salud asociados tales como coinfecciones con tuberculosis, hepatitis B y otras enfermedades de transmisión sexual.

4. OBJETIVOS

4.1 OBJETIVO GENERAL

Seleccionar diterpenos de tipo clerodano y/o caurano como potenciales inhibidores contra la proteasa de VIH-1 mediante el uso de una herramienta quimioinformática basada en algoritmos de aprendizaje de máquina.

4.2 OBJETIVOS ESPECÍFICOS

- Construir un banco de moléculas de estructuras con actividad inhibitoria (CI_{50}) reportada contra la enzima Proteasa de VIH-1
- Diseñar un modelo clasificatorio predictivo para la proteasa de VIH-1
- Seleccionar potenciales metabolitos secundarios (diterpenos) como potenciales inhibidores de VIH-1

5. MARCO DE REFERENCIA

5.1 MARCO TEÓRICO

El Virus de Inmunodeficiencia Humana (VIH) pertenece a la familia de los lentivirus y se clasifica en dos tipos: VIH-1 y VIH-2 que tienen un 40-50% de homología genética y una organización genómica similar. El VIH-1 es el causante de la pandemia de SIDA mientras que el VIH-2, aunque también puede producir SIDA, se considera menos patogénico y transmisible.

Las cepas del VIH-1 se han clasificado en tres grandes grupos según su homología genética y se piensa que representan diferentes episodios de salto inter-especies. Estos son el grupo M (main o principal), el grupo O (outlier), y el grupo N (no M, no O). El grupo M se ha dividido en 9 subtipos (A, B, C, D, F, G, H, J, K) y en cepas recombinantes entre ellos, denominados CRF (formas recombinantes circulantes). Los CRF se forman por recombinación de fragmentos genómicos de distintos subtipos. Tanto el VIH-1 como el VIH-2 provienen de diferentes saltos inter-especie de virus que infectan en la naturaleza a poblaciones de simios en África.(Delgado, 2011)

El virus de la inmunodeficiencia humana tipo 1 (VIH-1) es el agente productor del SIDA, la cual es una enfermedad reconocida desde hace 30 años, que ha alcanzado proporciones pandémicas. Su origen se remonta a la transmisión a humanos de retrovirus que infectan a poblaciones de chimpancés en África central hace aproximadamente 100 años.

La infección en humanos por el VIH-1 probablemente se mantuvo inicialmente limitada a pequeños grupos de población hasta que alcanzó, seguramente a través del Río Congo, un núcleo urbano en rápida expansión como era la ciudad de Kinshasa. A partir de este punto el VIH se diseminó por el continente por contacto sexual, y muy probablemente por prácticas sanitarias con

material contaminado, hasta que se introdujo en el mundo desarrollado durante los años setenta, causando los primeros casos de SIDA detectados inicialmente en EE. UU. a principios de los ochenta. El VIH-1 grupo M es el responsable principal de la pandemia de SIDA. Dentro de este grupo, las cepas del subtipo B predominan en Europa y América y son poco frecuentes en África. (Delgado, 2011)

Desde su localización en América, Europa y África, su expansión a todo el mundo se ha dado de manera gradual y rápida. Más del 80% de los adultos infectados con VIH-1 se infectaron mediante la exposición de las superficies mucosas al virus; la mayor parte del 20% restante se infectó mediante inoculaciones percutáneas o intravenosas. (Delgado, 2011)

El riesgo de infección asociado con diferentes rutas de exposición varía, pero no importa cuál sea la ruta de transmisión, el momento de aparición de los marcadores de infección virales y del huésped es diferente y sigue un patrón ordenado. Inmediatamente después de la exposición y transmisión, a medida que el VIH-1 se replica en la mucosa, submucosa y tejidos linforreticulares de drenaje. (Esteban, s.f)

Según estimaciones de la OMS y ONUSIDA, 2.4 millones de personas viven con VIH en América Latina y el Caribe. El 81% de las personas estimadas que vivían con el virus en la región estaban diagnosticadas, el 65% recibían tratamiento y el 60% estaban con carga viral suprimida. Los mayores aumentos de la infección se produjeron en Brasil (21%), Costa Rica (21%), el Estado Plurinacional de Bolivia (22%) y Chile (34%). Al mismo tiempo, se observaron grandes descensos en El Salvador (-48%), Nicaragua (-29%) y Colombia (-22%). (UNAIDS, 2023).

En América Latina las nuevas infecciones por VIH no han descendido entre 2010 y 2020 y la reducción en el Caribe no ha ocurrido al ritmo necesario. Acelerar la introducción y escala de

nuevos métodos de prevención y tratamiento para la población en mayor riesgo es clave para retomar el rumbo y superar los retos presentados por la pandemia (UNAIDS, 2023). Aunque los casos en Colombia han descendido gradualmente, la investigación científica trabaja en la búsqueda de la erradicación definitiva del virus.

Las principales características del SIDA incluyen la destrucción de los linfocitos T CD4+ auxiliares y la consiguiente pérdida de competencia inmunológica. Desde el reconocimiento de este síndrome en 1981, se han realizado esfuerzos considerables para identificar el mecanismo por el cual el VIH-1 causa la enfermedad, y se han propuesto dos hipótesis principales. La primera hipótesis es que el VIH provoca la pérdida de linfocitos T CD4+ al infectar y matar directamente esas células. La segunda, basada en observaciones de que las células infectadas y no infectadas se ven afectadas, la infección por VIH-1 perjudica indirectamente la función celular, tal vez debido a una reacción aberrante de la respuesta inmune del huésped a la infección (Enrique, sf).

El avance más significativo en el tratamiento médico de la infección por VIH-1 ha sido el tratamiento de pacientes con fármacos antivirales, que pueden suprimir la replicación del VIH-1 hasta niveles indetectables. El descubrimiento del VIH-1 como agente causante del SIDA, junto con una comprensión cada vez mayor del ciclo de replicación del virus, han sido fundamentales en este esfuerzo al proporcionar a los investigadores el conocimiento y las herramientas necesarias para llevar adelante los esfuerzos de descubrimiento de fármacos centrados en la inhibición dirigida con fármacos específicos agentes.

Hasta la fecha, se dispone de 24 medicamentos aprobados por la Administración de Alimentos y Medicamentos (FDA) para el tratamiento de las infecciones por VIH-1. Estos medicamentos se distribuyen en seis clases distintas según su mecanismo molecular y perfiles de

resistencia: (1) inhibidores de la transcriptasa reversa análogos de nucleósidos (NNRTI), (2) inhibidores de la transcriptasa reversa no nucleósidos (NNRTI), (3) inhibidores de la integrasa, (4) inhibidores de proteasa (IP), (5) inhibidores de fusión y (6) antagonistas de correceptores. (Arts & Hazuda, 2012)

La proteasa del VIH-1 es la enzima responsable de la escisión de los precursores de las poliproteínas virales gag y gag-pol durante la maduración del virión (Arts & Hazuda, 2012). Debido a su papel vital en el ciclo de vida del VIH-1 y su tamaño relativamente pequeño (11 kDa), inicialmente se esperaba que la resistencia a los inhibidores de la proteasa fuera rara. Sin embargo, el gen de la proteasa tiene una gran plasticidad, observándose polimorfismos en 49 de los 99 codones y más de 20 sustituciones que se sabe que están asociadas con la resistencia (Morales Torrado, 2022).

La aparición de resistencia a los inhibidores de la proteasa probablemente requiere la acumulación gradual de mutaciones primarias y compensatorias (Vanegas-Otálvaro et al., 2014). Han existido otras aproximaciones en búsqueda de nuevas moléculas contra VIH-1, una de las más conocidas es el uso de modelos predictivos de *Machine Learning* los cuales permiten el diseño de fármacos basados en dianas, un caso de éxito muy conocido es Amprenavir, el cual inhibe la proteasa aspártica del virus de VIH demostrando en el 88% de los pacientes tratados, que en 16 semanas la carga viral decrece hasta ser inferior a los límites de detección (Cruz-Langarica, 2017)

Otro caso de éxito muy conocido es Raltitrexed el cual inhibe la multiplicación de las células tumorales previniendo la formación de tumores en el tracto intestinal. Se puede administrar de forma segura en pacientes con enfermedad cardiovascular, así como en pacientes con déficit de dihidropiridina deshidrogenasa («Raltitrexed quincenal versus trisemanal con oxaliplatino (con o

sin bevacizumab) en cáncer colorrectal metastásico de primera línea (Carreres-Prieto et al., 2015).

Ambos casos representan un gran avance en el uso de métodos predictivos e impulsan este tipo de alternativas en el desarrollo de fármacos dando lugar a los procesos ADMET. La calidad de los candidatos a fármacos de molécula pequeña, que abarca aspectos que incluyen su potencia, selectividad y características ADMET (absorción, distribución, metabolismo, excreción y toxicidad), es un factor clave que influye en las posibilidades de éxito en los ensayos clínicos. Es importante destacar que dichas características están bajo el control de los químicos durante la identificación y optimización de los compuestos principales. (Cumming et al.,2013)

Dadas estas cuestiones, desde hace mucho tiempo existe interés en el uso de enfoques computacionales para ayudar a guiar la selección y optimización de compuestos para síntesis y pruebas con el fin de reducir los riesgos de fallas relacionadas con sus propiedades fisicoquímicas. Estos enfoques computacionales pueden dividirse en términos generales en métodos basados en la física y métodos basados en la empírica. Los métodos empíricos se basan en patrones observados en los datos existentes, que se utilizan para guiar el diseño de compuestos futuros; ejemplos de tales métodos incluyen relaciones cuantitativas estructura-actividad (QSAR), sistemas basados en reglas y sistemas expertos. Los métodos QSAR utilizan enfoques basados en clasificación y regresión estadística para identificar patrones cuantitativos que están presentes en los datos existentes. (Cumming et al.,2013)

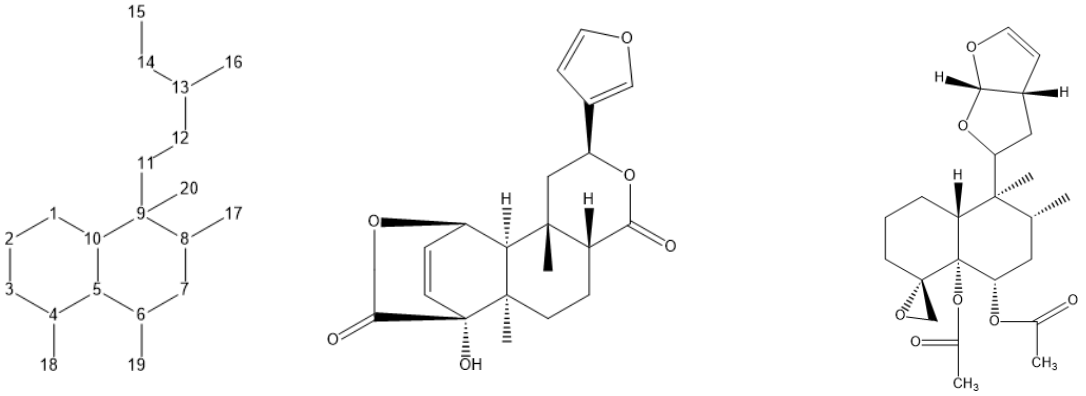
5.2 MARCO CONCEPTUAL

***5.2.1 Diterpenos Clerodanos*:** Los diterpenos de clerodano son diterpenoides bicíclicos. El esqueleto básico se divide en dos fragmentos: un resto de decalina anular fusionado. (C-1 – C-10) y una cadena lateral de seis carbonos en C-9 (C-11 – C-16, con C16 unido en C-13, es decir, 3-

metilpentilo). Los cuatro restantes los carbonos (C-17 – C-20) están unidos en C-8, C-4, C-5 y C-9. (Li et al., 2016)

Figura 1

Esqueleto de un diterpeno clerodano con dos ejemplos



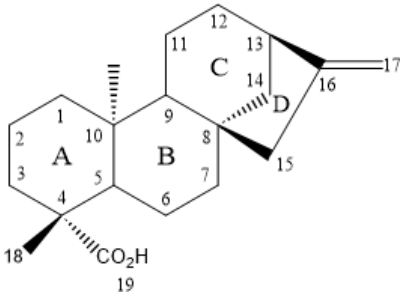
Esqueleto de un clerodano Columbina (1) Clerodina (2)

Nota: La figura muestra la estructura de dos tipos de clerodanos Columbina (1) y Clerodina (2), así como su esqueleto. Adaptado de *Clerodane diterpenes: sources, structures, and biological activities, por Li, R., Morris-Natschke, S. L., & Lee, K. H (2016). Natural Product Reports p.02, 1166-1226. https://doi.org/10.1039/c5np00137d*

***5.2.2 Diterpenos Cauranos*:** Son diterpenoides naturales aislados de varias familias de plantas, como las Asteraceae y Lamiaceae. Estos compuestos han atraído interés debido a sus estructuras y actividades biológicas, incluidas antitumorales, anti-VIH y antibacterianas. (Rosselli et al., 2007)

Figura 2.

