

TESIS DOCTORAL



UCAM

UNIVERSIDAD CATÓLICA
DE MURCIA

ESCUELA INTERNACIONAL DE DOCTORADO

Programa de Doctorado Ciencias de la Salud

Aplicación de técnicas de Dinámica Molecular y Machine Learning al descubrimiento de fármacos

Autor/a:

Miguel Carmena Bargueño

Directores/as:

Dr. D. Horacio Pérez Sánchez

Murcia, septiembre de 2024

TESIS DOCTORAL



UCAM

UNIVERSIDAD CATÓLICA
DE MURCIA

ESCUELA INTERNACIONAL DE DOCTORADO

Programa de Doctorado Ciencias de la Salud

Aplicación de técnicas de Dinámica Molecular y Machine Learning al descubrimiento de fármacos

Autor/a:

Miguel Carmena Bargueño

Directores/as:

Dr. D. Horacio Pérez Sánchez

Murcia, septiembre de 2024



AUTORIZACIÓN DEL DIRECTOR DE LA TESIS PARA SU PRESENTACIÓN

El Dr. D. Horacio Pérez Sánchez como Director de la Tesis Doctoral titulada “Aplicación de técnicas de Dinámica Molecular y Machine Learning al descubrimiento de fármacos” realizada por D. Miguel Carmena Bargueño en el Programa de Doctorado Ciencias de la Salud, **autoriza su presentación a trámite** dado que reúne las condiciones necesarias para su defensa.

Lo que firmo, para dar cumplimiento al Real Decreto 99/2011 de 28 de enero, en Murcia a 27 de septiembre de 2024.

A handwritten signature in blue ink, consisting of several overlapping loops and a vertical stroke, positioned above the name of the signatory.

Dr. Horacio Pérez Sánchez

RESUMEN

El descubrimiento de fármacos todavía es un desafío por resolver en la actualidad a pesar de los grandes avances en medicina y tecnología debido a que hay un gran número de enfermedades sin tratamiento, un aumento en las patologías crónicas y el incremento de microorganismos multirresistentes. Este desafío se debe principalmente a que los mecanismos moleculares que regulan los sistemas que participan en las patologías son muy complejos. Además, actualmente hay librerías de millones de compuestos que son candidatos para tratar diversas patologías. Por todo ello, se necesita una técnica rápida y de bajo coste económico para poder priorizar los ligandos a testar. Las técnicas *in silico* permiten obtener candidatos con un gran potencial de una forma rápida. Las herramientas con mayor proyección actualmente en el descubrimiento de fármacos son las simulaciones de Dinámica Molecular y el *Machine Learning*.

El objetivo de esta tesis es mejorar la capacidad predictiva de los diferentes métodos *in silico* que hay en la actualidad, concretamente los métodos conocidos como CAR (Computación de Alto Rendimiento). Para ello, se han utilizado diversidad de técnicas computacionales, centradas en la Dinámica Molecular y el Machine Learning. Posteriormente, como objetivo adicional se busca la validación e implementación de dichas técnicas en el tratamiento de enfermedades como el cáncer o la diabetes.

Los software utilizados en esta tesis fueron Maestro-Desmond para realizar cálculos de Dinámica Molecular, SIBILA para generar modelos de Machine Learning, así como diversos software para cribado virtual basado tanto en la estructura como en los ligandos.

Se ha conseguido desarrollar un programa que permite ejecutar y analizar simulaciones de Dinámica Molecular de forma masiva. Este programa llamado

TOLEDO (Throughput Optimization of Ligand-Protein systems Exploration through Dynamics simulation in Optimized HPC systems) ha sido utilizado para descubrir el mecanismo de actuación de proteínas como la rodopsina.

Este resultado junto con el uso de técnicas de cribado virtual y *Machine Learning*, permite concluir cómo se pueden seguir mejorando las técnicas CAR en diversos contextos tales como el cáncer, la resistencia a los antibióticos y la cicatrización. Además, se han ido desarrollando diversos protocolos que permiten optimizar la búsqueda de nuevos compuestos.

ABSTRACT

Nowadays, drug discovery remains a challenge that needs to be solved despite the significant advances in technology and medicine. The main reasons are the large number of untreated diseases, the increase in chronic pathologies, and the rise of multiresistant microorganisms. This problem is due to the complexity of the molecular mechanisms that regulate the systems involved in these pathologies. Moreover, there are currently libraries of millions of compounds that are candidates for treating various diseases. Therefore, rapid, and low-cost techniques are needed to prioritize the ligands to be evaluated. *In silico* tools allow candidates with great potential to be obtained faster than other techniques. The tools with the most significant projection currently in drug discovery are Molecular Dynamics simulations and Artificial Intelligence.

This thesis aims to improve the predictive capabilities of the different *in silico* methods available nowadays, specifically those known as HPC (High-Performance Computing). For this, various computational techniques have been used, focusing on Molecular Dynamics and Machine Learning. Subsequently, as an additional objective, the validation and implementation of these techniques in treating diseases such as cancer or diabetes are sought.

The software used in this thesis to perform Molecular Dynamics calculations was Maestro-Desmond. SIBILA was used to generate Machine Learning models, and various software was used for virtual screening based on structure and ligands.

Our software, TOLEDO (Throughput Optimization of Ligand-Protein systems Exploration through Dynamics simulation in Optimized HPC systems), allows the execution and analysis of massive molecular dynamics simulations. It has been used to discover the mechanism of proteins such as rhodopsin.

These results, along with the use of virtual screening and artificial intelligence techniques, allow concluding how HPC techniques can continue to be improved and used in various contexts such as cancer, antibiotic resistance, and healing. In addition, various protocols have been developed that allow optimizing the search for new drugs.

PALABRAS CLAVE

Diseño síntesis y estudio de nuevos fármacos, inteligencia artificial, farmacología, software

KEYWORDS

Artificial intelligence, drug discovery, pharmacology, software

AGRADECIMIENTOS

Quiero empezar agradeciendo a la UCAM por darme la oportunidad de realizar esta tesis doctoral, así como también a la UCLM y la UM por permitirme realizar el grado y el máster que me han habilitado para realizar esta tesis.

En el ámbito personal, agradecer tanto a mi director de tesis Horacio, como a mis compañeros del grupo BIO-HPC: Carlos, Antonio, Jochem, Alejandro, Maria, Elena y Jorge por haberme apoyado y ayudado tanto a nivel profesional como personal a la realización de esta tesis doctoral. Durante estos años me han ayudado a aprender enormemente tanto en programación como en técnicas computacionales. También una mención especial a mi tutor y profesor de TFG Juan Ángel por introducirme en el mundo de la bioinformática.

Agradecer a los colaboradores con lo he tenido el placer de trabajar y aprender durante estos cuatro años, con una mención especial al Dr. Pablo Pelegrín y al Dr. Juan José Hernández Morante con los que he aprendido conocimientos sobre inflamación y nutrición. También me gustaría agradecer al grupo del Dr. Thierry Langer por permite hacer la estancia en su laboratorio. Además del Dr. Langer quiero agradecer a Oliver, Nedra, Sara, Linda, Phillip, Jessica, Verena, Christian, Thomas, Ángel y Daniel por acogerme en el laboratorio los meses que estuve en Viena.

A mis compañeros de la UCAM: Juan Antonio, Carolina, los Joses y todo el departamento de profesores de enfermería por acogerme durante el primer año de la tesis y ayudarme en el tiempo que estuve allí. También agradecer a todos mis compañeros de HiTech en especial a Cindy, Ylenia, Pepa e Isabel por el apoyo, ayudas y consejos para la realización de esta tesis. Además, agradecer a mi tutora Merry por ayudarme en todas las dudas administrativas que me ha resuelto de forma eficaz y con rapidez.

Agradecer a mis amistades de Toledo su apoyo moral durante estos cuatro años; especialmente a Iván, Celia, Cynthia, Alejandro, Diego, Maeva, Irene y Rocío porque gracias a ellos he conseguido desconectar y coger fuerzas para realizar esta tesis.

También a mis amigos del máster JL, Belén, Adam, Alberto, Elena y Pedro por enseñarme a amar a la ciudad de Murcia y a su gente.

Y por último y más importante agradecer a mi familia, en especial a mis padres, Raimundo y Montse, y a mi hermana Julia por apoyarme durante toda mi vida y motivarme para estudiar y poder dedicarme a mi pasión por la biología y los ordenadores. Agradecer, además, al resto de familia, tíos y primos por apoyarme durante esta etapa.

También a mi abuela que me cuidó desde que nací y sé que estará viéndome orgullosa desde el cielo.

"Lo importante es no dejar de hacerse preguntas ". Albert
Einstein (1879-1955).