Esqueleto de un diterpeno tipo caurano



Nota: Adaptado de *The Remarkable Structural Diversity Achieved in ent-Kaurane Diterpenes Fungal Biotransformations*, por Takahashi, J., Gomes, D., Lyra, F., Santos, G. D., & Martins, L (2014). Molecules/Molecules Online/Molecules Annual p.1860, https://doi.org/10.3390/molecules19021856

5.2.3 Proteasa de VIH: Es una enzima encargada de favorecer la síntesis de diferentes aminoácidos que a su vez se transforman en glicoproteínas que favorecen la maduración del virus, algunas de estas, se encuentran presentes en la superficie del virus, las cuales facilitan la unión a los receptores de las células diana y desencadenan la fusión de membranas para provocar la infección. Las glicoproteínas del VIH 1 (gp120), (gp41) se sintetizan inicialmente por un precursor de gp160. (Anazodo et al., 1995)

5.2.4 Aprendizaje automatizado (Machine Learning): Es una rama en evolución de los algoritmos computacionales que están diseñados para emular la inteligencia humana aprendiendo del entorno circundante. Los grados de complejidad de estos procesos pueden variar y pueden involucrar varias etapas de interacciones sofisticadas entre humanos y máquinas y toma de decisiones, lo que naturalmente invitaría al uso de algoritmos de aprendizaje automático para optimizar y automatizar

estos procesos. (Naqa & Murphy, 2015)

***5.2.5 Modelos predictivos*:** Los métodos QSAR utilizan enfoques basados en clasificación y regresión estadística para identificar patrones cuantitativos que están presentes en los datos existentes. Este tipo de modelos permiten guiar la selección de candidatos a fármacos de mayor calidad, así como los factores culturales que pueden haber afectado su uso e impacto. (Cumming et al., 2013)

5.2.6 Tratamiento antirretroviral: Un tratamiento antirretroviral para el VIH 1 tiene la finalidad de suprimir la replicación viral a niveles tan bajos que el virus es incapaz de generar mutaciones de resistencia a los medicamentos. En teoría, una vez que se alcanza este nivel de supresión viral, el tratamiento debería funcionar indefinidamente y el riesgo a largo plazo de morbilidad y mortalidad relacionados con la inmunodeficiencia asociada al VIH se vuelve insignificante. (Deeks, 2006)

***5.2.7 Metabolitos secundarios*:** Los metabolitos secundarios son compuestos que no son necesarios para que una célula viva, pero que desempeñan un papel en la interacción de la célula (organismo) con su entorno. Estos compuestos suelen estar implicados en la protección de las plantas contra el estrés biótico o abiótico. Algunos metabolitos secundarios se utilizan especialmente como sustancias químicas, como medicamentos, aromas, fragancias, insecticidas y colorantes, y por tanto tienen un gran valor económico. Estas nuevas tecnologías servirán para ampliar y mejorar la utilidad continua de las plantas superiores como fuentes renovadoras de sustancias químicas, especialmente compuestos medicinales. (Pagare et al., 2015)

5.3 MARCO HISTÓRICO

La suramina fue un inhibidor de la transcriptasa reversa (RT) descrito en 1979 por su potente efecto inhibidor sobre la RT de varios virus tumorales de ARN (no humanos). Basándose en estas observaciones, Broder y sus colegas fueron los primeros en demostrar en 1984 que la suramina protegía las células in vitro contra la infectividad del HTLV-III, y la suramina fue también el primer compuesto antirretroviral que demostró ser eficaz para reducir el virus se produce in vivo, en humanos. Pero la carrera de la suramina en el tratamiento de las infecciones por VIH terminó bastante abruptamente, las dos razones principales fueron que se descubrió que la suramina era demasiado tóxica para uso sistémico y un nuevo agente antirretroviral (De Clercq, 2009)

En 1985, se desarrolló una prueba de diagnóstico de anticuerpos y se iniciaron ensayos clínicos con inhibidores didesoxinucleótidos de la transcriptasa inversa (INTI) de acción directa, siendo el primero la azidotimidina (AZT). En 1987, se encontró que el tratamiento con AZT se asociaba con una mayor supervivencia a las 24 semanas, pero este beneficio fue de corta duración; a las 48 semanas, ya no se observaron beneficios en la supervivencia. A pesar de sus limitaciones y efectos secundarios, el AZT, más tarde llamado zidovudina (ZDV), fue aprobado en 1987 para su uso en pacientes con VIH avanzado. En breve sucesión, se aprobaron otros tres NRTI para su uso en la infección por VIH-1: zalcitabina (ddC), didanosina (ddI) y estavudina (d4T). Cada uno tenía sus propias toxicidades particulares y ninguno se usa ampliamente en la actualidad. Para evitar los perfiles de toxicidad, se intentó administrar los fármacos de forma secuencial y alternar terapias. Estos enfoques no fueron muy efectivos y, clínicamente, los pacientes continuaron teniendo malos resultados, excepto por una reducción en algunas de las tasas de reacciones adversas (Sosa, s. f.).

En los últimos años, el progreso en la terapia antirretroviral se ha caracterizado por la disponibilidad de nuevos fármacos antirretrovirales potentes y relativamente más seguros que pertenecen a clases antiguas (INTI, ITINN e inhibidores de la proteasa (Sosa, s. f.).

Los inhibidores de proteasa constituyen la clase más grande de fármacos en la lucha contra el VIH. Los inhibidores de proteasa son inhibidores selectivos y competitivos de proteasa, una enzima crucial para la maduración viral, la infección y replicación. Se preparan para brotar de la célula, provocando la lisis celular y muerte. Justo antes de que las proteínas abandonen la célula, necesitan ser escindido por una enzima proteasa llamada “proteasa del VIH”. (Wynn et al., 2004)

6. TIPO DE INVESTIGACIÓN

El tipo de investigación es experimental, pues mediante el desarrollo y ejecución de dos modelos predictivos, se obtuvieron datos empíricos tales como los compuestos (diterpenos tipo clerodano y/o caurano) con mayor acción inhibitoria frente a la enzima proteasa del VIH 1.

La investigación experimental se utiliza generalmente en ciencias tales como la sociología y la psicología, la física, la química, la biología y la medicina, entre otras. Se trata de una colección de diseños de investigación que utilizan la manipulación y las pruebas controladas para entender los procesos causales. En general, una o más variables son manipuladas para determinar su efecto sobre una variable dependiente. (*Investigación Experimental*, s. f.)

Este tipo de investigación está integrada por un conjunto de actividades metódicas y técnicas que se realizan para recabar la información y datos necesarios sobre el tema a investigar y el problema a resolver. (Ruiz, s. f.)

.

7. DISEÑO METODOLÓGICO

7.1 BANCOS DE MOLÉCULAS*:*

Se utilizaron 20 artículos como punto de partida, para seleccionar estructuras que hayan reportado concentración inhibitoria 50% (CI_{50}) contra la proteasa de VIH-1. El objetivo fue construir una base de estructuras bidimensionales de aproximadamente 2850 moléculas. Una vez se seleccionaron las moléculas que cumplen con los parámetros de búsqueda, se diseñaron en ChemDraw (Perkin Elmer) y se guardaron en formato smiles y mol.

7.2 CREACIÓN DEL MODELO CLASIFICATORIO:

En la construcción del modelo clasificatorio se utilizó el banco de moléculas construido en la etapa anterior (estructuras con valores de CI_{50} reportados contra la proteasa de VIH-1. Los compuestos seleccionados se organizaron y clasificaron como activos e inactivos de acuerdo con su valor de pCI_{50} siendo el punto de corte $pCI_{50} \geq 7{,}0$. Todas las estructuras se guardaron en formato. smiles

Se realizó un proceso de limpieza de los datos (data curation), con el fin de remover estructuras duplicadas, evitar efectos de frontera y eliminar moléculas con información incompleta, todos los valores de CI_{50} (Los valores CI_{50}, describen la concentración de una sustancia dada requerida para inhibir el 50% del crecimiento del parásito) se estandarizaron usando unidades de molaridad; para las estructuras duplicadas, aquellas que tengan los mayores valores de pCI_{50} (-Log CI_{50}) serán eliminadas con el fin de generar modelos más restrictivos.

7.3 CÁLCULOS DE LOS DESCRIPTORES MOLECULARES:

Para el banco de moléculas de metabolitos secundarios, se diseñaron las estructuras tridimensionales, las cuales se usaron como dato de ingreso en formato .SDF al software Dragon 5.0, con el fin de obtener información química de las moléculas siendo que los descriptores moleculares incluidos en el software permitieran cubrir la mayoría de las aproximaciones teóricas («Handbook Of Molecular Descriptors. Methods And Principles In Medicinal Chemistry, Volume 11 Edited By Roberto Todeschini, Viviana Consonni (A Series Edited By R. Mannhold, H. Kubinyi, H. Timmerman), Wiley-VCH, Weinheim, 2000. 667 Pp.; E160.00», 2001)

7.4 CONSTRUCCIÓN DEL MODELO DE PREDICCIÓN CLASIFICATORIO:

Para construir el modelo de aprendizaje de máquina, se utilizó el programa KNIME 5.2.5 (KNIME 5.2.5, the Konstanz Information Miner Copyright, www.knime.org) (Lechner et al., 2024). Inicialmente se transformó el formato .smiles a formato 2D empleando la función *"Molecule Type Cast*" para posteriormente, generar descriptores tipo Fingerprint en Morgan con la función "*RDKit Fingerprint*" y utilizando la función ''*partitioning*'' se realizó un muestreo estratificado dividiendo las estructuras del banco de moléculas construido para la proteasa de VIH-1 en dos grandes grupos, un 80% conformó el grupo de entrenamiento y el restante 20% el grupo test. El modelo será generado utilizando un algoritmo *Random Forest* incluido en el software y utilizando los nodos de WEKA.

Se realizó un procedimiento de validación cruzada en este proceso, los datos fueron divididos (80/20) nuevamente cinco veces permitiendo la validación del modelo. Los parámetros

seleccionados para el algoritmo *Random Forest*, fueron definidos de acuerdo con el rendimiento alcanzado por el modelo. Una vez construido el modelo clasificatorio se determinaron los parámetros de calidad relacionados con la matriz de confusión, la curva ROC relacionando valores de verdaderos positivos con valores de falsos positivos y el coeficiente de correlación de Matthews (MCC), el cual relaciona los cuatro parámetros que conforman la matriz de confusión.

7.5 *DOCKING* MOLECULAR:

La estructura cristalina de la proteína proteasa wild type de VIH-1 fue extraída del Protein Data Bank (PDB), PDB ID: 2P3B, la proteína estaba en complejo con el ligando TL-3 (PDB ID: 3TL). La estructura tridimensional en formato .PDB, se usó como archivo de entrada en el software Molegro 6.0.1, para realizar los cálculos de *docking* molecular, utilizando TL-3 como inhibidor competitivo de la enzima modelada. En Molegro 6.0.1, inicialmente fueron eliminadas todas las aguas y cofactores asociados a la estructura, la enzima fue preparada usando la configuración predeterminada del mismo software (función de puntuación: MolDock score; evaluación del ligando: internal ES, internal H-Bond, sp2-sp2 torsion; número de corridas: 10; algoritmo:MolDock SE) los cálculos de *docking* fueron realizados usando un Grid con radio de 15 Å y una resolución de 0.30 Å para cubrir el sitio de unión al ligando de la enzima.

7.6 ANÁLISIS DE CONSENSO:

Los resultados obtenidos en el modelo clasificatorio fueron analizados en conjunto con los resultados provenientes de los cálculos de *docking* molecular mediante un análisis de consenso el cual relaciona los valores de probabilidad de cada aproximación, de acuerdo con las ecuaciones propuestas por Herrera-Acevedo y Scotti (Herrera-Acevedo et al., 2021). Los resultados de probabilidad obtenidos del modelo clasificatorio tuvieron mayor relevancia puesto que estos son

basados en valores experimentales (resultados In vitro). Las moléculas que fueron clasificadas como activas por las dos aproximaciones fueron las seleccionadas como potenciales inhibidores de la proteasa de VIH-1.

8. RESULTADOS Y DISCUSIÓN

Inicialmente, se construyó una base de datos de 2840 estructuras bidimensionales con moléculas que presentan valores de actividad inhibitoria de ensayos in vitro contra la enzima proteasa 2P3B del VIH-1 (concentración inhibitoria 50% - CI_{50}) reportadas en artículos científicos. Para clasificarlos como activos e inactivos se tomó un punto de corte de pCI_{50}=7,0. Siendo crucial para la construcción del modelo pues nos permite clasificar de manera clara y precisa los compuestos como activos o inactivos basándonos en un criterio objetivo.

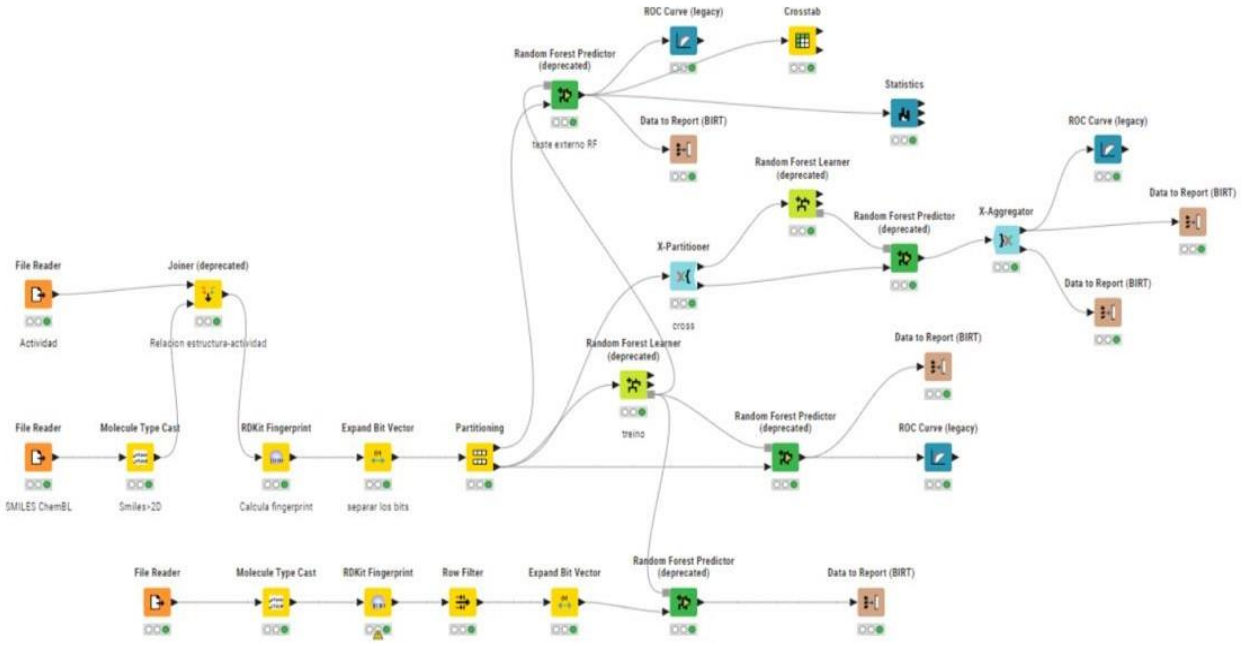
Posteriormente, se realizó la limpieza de los datos con el fin de remover estructuras duplicadas, evitar efectos de frontera y eliminar moléculas con información incompleta. Mientras que, a su vez, los normaliza con unidades de molaridad respectivas para los valores de pCI_{50}.

8.1 DESCRIPCIÓN DEL MODELO DE *MACHINE LEARNING*

Se construyó el modelo de *Machine Learning* usando el programa KNIME 5.2.5 basado en cuatro etapas: preparación de datos, desarrollo y evaluación del modelo, prueba y comparación de predicciones, y finalmente generación de resultados (Figura 3). Mediante la implementación de 30 nodos totales, se construyeron árboles de decisión. El programa evaluó el conjunto de modelado que representa el 80% y el conjunto de validación externa que representa el 20% restante del total de las moléculas procesadas.

Figura 3.

Modelo predictivo diseñado en el programa KNIME 5.2.5



Para la **preparación de datos** respecto al **conjunto de modelado**(80%), inicialmente, se usó el nodo ubicado en la parte superior "*File Reader*" que lee las moléculas activas e inactivas, mientras que el nodo "*File Reader*" ubicado en la parte inferior, permitió importar los smiles generados por medio de ChEMBL para posteriormente usar el nodo "*Molecule Type Cast*" que permite transformar el formato smiles a formato 2D, para a su vez, ser unidos al nodo "*Joiner*" el cual permite la combinación de los smiles en formato 2D junto con las moléculas activas e inactivas provenientes de la base de datos. Adicionalmente se emplea el nodo "*Fingerprint*" que permite generar huellas moleculares (vectores) en Morgan de las moléculas para posteriormente usar el nodo "*Expand Bit Vector*" que toma las huellas digitales generadas en Morgan y genera una representación más detallada de las moléculas para su posterior análisis mediante la expansión de bits con el fin de emplear el nodo "*Partitioning*" que facilita la división de los datos expandidos

para su posterior análisis.

La validación del modelo es la parte más importante de la construcción de un modelo supervisado. Para encontrar un conjunto óptimo de parámetros del modelo, es necesario dividir los datos en conjuntos de entrenamiento y validación (Xu & Goodacre, 2018). Con el fin de efectuar la validación del modelo, se emplearon las metodologías de entrenamiento, validación cruzada y prueba externa respectivamente. Para cada una de ellas, se desarrollaron las etapas de desarrollo y evaluación del modelo, prueba y comparación de predicciones y generación de resultados las cuales son explicadas a continuación para cada caso.

En lo referente al entrenamiento, se utiliza para construir el modelo con múltiples configuraciones de parámetros (Xu & Goodacre, 2018). Para la fase de **desarrollo y evaluación de la metodología**, inicialmente se empleó un nodo "*Random Forest* Learner" que define los parámetros favoreciendo la generalización del modelo como se ha mencionado anteriormente, mientras que el nodo "*Random Forest* Predictor" ejecutó la predicción del modelo parametrizado teniendo en cuenta los datos provenientes del entrenamiento del anterior nodo y de la fase de preparación de datos provenientes del nodo "*Partitioning*", para así mismo ejecutar la fase de **prueba y comparación de predicciones** mediante la implementación del nodo "*ROC Curve*" con el cual podemos evaluar el rendimiento del modelo ejecutado para finalmente pasar a la **generación de resultados** relacionados con los datos evaluados mediante el nodo "*Data to report*".

Por otra parte, la validación cruzada es un esquema con el que se trata de obtener una estimación de la actuación de las máquinas. Para ello se comprueba que distintas máquinas entrenadas con un conjunto de datos distintos producen resultados consistentes (Aprendizaje

Automático y Técnicas de Validación, 2020). Para modelar este método de evaluación, en la etapa de **desarrollo y evaluación de la metodología** inicialmente, se empleó el nodo "*X Partitioner*" el cual permitió dividir los datos en subconjuntos (folds) para evaluar la validación cruzada, posteriormente, se empleó el nodo "*Random Forest Learner*" que permite construir múltiples árboles de decisión y así mismo favorece la generalización del modelo configurando los parámetros de este. A su vez, se empleó el nodo "*Random Forest* Predictor" el cual evalúa el modelo en los subconjuntos desarrollados previamente. Adicionalmente, se empleó el nodo "*X Aggregator*" el cual se encargó de combinar todos los datos obtenidos después de evaluar el modelo en cada subconjunto. Seguido de esto, pasamos a la fase de **prueba y comparación de predicciones** durante la cual se empleó el nodo "*ROC Curve*" que nos permite evidenciar de forma gráfica el desempeño del modelo evaluado para así mismo pasar a la fase de **generación de resultados** relacionados con el ítem anteriormente mencionado mediante el nodo "*Data to Report*" con el cual podemos evidenciar datos referentes a la curva de ROC y otros parámetros asociados a la misma. Mientras que para el nodo de "*Data to Report*" conectado a la parte inferior del nodo "*X Aggregator*" podemos generar datos relacionados con los subconjuntos generados en la validación cruzada.

Así mismo la prueba externa, suele proporcionar una evaluación más realista del rendimiento del modelo en comparación con los conjuntos de validación internos, que tienden a ser optimistas (Xu & Goodacre, 2018). En la fase de **desarrollo y evaluación de la metodología**, se implementó el nodo "*Random Forest* Learning" el cual permite entrenar el modelo a predecir mediante el establecimiento de los parámetros a tener en cuenta mientras que el nodo "*Random Forest* Predictor" permite hacer las predicciones necesarias y así mismo evaluar el modelo construido mediante la implementación de otros nodos. Seguidamente, **durante la prueba y**

comparación de predicciones se emplean los nodos como el de "*ROC Curve*" el cual nos permite evaluar el área bajo la curva del gráfico que relaciona verdaderos positivos y falsos positivos, "*Crosstab*" el cual nos permite desarrollar la matriz de confusión que a su vez determina el desempeño del modelo y "*Statistics*" el cual nos permite determinar las métricas de clasificación (sensibilidad, exactitud, especificidad, F1). Finalmente, para la **generación de resultados** se empleó el nodo "*Data to Report*" que permite generar los resultados obtenidos del modelo.

En cuanto a la **preparación de datos** del **conjunto de validación externo** (20%) inicialmente se emplea un nodo "*File Reader*" que nos permite importar o leer las moléculas provenientes del banco, adicionalmente se emplea el nodo "*Molecule Type Cast*" el cual permite transformar el formato smiles a formato 2D para posteriormente añadir el nodo "*RDKit Fingerprint*" que nos permite generar las huellas moleculares con RDKit en huellas circulares (Morgan), las cuales representan las moléculas en términos de los subgrupos atómicos presentes. Para seguidamente, emplear el nodo "*Row Filter*" que nos permite el filtrado de los fingerprint generados con haciendo una limpieza de los mismos en caso de la existencia de errores, para seguidamente poder emplear el nodo "*Expand Bit Vector*" que toma las huellas generadas en Morgan y expande sus bits para pasar a la fase de **desarrollo y evaluación del modelo** en la que se empleó el nodo "Random Forest Learner" que hace un entrenamiento de los datos para posteriormente emplear el nodo "*Random Forest* Predictor" el cual ejecuta la predicción del modelo planteado bajo los parámetros establecidos en el entrenamiento así como los datos provenientes de la preparación de datos para finalmente pasar a la fase de **generación de informes** en la cual mediante el nodo **"***Data to report***"** se generan los informes correspondientes al modelo evaluado.