ÍNDICE GENERAL

RESUMEN	7
I - INTRODUCCIÓN	29
1.1. Descubrimiento de fármacos	29
1.2. Enfermedades actuales	30
1.3. Desafíos para resolver el problema	32
1.3.1. Complejidad de mecanismos.....	32
1.3.2. Desigualdades en el acceso a medicamentos	33
1.3.3. Limitaciones de los tratamientos actuales	35
1.3.4. Problemas éticos y regulatorios	36
1.3.5. Integración de tecnologías y ciencias de datos en la medicina.....	38
1.4. Diseño de fármacos	38
1.4.1. Diseño racional de fármacos.....	38
1.4.2. Química computacional	39
1.4.2.1. Orígenes de la química computacional.....	39
1.4.2.2. Desarrollo de la computación de alto rendimiento	41
1.4.2.3. Machine Learning: Desafíos de la química computacional.....	46
II - JUSTIFICACIÓN	49
III - OBJETIVOS	55
IV - MATERIAL Y MÉTODO	59
4.1. High Performance computing	59
4.2. Modelado molecular	61
4.2.1. Preparación de Estructuras en Formato mol2.....	61

4.2.2.	Preparación de Estructuras en Formato pdbqt	62
4.2.3.	Preparación de Estructuras en Formato pmz/pml	63
4.2.4.	Librerías de compuestos.....	63
4.3.	Predicción de estructuras de proteínas.....	64
4.4.	Cribado virtual.....	65
4.5.	Técnicas de cribado virtual basadas en ligandos	66
4.5.1.	Modelado farmacofórico	67
4.5.2.	Modelado basado en fingerprints.....	70
4.6.	Técnicas de cribado virtual basadas en la estructura	71
4.6.1.	Docking Molecular.....	71
4.6.1.1.	AutoDock Vina	72
4.6.1.2.	Lead Finder	72
4.6.1.3.	Docking Sucesivos.....	73
4.6.1.4.	Docking Consenso.....	74
4.6.1.5.	Blind Docking	74
4.6.2.	Dinámica Molecular.....	75
4.6.2.1.	MMGBSA	76
4.6.2.2.	PSN.....	76
4.7.	Machine learning	76
4.8.	Experimentación.....	77
V -	RESULTADOS Y DISCUSIÓN.....	81
5.1.	Implementación de software para dinámica molecular.....	81
5.1.1.	Script Lanzador	82
5.1.2.	Script Creador de Complejos.....	83
5.1.3.	Script Elongador de Dinámicas.....	83
5.1.4.	Script de Unión de Dinámicas.....	84

5.1.5.	Script de Análisis de Dinámicas.....	84
5.1.6.	Casos de Estudio	85
5.1.6.1.	Virtual Screening.....	86
5.1.6.2.	Inverse Virtual Screening.....	87
5.1.6.3.	Blind Docking combinado con simulaciones de Dinámica Molecular	88
5.2.	Cáncer.....	90
5.2.1.	Casos Especies Reactivas de Oxígeno	90
5.2.2.	Quimioresistencia a tumores	94
5.2.3.	Metástasis y angiogénesis	100
5.3.	Alzheimer	103
5.3.1.	Acetilcolinesterasa.....	104
5.3.2.	Butirilcolinesterasa.....	107
5.3.3.	Monooxigenasas.....	110
5.4.	Retinitis pigmentosa.....	112
5.5.	Cicatrización.....	115
5.6.	Diabetes	117
5.7.	Resistencia a los antibióticos	120
5.8.	Otros	123
5.8.1.	Valoración del uso de herramientas de ML en el ámbito de la trombosis 123	
5.8.2.	Desarrollo de MetaScreener.....	129
VI -	CONCLUSIONES	133
VII -	LIMITACIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN.....	137
VIII -	REFERENCIAS BIBLIOGRÁFICAS	141
IX -	ANEXOS	167

SIGLAS Y ABREVIATURAS

- AChE**, Acetilcolinesterasa
AF, AlphaFold
AV, AutoDock Vina
BChE, Butirilcolinesterasa
BD, *Blind Docking*
CAR, Computación de Alto Rendimiento
DM, Dinámica Molecular
FEL, *Free Energy Landscape*
GUI, Interfaz Gráfica de Usuario
HPC, *High Performance Computing*
HSA, Seroalbúmina Humana
IDO1, Indolamina 2,3-dioxigenasa 1
KMO, Quinurenina 3-monooxigenasa
LF, Lead Finder
ML, *Machine Learning*
MM-GBSA, *Molecular Mechanics-Generalized Born Surface Area*
PDB, Protein Data Bank
POPC, *1-palmitoyl-2-oleoyl-glycero-3-phosphocholine*
PSN, *Protein Structure Network*
QSAR, *Quantitative Structure-Activity Relationship*
RMSD, *Root Mean Square Deviation*
RMSF, *Root Mean Square Fluctuation*
ROS, Especies Reactivas de Oxígeno
XO, Xantina Oxidasa