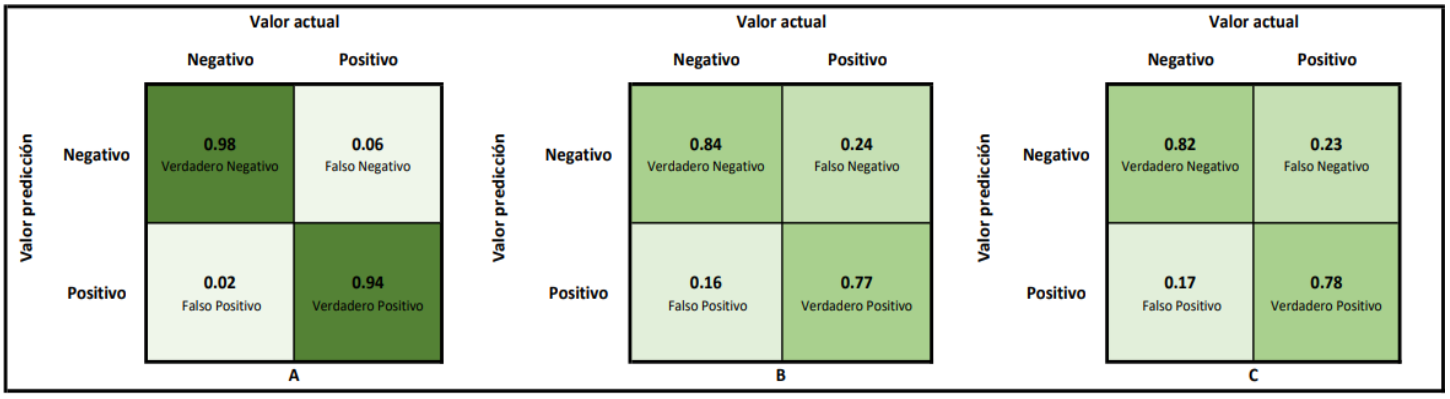
8.2 RESULTADOS DEL MODELO DE *MACHINE LEARNING*

Para la evaluación de los resultados obtenidos del modelo predictivo se emplearon las métricas de evaluación derivadas de la curva ROC tales como: Coeficiente de Correlación de Matthews (MCC), sensibilidad (VPR), exactitud (ACC), especificidad (SPC), valor predictivo positivo (PPV), valor predictivo negativo (NPV) y F1 respectivamente (Figura 4 y 5). Para evaluar dichas métricas, se construyó una matriz de confusión a cada metodología de evaluación del modelo (entrenamiento, validación cruzada y prueba externa).

En cuanto a la metodología de evaluación de validación cruzada y de entrenamiento, el número de moléculas evaluadas para cada una fue de 604, para las cuales, en ambos casos, 308 moléculas fueron clasificadas como activas mientras que 296 moléculas fueron clasificadas como inactivas, entre tanto para la metodología de prueba externa, el número de moléculas evaluadas fue de 150 de las cuales 76 fueron clasificadas como activas y 74 como no activas.

Figura 4.

Matrices de confusión



Nota: Matrices de confusión construidas con los resultados del modelo de predicción para cada metodología de evaluación a) Entrenamiento, b) Validación Cruzada, c) Prueba externa.

Figura 5.

Resultados de parámetros de métricas de evaluación

Parámetro	Valor
MCC	0.92
Sensibilidad o Razón de Verdaderos Positivos (VPR)	0.94
Exactitud (ACC)	0.96
Especificidad (SPC) o Razón de Verdaderos Negativos	0.98
Valor Predictivo Positivo (PPV)	0.98
Valor Predictivo Negativo (NPV)	0.94
F1	0.96

A

Parámetro	Valor
MCC	0.61
Sensibilidad o Razón de Verdaderos Positivos (VPR)	0.76
Exactitud (ACC)	0.80
Especificidad (SPC) o Razón de Verdaderos Negativos	0.84
Valor Predictivo Positivo (PPV)	0.83
Valor Predictivo Negativo (NPV)	0.78
F1	0.80

B

Parámetro	Valor
MCC	0.60
Sensibilidad o Razón de Verdaderos Positivos (VPR)	0.77
Exactitud (ACC)	0.80
Especificidad (SPC) o Razón de Verdaderos Negativos	0.83
Valor Predictivo Positivo (PPV)	0.82
Valor Predictivo Negativo (NPV)	0.78
F1	0.79

C

Nota: Resultados de las métricas de evaluación determinadas gracias a las matrices de confusión construidas para cada metodología de evaluación a) Entrenamiento, b) Validación Cruzada, c) Prueba Externa.

En cuanto a los resultados obtenidos de manera general para cada uno de los parámetros evaluados, podemos identificar que todos presentan valores superiores al 0,60 lo cual indica un buen desempeño del modelo.

Respecto al Coeficiente de Correlación de Matthews (MCC), podemos encontrar un número con valores entre -1 y 1. Cuando los valores de la correlación son mayores a cero y menores a uno implican una correlación positiva, significa que, si un suceso ocurre, probablemente el otro también lo hará. (Zambrano & Del Rosario de la Torre Cruz, s. f.) De modo que, para los valores obtenidos en el caso de la prueba externa y la validación cruzada siendo de 0,60 y 0,61 respectivamente, podemos encontrar una alta relación en los valores para verdaderos positivos y verdaderos negativos pues ambos casos presentan valores superiores a 0,77 sin embargo, así

mismo podemos evidenciar que los valores para falso negativo y falso positivo en ambos casos presentan un valor superior a 0,17 lo cual impacta de forma significativa el valor del MCC pues este valor relaciona todos los parámetros obtenidos en la matriz de confusión. En cuanto a la metodología de entrenamiento, podemos evidenciar resultados completamente favorables para el MCC pues se obtuvo un valor de 0,92 el cual es un número muy cercano a uno e indica una alta correlación en los verdaderos positivos y verdaderos negativos obteniendo valores de 0,94 y 0,98 para cada uno respectivamente, y valores muy bajos para los valores de falsos positivos y falsos negativos siendo de 0,02 y 0,06 respectivamente, lo cual nos indica una adecuada detección y clasificación de los datos que son identificados como activos e inactivos por el modelo de predicción.

Con relación al parámetro de sensibilidad (VPR), para todas las metodologías obtenemos valores superiores al 0,76 lo cual indica que el modelo logra identificar el 76% de verdaderos positivos en el caso de la validación cruzada y el 88% para la prueba externa, siendo la metodología de entrenamiento la que presenta el mejor resultado con un valor del 94%. En un modelo perfecto la sensibilidad es igual a 1 para cada clase. Desde el punto de vista analítico un investigador busca aumentar la sensibilidad sin afectar el valor de la exactitud (Drzewiecki, 2017). Y tal como podemos evidenciar, los valores de exactitud (ACC) obtenidos son iguales o superiores al 80% lo cual nos indica que el modelo está clasificando adecuadamente las moléculas como activas e inactivas siendo la metodología de entrenamiento la que presenta el mejor resultado con un valor del 96%. Ya que los resultados para ambos parámetros son favorables en cuanto a las metodologías evaluadas, es correcto afirmar que el modelo clasifica de manera adecuada las predicciones positivas y negativas lo que indica un correcto desempeño del mismo.

El parámetro de especificidad (SPC), puede ser relacionado con la sensibilidad mediante la siguiente definición: la sensibilidad mide la proporción de módulos defectuosos reales que se identifican correctamente y la especificidad mide la proporción de módulos no defectuosos que se identifican correctamente. (Catal, C. 2012). Para los valores que obtuvimos en cada uno de los casos, podemos evidenciar que, en todas las tablas, el valor de especificidad es superior al valor de sensibilidad, para lo cual presenta valores superiores al 83%. Esto nos indica que el modelo tiene una capacidad muy alta para identificar los verdaderos negativos lo cual, a su vez, favorece la reducción de falsos negativos.

Los parámetros, valor predictivo positivo (PPV) y valor predictivo negativo (NPV), se pueden definir como parámetros que permiten relacionar la proporción de verdaderos positivos en cuanto al total de positivos y la proporción de verdaderos negativos en cuanto al total de negativos evaluados (Catal, C. 2012). Para todos los casos, el valor de PPV fue mayor al valor de NPV, siendo en superior o igual a 82% lo cual es realmente bueno pues indica una correcta predicción de verdaderos positivos en el modelo. Así mismo, en cuanto al valor NPV obtenemos resultados positivos iguales o superiores a 78%, demostrando que el modelo predice correctamente los verdaderos negativos y que para las predicciones de ambos casos el modelo presenta valores bastante estables demostrando su confiabilidad.

Finalmente, en cuanto al valor de F1, que sirve como métrica única que resume el rendimiento de un clasificador en términos tanto de precisión como de recuperación, garantizando que no se ignora ninguna a expensas de la otra (*Métricas de Evaluación de Modelos En el Aprendizaje Automático,* 2023). Para el caso de los modelos evaluados, se obtuvieron valores iguales o superiores al 79% lo cual indica que en todos los casos existe una alta sensibilidad y precisión en el modelo de predicción, para lo cual el resultado más favorable fue el de la

metodología de entrenamiento con un 96%.

De manera global, podemos ver, que de acuerdo a los resultados obtenidos para cada uno de los parámetros evaluados, el modelo de predicción es un modelo que presenta un excelente rendimiento y desempeño pues identifica y clasifica oportunamente compuestos activos e inactivos presentando valores favorables para la exactitud, sensibilidad y especificidad las cuales son métricas de evaluación muy importantes para la definición del correcto funcionamiento del mismo, por lo cual los resultados de predicción obtenidos son confiables.

8.3 DESCRIPCIÓN DEL *DOCKING* MOLECULAR

El *docking* molecular inicia con la identificación del objetivo, que está respaldada por metodologías de genética, biología molecular y bioinformática. A continuación, se lleva a cabo la extracción y purificación de proteínas. Se realiza una determinación estructural del objetivo, principalmente mediante RMN, cristalografía de rayos X y Cry-EM; para aquellas proteínas cuya estructura cristalina no está definida, se construyen modelos de homología mediante software. (Herrera-Acevedo et al., 2022). El *docking* molecular desarrollado, consta de las siguientes etapas: preparación de la estructura de la molécula Diana (enzima proteasa), generación de conformaciones, evaluación de las interacciones y selección de las mejores cinco conformaciones.

Inicialmente, se identificó el objetivo, siendo en este caso la estructura cristalina de la proteína proteasa wild type de VIH-1 la cual fue extraída del Protein Data Bank (PDB), PDB ID: 2P3B que estaba en complejo con el ligando TL-3 (PDB ID: 3TL). Seguido de esto, se realizó la preparación de la proteína en el software Molegro Virtual Docker (MVD), para esto, se determinó el sitio de unión en las coordenadas X: 16.87, Y: -0.05, Z: 0.74. Se eliminaron todas las aguas y cofactores asociados a la estructura. Sucesivamente, se procedió a la generación de

conformaciones, para esto, se realizó el re*docking* del ligando presente en la proteína con el objetivo de encontrar la conformación que mejor se ajuste a la pose original del ligando. Y así mismo, la pose que cuenta con un menor nivel de energía para verificar la precisión del algoritmo. Posteriormente, se procedió a realizar el *docking* de las 482 moléculas procedentes de la base de datos con el fin de determinar los compuestos que presentan mayor afinidad en el sitio activo de la proteína, cada una de ellas es evaluada cinco veces. Una vez realizado dicho proceso, se evaluaron las energías de interacción entre el ligando y el sitio activo de la proteasa de VIH-1. Un nivel de energía bajo indica una alta interacción ligando-proteína en el sitio activo. Con la energía total proveniente del *docking*, se calculó la probabilidad para cada molécula de ser activa (Ps) relacionando el menor valor de energía total obtenido con cada uno de los valores de energía resultantes del *docking* tal como se puede evidenciar en la ecuación 1.

Ecuación 1. Probabilidad del *docking* molecular

$$Ps = \frac{E_M}{E_X}$$

Donde E_X es el menor valor de energía obtenido después del *docking* y E_M es el valor de energía de cada molécula evaluada en el *docking*.

Para posteriormente calcular la probabilidad combinada (Pc) con la probabilidad del modelo de predicción de Knime (Pk) y así mismo, identificar aquellas moléculas activas en los dos modelos predictivos.

Ecuación 2. Probabilidad combinada

$$\mathrm{Pc} = \frac{2 * (\mathrm{Pk}) + \mathrm{Ps}}{3}$$

Finalmente, se procedió a seleccionar las mejores cinco conformaciones con base en la probabilidad combinada obtenida, siendo estas las moléculas **109**, **234**, **235**, **230** y **231** respectivamente.

8.4 RESULTADOS *DOCKING* MOLECULAR

En cuanto a los resultados del *docking* realizado, se identificó que, de las 482 moléculas evaluadas, solo 131 resultaron siendo activas pues presentaron valores de probabilidad iguales o superiores al 0,50. Respecto al número de moléculas activas en los modelos de predicción de Knime y *docking*, únicamente 111 lograron cumplir con esta condición, para lo cual fueron seleccionadas las mejores cinco moléculas, dicha información es mostrada en la Tabla 1 para las cuales los valores de probabilidad combinada oscilaron entre 0,70 y 0,81.

Tabla 1.

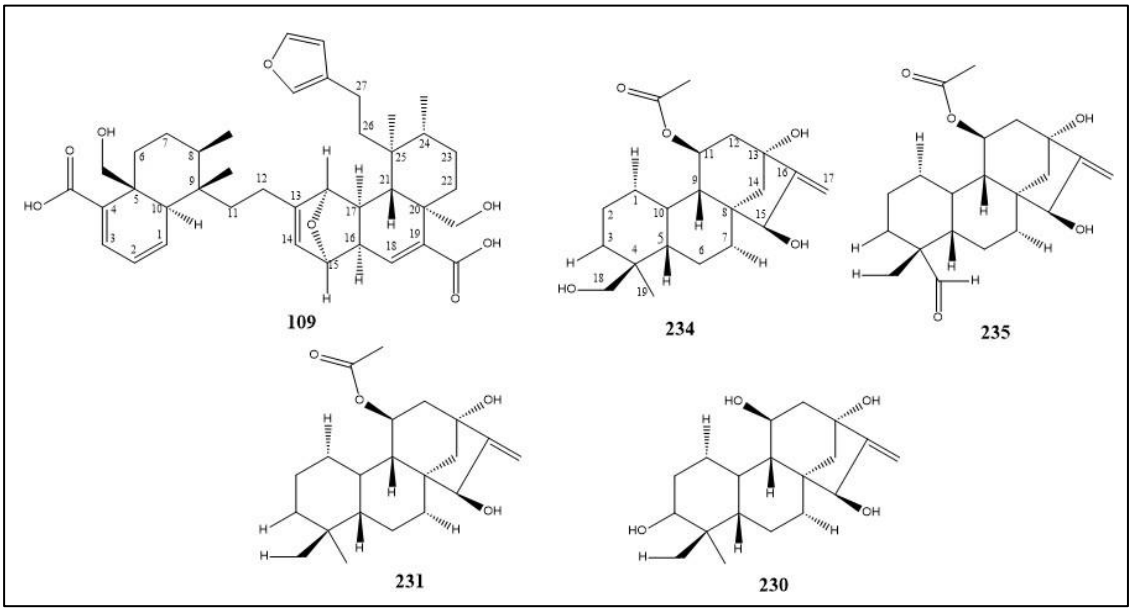
Resultados de probabilidad combinada

Clasificación	Molecula	Probabilidad docking	Probabilidad modelo predicción	Probabilidad combinada
1	**109**	1.00	0.71	0.81
2	**234**	0.59	0.77	0.71
3	**235**	0.61	0.75	0.70
4	**230**	0.54	0.78	0.70
5	**231**	0.54	0.78	0.70

Nota: Resultados de probabilidad combinada para los mejores 5 ligandos

Figura 6.

Moléculas mejor clasificadas mediante el análisis de consenso



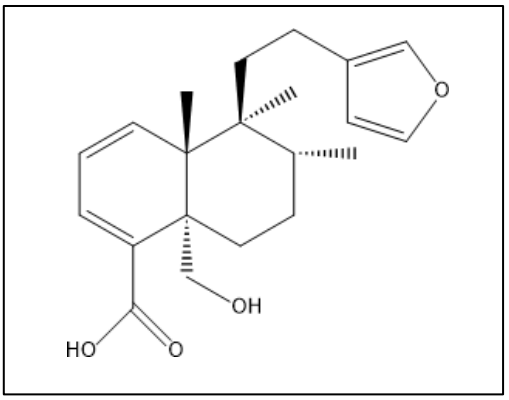
Nota: Estructura química de los ligandos con mejor interacción ligando-proteína del algoritmo de *docking* evaluado. Adaptado de *Synergism between Terbinafine and a Neo-clerodane Dimer or a Monomer Isolated from Baccharis flabellata against Trichophyton rubrum* por Rodriguez, M. V., Butassi, E., Funes, M., & Zacchino, S. A. (2019). Natural Product Communications, p.01, vol-14 https://doi.org/10.1177/1934578x1901400101 *y Structurally Diverse Diterpenoids from Isodon scoparius and Their Bioactivity* por Jiang, H., Wang, W., Tang, J., Liu, M., Li, X., Hu, K., Du, X., Li, X., Zhang, H., Pu, J., & Sun, H. (2017), Journal Of Natural Products, p.2027. https://doi.org/10.1021/acs.jnatprod.7b00163

En lo relacionado a las moléculas seleccionadas, destaca que la molécula **109** (Figura 6) es la que presentó la mayor afinidad siendo un diterpeno clerodano que cuenta con grupos funcionales

tales como ácidos carboxílicos, epóxidos, éteres y alcoholes. El compuesto **109** se formó mediante la fotocicloadición de dos moléculas de 5,16-epoxi-19-hidroxi-1,3,13(16),14-ácido clerodatetraen-18-oico (DAC) (Figura 7). Los dímeros diterpenoides son una subclase bastante poco común de diterpenoides que se componen de dos unidades diterpenoides de 20 carbonos unidas a través de uno o dos enlaces C-C, un enlace éster o un resto de anillo y se sintetizan naturalmente principalmente mediante una cicloadición de Diels Alder catalizada por enzimas. Este tipo de moléculas son conocidas por su acción citotóxica, antibacteriana y antiinflamatoria (Lin L-G,2016). Los ácidos carboxílicos presentes en los carbonos C-29 y C-31 pueden protonar el medio (pH fisiológico) lo cual provoca la desestabilización de la pared celular y la desnaturalización de proteínas (Jarboe et al., 2013). Respecto al epóxido del C-15, tiene funciones muy importantes, pues tiene la capacidad de reaccionar con los nucleófilos presentes en las proteínas llevando a la muerte celular. (Salomatina et al., 2022) En cuanto al éter del ciclopentadieno conectado al C-27, influye en la permeabilidad de la membrana favoreciendo la interacción de la molécula. Por otro lado, los grupos hidroxilo de C-30 y C-18 pueden formar enlaces de hidrógeno con las proteínas, lo cual afecta su funcionalidad e induce a la muerte celular (Derewenda, 2023).

Figura 7.

Molécula de DAC



Nota: Adaptado de Synergism between Terbinafine and a Neo-clerodane Dimer or a Monomer Isolated from Baccharis flabellata against Trichophyton rubrum por Rodriguez, M. V., Butassi, E., Funes, M., & Zacchino, S. A. (2019).. Natural Product Communications, p.02, vol-14, https://doi.org/10.1177/1934578x1901400101

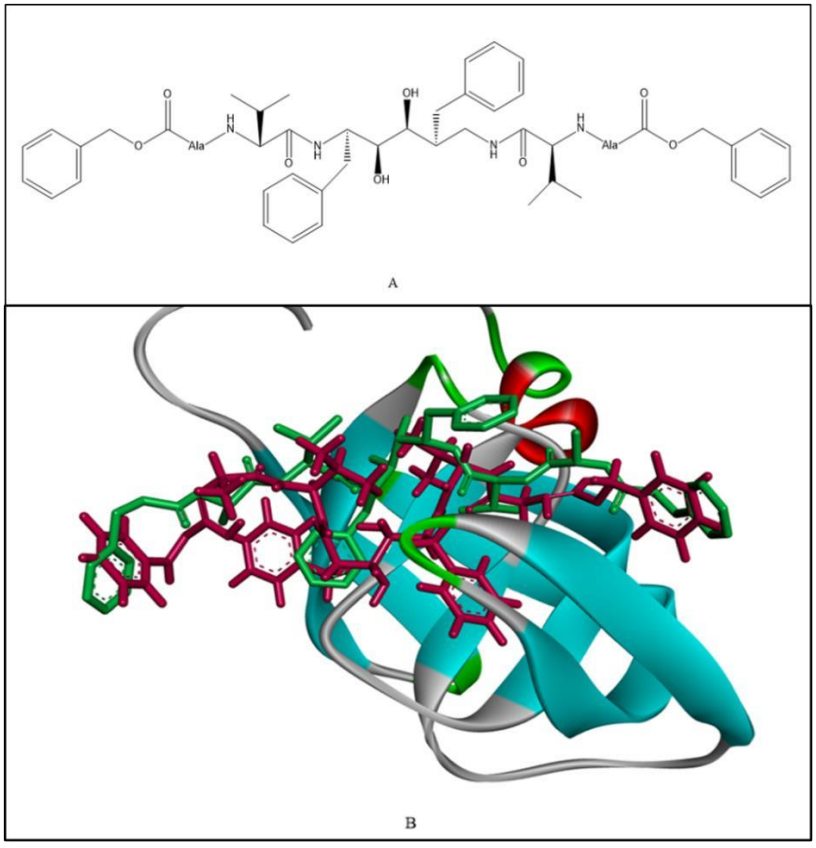
Respecto a los otros cuatro compuestos, se identificó que todos están clasificados como cauranos diterpeno. Las estructuras **234**, **235** y **230** poseen grupos acetoxi unidos a C-11. De igual forma, las estructuras **234**, **230** y **231** contienen tres grupos hidroxilo unidos a C-8, C-13 y C-18, mientras que la estructura **235** contiene dos grupos hidroxilo unidos a C-8 y C-13. En cuanto al grupo aldehído, la estructura **235** es la única que lo posee en C-19. El compuesto **235** es un análogo estructural de **234** según el autor de su artículo original, excepto por la presencia de un grupo formilo en el compuesto **235** en lugar del metileno oxigenado en el compuesto **234**, lo que provocó la desprotección de un grupo metilo (Jiang et al., 2017). En cuanto a los compuestos **230** y **231**, se identificó que estos compuestos eran análogos estructurales según el autor de su artículo original y que la única diferencia fue la presencia de un grupo hidroxilo en el compuesto **231**, que

reemplazó al grupo acetiloxi en el compuesto **230** (Jiang et al., 2017). Los grupos aldehído reaccionan con los nucleófilos de las proteínas induciendo la muerte celular, por otro lado (Singh et al., 2013), los grupos acetiloxi pueden ser hidrolizados por enzimas presentes en cierto tipo de células microbianas, lo cual favorece la liberación del principio activo del compuesto, de igual forma aumentan la estabilidad del compuesto (Arnold, 1969). En cuanto a los grupos hidroxilo, como se mencionó anteriormente, puede formar puentes de hidrógeno con las proteínas presentes en la célula.

Los sitios de unión de proteínas son módulos que interactúan con otras macromoléculas, y pequeños ligandos. Estas interacciones son responsables de la formación de complejos de proteínas, así como de la regulación de diversas vías biológicas (MONTES GRAJALES, s. f.). La naturaleza de los sitios de unión se caracteriza por ser generalmente de tipo hidrofóbico, con un número mayor de residuos con grupos de esta naturaleza expuestos en su superficie y con grandes pero variables extensiones de área superficial apolar (MONTES GRAJALES, s. f.).

Figura 8.

Redocking molecular



Nota: A) Estructura química del ligando de referencia TL-3, B) Resultado de re*docking* para el ligando TL-3, la posición original está resaltada en verde y la posición de re*docking* está resaltada en rosa. La figura A fue adaptada de *Structural Characterization of B and non-B Subtypes of HIV-Protease: Insights into the Natural Susceptibility to Drug Resistance Development* por Sanches, M., Krauchenco, S., Martins, N. H., Gustchina, A., Wlodawer, A., & Polikarpov, I. (2007), Journal Of Molecular Biology/Journal Of Molecular Biology, p.1032, https://doi.org/10.1016/j.jmb.2007.03.049

La pose de re*docking* presentó una diferencia significativa comparada con la pose original del ligando (Figura 7). Esto se debe a que, por su tamaño y flexibilidad, el ligando adoptó varias

posiciones en niveles bajos de energía lo que indicaría dificultades para que se ubique en una única pose. Por lo general, el ligando estabiliza un subconjunto de varias conformaciones posibles del receptor, desplazando el equilibrio hacia las estructuras de mínima energía (Ferreira et al., 2015). Si bien la flexibilidad de los ligandos se ha incorporado en muchos esquemas de acoplamiento, la mayoría de los programas todavía tratan a los receptores como objetos rígidos. En general, los ligandos pueden unirse a conformaciones del receptor que ocurren con poca frecuencia en el receptor sin ligando; por lo tanto, esta suposición de cuerpo rígido no logrará encontrar modos correctos de unión ligando-receptor (Lin et al., 2002).

Tabla 2.

Niveles de energía total

Clasificación	Molecula	Energía (kJ/mol)
1	**109**	-128.49
2	**234**	-75.88
3	**235**	-77.75
4	**230**	-69.91
5	**231**	-68.85
-	**Ligando TL-3**	-297.18

Nota: Niveles de energía total provenientes del *docking* para los mejores 5 ligandos y el ligando original TL-3.

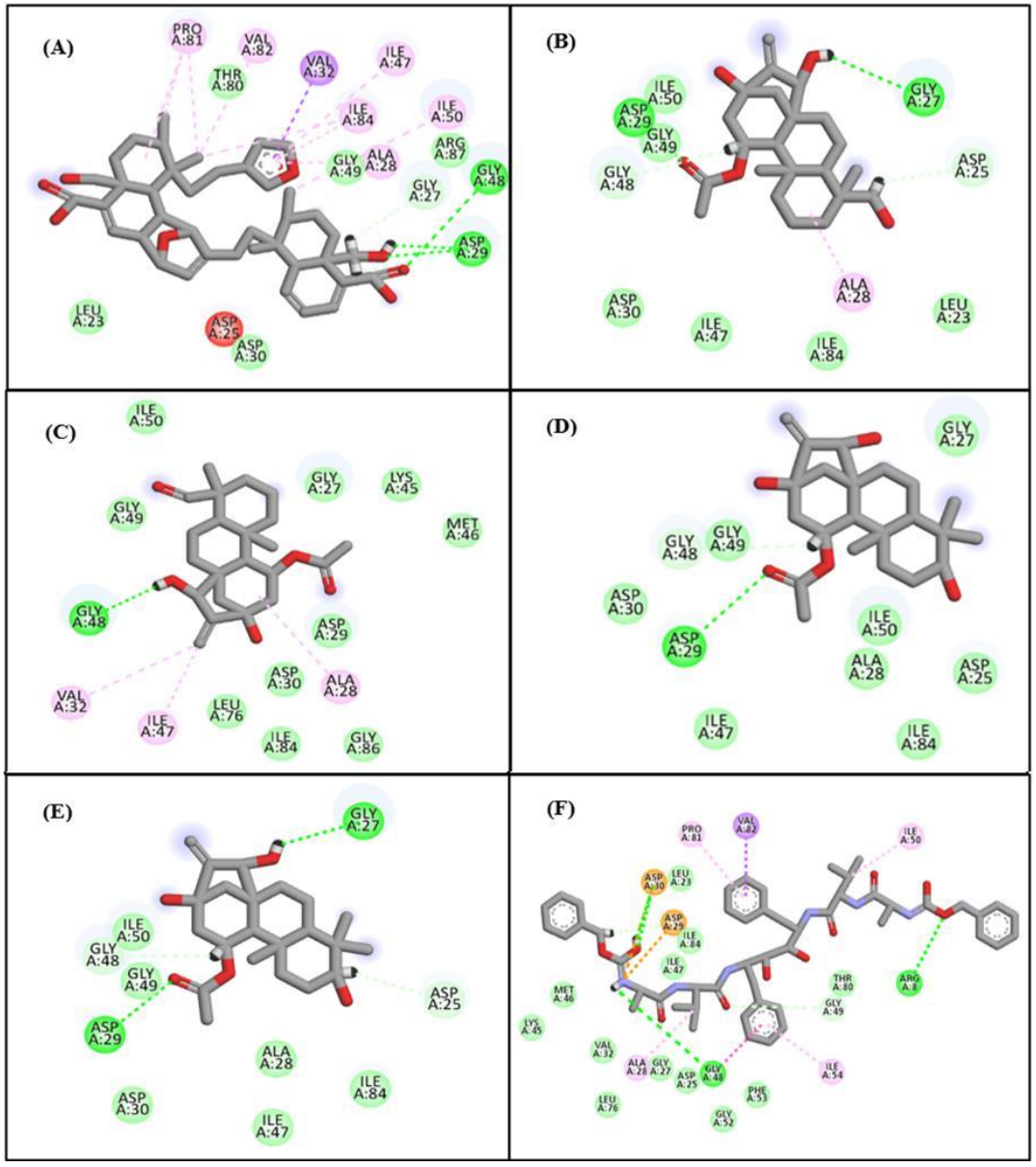
Los resultados mostrados en la Tabla 2 permitieron evidenciar que el valor más bajo de energía total es el de el ligando TL-3, para lo cual la molécula **109** es la más cercana. Mientras que el valor más alto de energía es el de la molécula **231** lo que corresponde con su clasificación. El valor de de energía total obtenido para el ligando **109** es el más bajo comparado con los demás ligandos, teniendo que en cuenta que a niveles más bajos de energía existe mayor estabilidad, podemos identificar que esta es una de las razones por las cuales es el ligando más que presentó

más afinidad en el sitio activo. En cuanto al ligando **234** y **235**, podemos encontrar valores de energía similares, sin embargo, el valor de energía para el compuesto **235** es más alta que la del compuesto **234**. Respecto a los compuestos **230** y **231**, podemos mencionar que también presentaron niveles de energía similares que corresponden a su clasificación.

Los puentes de hidrógeno favorecen la interacción proteína-ligando influyendo de manera significativa en la estabilidad del ligando en el sitio activo fortaleciendo la unión del ligando. Los enlaces de hidrógeno contribuyen a la energía libre de unión, ya que estos enlaces son particularmente fuertes e importantes en el entorno hidrofóbico de la membrana celular (Paucara & Torrez, 2019). Estos enlaces acortan la distancia entre el protón y el núcleo al que pertenece el par solitario de electrones (Th. Zeegers, 2011). Los compuestos **109**, **234**, **235**, **231** y el ligando TL-3 presentaron interacciones de puentes de hidrógeno con el residuo aminoácido glicina A:27 y A:48 (Figura 8).

Figura 9.

Diagramas de interacción proteína-diterpeno



Nota: Diagramas de interacción residual para los compuestos A)**109**, B) **234**, C) **235**, D) **230**, E) **231**, F) Ligando TL-3. Los residuos que interactúan se muestran en círculos de colores y líneas discontinuas dependiendo del tipo de interacción: enlace H (lima), Van der Waals (verde), π–π (púrpura), π–alquilo (rosa), enlace de carbono H (verde claro), carga atractiva (naranja), amida-π apilada (fucsia).

Respecto al compuesto **109** (Figura 8a), la interacción se realizó con el oxígeno del ligando del C-28, en cuanto a este tipo de interacción esta unión sería la única de esta clase, por lo que se puede deducir que tiene relación con el tipo de compuesto, pues la molécula **109** es un clerodano lo que resalta un aspecto favorable. En cuanto a los compuestos **234**, **230** y **231**, esta interacción se da con grupos hidroxilo en C-15 pues este grupo puede actuar como donante o receptor en este tipo de enlaces. Los compuestos **109**, **234**, **230** y **231** presentaron igualmente interacciones de puentes de hidrógeno con el residuo aminoácido de ácido aspártico A:29 en el caso de la molécula **109** esta interacción se da en C-28 y C-31 mientras que en las moléculas **234**, **230** y **231** la interacción se da en C-11. La interacción se forma con el grupo hidroxilo y el ácido carboxílico.

El compuesto **109** y el ligando TL-3 cuentan con anillos aromáticos en su estructura como el ciclopentadieno y el benceno, respectivamente. La aromaticidad como característica estructural, favorece las interacciones pi-sigma, las cuales relacionan enlaces tipo C-C, C-H, enlaces con grupos aromáticos, así como con alquenos y alquinos conjugados. Dicha interacción se dio con el residuo aminoácido valina A:32 el cual es un aminoácido que cuenta con cadena lateral no aromática. Con respecto al compuesto **109**, los dobles enlaces conjugados presentes en el ciclopentadieno favorecen la formación de este tipo de interacción, así como las fuerzas de Van der Waals. En lo referente al ligando TL-3, podríamos evaluar una situación similar, pues los enlaces dobles conjugados del benceno favorecieron este tipo de interacción. Este tipo de uniones contribuyen a la afinidad del ligando en su sitio de acción. Esto representó un aspecto bastante relevante para los resultados ya que indicaría la razón por la cual el compuesto **109** tiene un mejor acoplamiento en cuanto a los demás.

En cuanto a las interacciones alquilo y pi-alquilo, los compuestos **109**, **234**, **235** y el ligando TL-3 presentaron este tipo de enlace. Para que dicha interacción se forme, es importante la acción

de las fuerzas de Van der Waals, las cuales se encuentran presentes de forma significativa en las estructuras **109**, **234**, **235** y el ligando TL-3. Respecto al compuesto **109**, se encontraron 6 interacciones de este tipo para las cuales el enlace se da con 4 tipos de residuos aminoácidos: valina A:82(1), prolina A:81(1), isoleucina A:47, A:84, A:50 (3), alanina A:28 (1). En cuanto al compuesto **234** el enlace solo se da con el residuo aminoácido alanina A:28. Respecto al compuesto **235**, el enlace se da con 3 tipos de residuos aminoácidos: valina A:32(1), isoleucina A:47 (1), alanina A:28(1). Y en cuanto al ligando TL-3 el enlace se da con tres tipos de residuos aminoácidos: prolina A:81 (1), isoleucina A:50, A:54(2), alanina A:28 (1).

Para la interacción valina-hidrogeno (compuesto **109** y **235**) se debe tener en cuenta que cerca del hidrogeno existen anillos aromáticos por lo cual se puede dar una interacción pi-alquilo. En lo referente a la interacción prolina-anillos aromáticos (compuesto **109** y ligando TL-3), gracias a la cadena lateral cíclica de la prolina se pueden formar interacciones alquilo, mientras que las pi-alquilo se dan gracias a la interacción de la cadena lateral alifática de la prolina con el sistema pi del anillo aromático. En cuanto al enlace isoleucina-hidrogeno (compuesto **209** y **235**) se produciría un efecto similar al explicado para la interacción valina-hidrógeno debido a la cadena no alifática de la isoleucina. Respecto al enlace isoleucina-anillos aromáticos (compuesto **109** y ligando TL-3) el enlace se dió gracias a las fuerzas de Van der Waals en la interacción del sistema pi del anillo con la cadena no alifática del residuo aminoácido. Respecto a la interacción alanina-hidrógeno (Ligando TL-3) la alanina contiene una cadena corta alifática que favorece la formación de interacciones alquilo. (Nelson & Cox, 2008)

Así mismo, en cuanto a la interacción alanina-anillos aromáticos (compuesto **109** y **234**) gracias al sistema pi de los anillos aromáticos, la cadena alifática puede interactuar para formar el enlace pi-alquilo sin dificultad alguna. El compuesto que presentó más similitud con el ligando

original fue el **109**, pues cuenta con varias interacciones similares. Este tipo de enlaces favorecen la estabilidad y afinidad del compuesto determinando la función y estructura del compuesto.

En lo referente a los enlaces carbono hidrógeno, son de gran importancia pues permiten definir la orientación del ligando en el sitio activo, aunque suelen ser menos fuertes que los puentes de hidrógeno. La razón por la cual este enlace es más débil es gracias a factores como la electronegatividad del carbono, la cual es bastante baja en comparación a la electronegatividad que tienen átomos como el nitrógeno o el oxígeno, que son los átomos que normalmente conforman un enlace H-H (Nelson & Cox, 2008). Este tipo de enlaces C-H se encuentran presentes en los compuestos **109**, **234**, **231**,**230** y en el ligando TL-3. Respecto al compuesto **109**, este tipo de enlace se encuentra en el hidrógeno que se encuentra C-28 y el residuo aminoácido glicina A:27. En cuanto al compuesto **234**, se encuentra en el hidrógeno de la posición C-11 junto con el residuo aminoácido glicina A:48, así como en el hidrógeno de la posición C-18 con el residuo aminoácido ácido aspártico A:25. Con relación al compuesto **231** la relación se da en las mismas posiciones y con los mismos aminoácidos del compuesto **234**. En cuanto al compuesto **230** la interacción se da con el átomo de hidrógeno de C-11 junto con el residuo aminoácido glicina A:49. En relación al ligando LT-3, dicha relación se presenta con el residuo aminoácido glicina A:49.

Las fuerzas de Van der Waals aumentan su fuerza, a medida que aumenta el tamaño de la molécula (StudySmarter, s. f.). Este tipo de fuerza intermolecular depende de las fuerzas de dispersión, dipolo-dipolo y dipolo inducido-dipolo inducido de los átomos presentes en las moléculas provocando que se atraigan entre sí por el efecto electrostático generado por la atracción del polo positivo de una molécula con el polo negativo de otra (Fuerzas de Van Der Waals, 2018). Todos los compuestos evaluados cuentan con fuerzas de fuerzas de Van der Waals siendo respectivamente el compuesto **109**, **234**, **235**, **230**, **231** y el ligando LT-3.

En cuanto al compuesto **109** este tipo de interacción se presenta en 5 residuos aminoácidos tales como leucina A:23, ácido aspártico A:30, treonina A:80, glicina A:49, arginina A:87. Respecto al compuesto **234**, encontramos 6 residuos aminoácidos tales como isoleucina A:50, A:47, A:84, glicina A:49, ácido aspártico A:30 y leucina A:23. A su vez, para el compuesto **235** este tipo de interacción se encuentra en 10 residuos aminoácidos tales como isoleucina A:50 y A:84, glicina A:49 y A:27, leucina A:76, ácido aspártico A:30, A:29, lisina A:45, metionina A:46. Sobre el compuesto **231**, los 6 aminoácidos presentes en esta fuerza intermolecular son isoleucina A:84, A:47, ácido aspártico A:30, alanina A:28.

En lo referente a la molécula **230** podemos encontrar 8 residuos aminoácidos glicina A:49, A:27, isoleucina A:50, A:84, ácido aspártico A:30, A:25, alanina A:28. Respecto a la molécula **231** este tipo de interacciones se encuentran presentes en 12 residuos aminoácidos tales como isoleucina A:50, A:47, A:84, glicina A:49, ácido aspártico A:30, alanina A:28. Y finalmente, respecto al ligando TL-3, encontramos 11 residuos aminoácidos relacionados con esta interacción tales como isoleucina A:84, A:47, leucina A:23, A:76, lisina A:45, metionina A:46, valina A:32, ácido aspártico A:25, fenilalanina A:53.

Las fuerzas de Van der Waals más fuertes se presentan en los aminoácidos que cuentan con cadenas laterales largas como la leucina, isoleucina, arginina, lisina, metionina, fenilalanina (Nelson & Cox, 2008). Los residuos aminoácidos leucina, isoleucina y metionina cuentan con cadenas grandes y apolares lo que favorece la formación de fuerzas de dispersión, respecto al residuo aminoácido arginina y lisina, presentan cadenas grandes pero se encuentran cargadas positivamente lo que favorece la formación de fuerzas de Van der Wals tipo dispersión, iónicas y dipolo-dipolo y en lo referente al residuo aminoácido fenilalanina, este cuenta con un una cadena grande y aromática, lo cual favorece la formación de fuerzas de dispersión (Nelson & Cox, 2008)

9. CONCLUSIONES

De acuerdo a la investigación realizada se pueden concluir varios aspectos relacionados con los objetivos inicialmente planteados. En cuanto al modelo de predicción desarrollado en KNIME 5.2.5, se obtuvo un modelo de predicción capaz de clasificar oportunamente compuestos activos en inactivos lo que permite obtener resultados favorables para métricas de evaluación como exactitud, precisión y sensibilidad gracias a la adecuada identificación de verdaderos positivos y falsos negativos. En lo referente al modelo de predicción de *docking* molecular, se identificaron 5 compuestos con acción inhibitoria para la proteasa del VIH-1, siendo respectivamente las moléculas **109**, **234**, **235**, **231** y **230** que presentan mayor afinidad con el sitio activo de la proteína.

La molécula **109** es un diterpeno clerodano y obtuvo una energía total de -128,49 KJ/mol lo cual la hizo la molécula con menor valor de energía y con mejor afinidad, mientras que las moléculas **234**, **235**, **231** y **230** obtuvieron valores de energía de -75,88 KJ/mol, -77,75 KJ/mol, -69,91 KJ/mol, -68,85 KJ/mol respectivamente. Al realizar el respectivo análisis de los diagramas de interacción residual se logró identificar que una de las razones por las cuales el compuesto **109** presenta mayor afinidad con la proteína es por algunas fuerzas intermoleculares características presentes en dicha interacción tales como la interacción pi-sigma, pi-alquilo y alquilo las cuales brindan estabilidad al complejo ligando proteína favoreciendo la energía de unión para obtener niveles bajos de energía provenientes del *docking*.

En cuanto a las moléculas **234**, **235**, **231** y **230** existe una característica respecto a sus interacciones moleculares con los residuos aminoácidos pues todos presentan enlaces de hidrógeno en posiciones similares la cual es una de las razones por las cuales este tipo de compuestos fueron clasificados como los mejores ligandos, teniendo en cuenta que los enlaces de hidrógeno representan interacciones bastante fuertes que favorecen la estabilidad del ligando. Sin embargo,

uno de los aspectos que hace mejor a los compuestos **234** y **235** es el enlace alquil presente en las interacciones intermoleculares que tal como se mencionó anteriormente, favorecen la afinidad del ligando y son fuerzas más fuertes que los enlaces de hidrógeno. También se presentaron otro tipo de interacciones intermoleculares tales como las fuerzas de Vander Waals, las cuales estuvieron presentes de forma significativa en todos los ligandos seleccionados. En cuanto al ligando TL-3, presentó fuerzas intermoleculares como la pi-amida y la carga atractiva que no fueron encontradas en otros compuestos, razón por la cual se puede afirmar que presenta más afinidad y estabilidad que los compuestos evaluados.

Por tanto, el metabolito secundario que puede funcionar como potencial inhibidor de la enzima proteasa del VIH-1 es la molécula **109** el cual es proveniente de la planta *Baccharis flabellata* perteneciente a la familia Asteraceae con origen en América Latina.

10. REFERENCIAS (BIBLIOGRAFÍA)

Anazodo, M. I., Salomon, H., Friesen, A. D., Wainberg, M. A., & Wright, J. A. (1995). *Antiviral activity and protection of cells against human immunodeficiency virus type-1 using an antisense oligodeoxyribonucleotide phosphorothioate complementary to the 5'-LTR region of the viral genome.* Gene, 166(2), 227-232. https://doi.org/10.1016/0378-1119(95)00582-x

Aprendizaje Automático y Técnicas de Validación. (2020*). [Escuela Técnica Superior de Ingeniería Universidad de Sevilla].* https://biblus.us.es/bibing/proyectos/abreproy/92987/fichero/TFG-2987+SOTO+MARCHENA%2C+DAVID.pdf

Arnold, B. H. (1969). *HYDROLYSES OF THE ACETOXYMETHYL GROUP AT THE 3-POSITION OF THE CEPHALOSPORN NUCLEUS.* United States Patent Office. https://patentimages.storage.googleapis.com/dc/9c/49/9f099a6037cec4/US3436310.pdf

Arts, E. J., & Hazuda, D. J. (2012). *HIV-1 antiretroviral drug therapy. Cold Spring Harbor Perspectives In Medicine*, 2(4), a007161. https://doi.org/10.1101/cshperspect.a007161

Bartuzi, D., Kaczor, A., Targowska-Duda, K., & Matosiuk, D. (2017). *Recent Advances and Applications of Molecular to G Protein-Coupled Receptors. Molecules/Molecules Online/Molecules Annual,* 22(2), 340. https://doi.org/10.3390/molecules22020340

Cachay, E. R. (2023, 9 febrero). *Infección por el virus de la inmunodeficiencia humana (VIH). Manual MSD Versión Para Público General*. https://www.msdmanuals.com/es/hogar/infecciones/infecci%C3%B3n-por-el-virus-de-la-inmunodeficiencia-humana-vih/infecci%C3%B3n-por-el-virus-de-la-inmunodeficiencia-humana-vih

Carreres-Prieto, M., López-Sisamón, D., & Layos-Romero, L. (2015). *[Interstitial lung disease induced by raltitrexed-oxaliplatin based chemotherapy for colorectal cancer: a case report*]. DOAJ (DOAJ: Directory Of Open Access Journals), 39(2), 118-119.

Cumming, J. G., Davis, A. M., Mureşan, S., Haeberlein, M., & Chen, H. (2013). *Chemical predictive modelling to improve compound quality. Nature Reviews Drug Discovery*, 12(12), 948-962. https://doi.org/10.1038/nrd4128

De Clercq, E. (2009). *The history of antiretrovirals: key discoveries over the past 25 years. Reviews In Medical Virology*, 19(5), 287-299. https://doi.org/10.1002/rmv.624

Deeks, S. G. (2006). *Antiretroviral treatment of HIV infected adults*. The BMJ, 332(7556), 1489. https://doi.org/10.1136/bmj.332.7556.1489

Delgado, R. (2011). *Características virológicas del VIH. Enfermedades Infecciosas y Microbiología Clínica*, 29(1), 58-65. https://doi.org/10.1016/j.eimc.2010.10.001

Derewenda, Z. S. (2023). C-H *Groups as Donors in Hydrogen Bonds: A Historical*

Overview and Occurrence in Proteins and Nucleic Acids. International Journal Of Molecular Sciences, 24(17), 13165. https://doi.org/10.3390/ijms241713165

Drzewiecki, W. (2017). *Thorough statistical comparison of Machine Learning regression models and their ensembles for sub-pixel imperviousness and imperviousness change mapping.* Geodesy an Cartography, 66(2), 171-209.

Enrique, S. R. L. (s. f.). *Mecanismos patogénicos de la infección por VIH.* https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0034-83762004000200005

Ferreira, L., Santos, R. D., Oliva, G., & Andricopulo, A. (2015). *Molecular Docking and Structure-Based Drug Design Strategies. Molecules/Molecules Online/Molecules Annual*, 20(7), 13384-13421. https://doi.org/10.3390/molecules200713384

Franzusoff, A., Volpe, A. M., Josse, D., Pîchuantes, S., & Wolf, J. R. (1995). *Biochemical and Genetic Definition of the Cellular Protease Required for HIV-1 gp160 Processing. Journal Of Biological Chemistry*, 270(7), 3154-3159. https://doi.org/10.1074/jbc.270.7.3154

Fuerzas de Van der Waals. (2018, 21 febrero). *Portal Académico del CCH*. https://e1.portalacademico.cch.unam.mx/alumno/quimica1/unidad2/tiposdeenlaces/vanderwaals

Halperin, I., Wolfson, H., Nussinov, R., sitelight: *Binding-site prediction using phage display libraries. Protein Science* 2003, 12 (7), 1344-1359.

Handbook of molecular descriptors. Methods and principles in medicinal chemistry, volume 11 Edited by Roberto Todeschini, Viviana Consonni (a series edited by R. Mannhold, H. Kubinyi, H. Timmerman), Wiley-VCH, Weinheim, 2000. 667 pp.; E160.00. (2001). European Journal Of Medicinal Chemistry, 36(11-12), 966. https://doi.org/10.1016/s0223-5234(01)80018-0

Hassan, M. (2023, 15 agosto). *Documentary Research - Types, Methods and Examples.* Research Method. https://researchmethod.net/documentary-research/

Herrera-Acevedo, C., Dos Santos Maia, M., Cavalcanti, E. B. V. S., Coy-Barrera, E., Scotti, L., & Scotti, M. T. (2021). *Selection of antileishmanial sesquiterpene lactones from SistematX database using a combined ligand-/structure-based virtual screening approach. Molecular Diversity*, 25, 2411-2427.

Herrera-Acevedo, C., Perdomo-Madrigal, C., De Sousa Luis, J. A., Scotti, L., & Scotti, M. T. (2022). *Drug discovery Paradigms: Target-Based drug discovery. En Springer eBooks* (pp. 1-24). https://doi.org/10.1007/978-3-030-95895-4_1

Investigación experimental. (s. f.). https://explorable.com/es/investigacion-experimental

Jarboe, L. R., Royce, L. A., & Liu, P. (2013). *Understanding biocatalyst inhibition by carboxylic acids. Frontiers In Microbiology*, 4. https://doi.org/10.3389/fmicb.2013.00272

Jarboe, L. R., Royce, L. A., & Liu, P. (2013). *Understanding biocatalyst inhibition by carboxylic acids. Frontiers In Microbiology*, 4. https://doi.org/10.3389/fmicb.2013.00272

Jiang, H., Wang, W., Tang, J., Liu, M., Li, X., Hu, K., Du, X., Li, X., Zhang, H., Pu, J., & Sun, H. (2017). *Structurally Diverse Diterpenoids from Isodon scoparius and Their Bioactivity. Journal Of Natural Products*, 80(7), 2026-2036. https://doi.org/10.1021/acs.jnatprod.7b00163

KNIME Press | KNIME. (s. f.). KNIME. https://www.knime.com/knimepress?pk_vid=881a2cfcf27900ab1719972229b5db68#c heat-sheets

Lechner, M., Moser, S., Pander, J., Geist, J., & Lewalter, D. (2024). *Learning scientific observation with worked examples in a digital learning environment.* Frontiers In Education, 9. https://doi.org/10.3389/feduc.2024.1293516

Li, R., Morris-Natschke, S. L., & Lee, K. H. (2016). *Clerodane diterpenes: sources, structures, and biological activities. Natural Product Reports*, 33(10), 1166-1226. https://doi.org/10.1039/c5np00137d

Lin L-G, Lam Ung CO, Feng Z-L, Huang L, Hu H. (2016) *Naturally occurring diterpenoid dimers: source, biosynthesis, chemistry and bioactivities.Planta Medica*, 82, 1309-1328.

Lin, J., Perryman, A. L., Schames, J. R., & McCammon, J. A. (2002). *Computational drug design Accommodating receptor flexibility: the relaxed complex scheme.* Journal Of The American Chemical Society, 124(20), 5632-5633. https://doi.org/10.1021/ja0260162

Mathematical Model considering Antiretroviral Administration. https://doi.org/10.17488/rmib.38.3.5

Métricas de evaluación de modelos en el aprendizaje automático. (2023, 25 septiembre). DataSource.ai. https://www.datasource.ai/es/data-science-articles/metricas-de-evaluacion-de-modelos-en-el-aprendizaje-automatico

MONTES GRAJALES, D. (s. f.). *ORGANIZACIÓN ESTRUCTURAL DE PROTEÍNAS.* Universidad de Cartagena.

Morales Torrado, O. (2022). *MUTACIONES POR LA RESISTENCIA AL TRATAMIENTO ANTIRRETROVIRAL EN PACIENTES CON VIH+1 DE UNA IPS DE LA CIUDAD DE BARRANQUILLACOLOMBIA EN EL PERIODO DE 2009 a 2013* [Tesis Magister, Universidad del norte]. https://manglar.uninorte.edu.co/bitstream/handle/10584/11203/73106272.pdf?sequence=1&isAllowed=y

Naqa, I. E., & Murphy, M. J. (2015). *What is Machine Learning? En Springer eBooks* (pp. 3-11). https://doi.org/10.1007/978-3-319-18305-3_1

Nelson, D., & Cox, M. (2008). *Lehninger Principles of Biochemistry (5th Edition). Macmillan Learning.* https://repository.tudelft.nl/view/MMP/uuid:ce33aac8-a458-478e-9f5b-

84f7b5bc75f0/

OMS. (2020). *Terapia antirretroviral. OPS/OMS* | Organización Panamericana de la Salud. https://www.paho.org/es/temas/terapia-antirretroviral

OMS. (2024, 27 mayo). *VIH/SIDA. OPS/OMS* | Organización Panamericana de la Salud. https://www.paho.org/es/temas/vihSIDA

Pagare, S., Bhatia, M., Tripathi, N., Pagare, S., & Bansal, Y. K. (2015). *Secondary Metabolites of Plants and their Role: Overview. Current Trends In Biotechnology And Pharmacy*, 9(3), 293-304. https://www.indianjournals.com/ijor.aspx?target=ijor:ctbp&volume=9&issue=3&article=011

Paucara, W. G. B., & Torrez, R. E. G. (2019). *Acomplamiento molecular: criterios prácticos para la selección de ligandos biológicamente activos e identificación de nuevos blancos terapéuticos.* Revista CON-CIENCIA, 7(2), 55-72. http://www.scielo.org.bo/pdf/rcfb/v7n2/v7n2_a06.pdf

Rodriguez, M. V., Butassi, E., Funes, M., & Zacchino, S. A. (2019*). Synergism between Terbinafine and a Neo-clerodane Dimer or a Monomer Isolated from Baccharis flabellata against Trichophyton rubrum. Natural Product Communications,* 14(1), 1934578X1901400. https://doi.org/10.1177/1934578x1901400101

Rosselli, S., Bruno, M., Maggio, A., Bellone, G., Chen, T. H., Bastow, K. F., & Lee, K. (2007). *Cytotoxic Activity of Some Natural and Synthetic ent-Kauranes. Journal Of Natural Products*, 70(3), 347-352. https://doi.org/10.1021/np060504w

Ruiz, L. (s. f.). Investigación experimental. En *monografias.com*. Recuperado 18 de julio de 2024, de https://www.scientific-european-federation-osteopaths.org/wp-content/uploads/2019/01/Investigaci%C3%B3n-experimental.pdf

Salomatina, O. V., Sen'kova, A. V., Moralev, A. D., Savin, I. A., Komarova, N. I., Salakhutdinov, N. F., Zenkova, M. A., & Markov, A. V. (2022). *Novel Epoxides of Soloxolone Methyl: An Effect of the Formation of Oxirane Ring and Stereoisomerism on Cytotoxic Profile, Anti-Metastatic and Anti-Inflammatory Activities In Vitro and In Vivo.* International Journal Of Molecular Sciences, 23(11), 6214. https://doi.org/10.3390/ijms23116214

Sanches, M., Krauchenco, S., Martins, N. H., Gustchina, A., Wlodawer, A., & Polikarpov, I. (2007). *Structural Characterization of B and non-B Subtypes of HIV-Protease: Insights into the Natural Susceptibility to Drug Resistance Development. Journal Of Molecular Biology/Journal Of Molecular Biology*, 369(4), 1029-1040. https://doi.org/10.1016/j.jmb.2007.03.049

Singh, S., Brocker, C., Koppaka, V., Chen, Y., Jackson, B. C., Matsumoto, A., Thompson, D. C., & Vasiliou, V. (2013). *Aldehyde dehydrogenases in cellular responses to oxidative/electrophilicstress. Free Radical Biology & Medicine*, 56, 89-101. https://doi.org/10.1016/j.freeradbiomed.2012.11.010

Sosa, N. (s. f.). *Avances en VIH/SIDA y complicaciones de la terapia antirretroviral*. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-24482007000300016

StudySmarter. (s. f.). Fuerzas intermoleculares: Tipos y explicaciones | StudySmarter.

StudySmarter ES. https://www.studysmarter.es/resumenes/quimica/enlaces-quimicos/fuerzas-intermoleculares/

Takahashi, J., Gomes, D., Lyra, F., Santos, G. D., & Martins, L. (2014b). *The Remarkable Structural Diversity Achieved in ent-Kaurane Diterpenes by Fungal Biotransformations. Molecules/Molecules Online/Molecules Annual*, 19(2), 1856-1886. https://doi.org/10.3390/molecules19021856

Th. Zeegers. (2011). *Intermolecular and Surface Forces. En Elsevier eBooks* (p. iii). https://doi.org/10.1016/b978-0-12-391927-4.10024-6

UNAIDS. (2023). *Hoja informativa — Últimas estadísticas sobre el estado de la epidemia* de SIDA. https://www.unaids.org/es/resources/fact-sheet

Vanegas-Otálvaro, D., Acevedo-Sáenz, L., Díaz-Castrillón, F. J., & Velilla-Hernández, P. A. (2014). *Resistencia a antirretrovirales: bases moleculares e implicaciones farmacológicas.* Revista CES Medicina, 28(1), 91-106. https://dialnet.unirioja.es/descarga/articulo/4804457.pdf

Vella, S., Schwartländer, B., Sow, S., Eholié, S., & Murphy, R. L. (2012). *The history of antiretroviral therapy and of its implementation in resource-limited areas of the world. AIDS,* 26(10), 1**231**-1241. https://doi.org/10.**109**7/qad.0b013e32835521a3

W&B. (2024, 30 junio). *Weights & biases*. W&B. https://wandb.ai/mostafaibrahim17/ml-

articles/reports/An-Introduction-to-the-F1-Score-in-Machine-Learning--Vmlldzo2OTY0Mzg1

Wynn, G. H., Zapor, M., Smith, B., Wortmann, G., Oesterheld, J. R., Armstrong, S. C., & Cozza, K. L. (2004). Antiretrovirals, *Part 1: Overview, History, and Focus on Protease Inhibitors. Psychosomatics*, 45(3), 262-270. https://doi.org/10.1176/appi.psy.45.3.262

Xu, Y., & Goodacre, R. (2018b). *On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning. Journal Of Analysis And Testing*, 2(3), 249-262. https://doi.org/10.1007/s41664-018-0068-2

Zambrano, M. M., & Del Rosario de la Torre Cruz, M. (s. f.). *El mundo del trabajo: estudiantes, egresados y empleadores desde la perspectiva de género*. ResearchGate. https://www.researchgate.net/publication/371446328_El_mundo_del_trabajo_estudiantes_egresados_y_empleadores_desde_la_perspectiva_de_genero