ÍNDICE DE FIGURAS, DE TABLAS Y DE ANEXOS**ÍNDICE DE FIGURAS**

Figura I.1. Fases del descubrimiento y desarrollo del descubrimiento de fármacos.	30
Figura I.2. Porcentaje de fármacos aprobados por la FDA en el año 2023 según la patología.	37
Figura I.3. Diagrama de flujo del diseño racional de fármacos.	39
Figura I.4. Modelo llave cerradura de Fischer.	43
Figura I.5. Resultado obtenido tras lanzar un cálculo de Virtual Screening basado en modelos farmacofóricos realizado por el software LigandScout.	45
Figura IV.1 Diagrama de trabajo en HPC.	60
Figura IV.2. Diagrama de flujo de métodos de diseño de fármacos.	66
Figura IV.3 Ejemplo de modelo farmacofórico obtenido por Ligand Scout.	68
Figura IV.4. Diagrama de flujo de cribado basado en modelado farmacofórico utilizando un Virtual Screening consenso.	70
Figura IV.5. Representación en MACCS fingerprints del paracetamol.	71
Figura IV.6. Esquema de cálculos de Dockings sucesivos.	73
Figura V.1. Blind Docking de la neobavaisoflavona contra la Xantina Oxidasa.	92
Figura V.2. Blind Docking de los ácidos del café contra la Xantina Oxidasa.	94
Figura V.3. Blind Docking de la 4,7-dicloroquinona contra la indolamina 2,3-dioxigenasa 1.	95
Figura V.4. Dockings sucesivos de los derivados monosustituidos del ácido cafeoilquínico contra la Topoisomerasa II α	99
Figura V.5. Dockings sucesivos de los derivados disustituidos del ácido cafeoilquínico contra la Topoisomerasa II α	99

Figura V.6. Evolución de RMSD de los compuestos de interés frente a la lectina CLEC-2.....	101
Figura V.7. Docking de los curcuminoides frente a VEGFR.	102
Figura V.8. Docking sucesivo de la lactucina y la lactupicrina frente a la AChE.	105
Figura V.9. Docking sucesivo de la lactupicrina y la lactucina frente a la AChE.	105
Figura V.10. Dockings sucesivos de los derivados monosustituidos del ácido cafeoilquínico frente a la BChE.	108
Figura V.11. Dockings sucesivos de los derivados monosustituidos del ácido cafeoilquínico frente a la BChE.	108
Figura V.12. Dockings sucesivos de los derivados disustituidos del ácido cafeoilquínico frente a la BChE.	109
Figura V.13. Dockings sucesivos de los derivados disustituidos del ácido cafeoilquínico frente a la BChE.	109
Figura V.14. Evolución de RMSD de los compuestos de interés frente a la KMO.	111
Figura V.15. Evolución de RMSF durante las simulaciones de la rodopsina.	113
Figura V.16. Imágenes de la retina de ratón tomada mediante tomografía de coherencia óptica (OCT).....	114
Figura V.17. Evolución de RMSD de los compuestos de interés frente al receptor de EP3.....	117
Figura V.18. Diagrama de RMSD de cada proteína de interés durante la simulación con el carvacrol.....	120
Figura V.19. Docking de los ligandos de interés frente a la histona deacetilasa 3.	122
Figura V.20. Diagrama de RMSD de la antitrombina.	129

ÍNDICE DE TABLAS

Tabla I.1. Prevalencia de enfermedades crónicas en el mundo según la OMS (Organización Mundial de la Salud).	31
Tabla I.2. Prevalencia de enfermedades transmitidas por vectores según la OPS (Organización Panamericana de la Salud).....	34
Tabla V.1. Valores máximos de RMSD y RMSF obtenidos tras el análisis del Virtual Screening.	87
Tabla V.2. Valores máximos de RMSD y RMSF obtenidos tras el análisis del Inverse Virtual Screening.....	88
Tabla V.3. Valores máximos de RMSD y RMSF obtenidos tras los cálculos de Blind Docking y la simulación de dinámica molecular.....	89
Tabla V.4. Resumen de las interacciones entre los derivados del ácido cafeoilquínico y la Topoisomerasa II α	98
Tabla V.5. Resumen de las interacciones entre los curcuminoides y VEGFR2.....	103
Tabla V.7. Resumen de las interacciones entre la lactupicrina y la lactucina contra la AChE.....	106
Tabla V.8. Resultados obtenidos por la herramienta DIA-DB del carvacrol frente a distintas proteínas implicadas en diabetes.....	119
Tabla V.9. Tabla comparativa de la estructura secundaria de las estructuras de antitrombina.....	125
Tabla V.10. Métrica RMSD de las estructuras predichas por AF en comparación con los cristales obtenidos de la PDB.....	126
Tabla V.11. Tabla con las diferentes métricas de las mutaciones M1-M3 predichas por AF2.	126
Tabla V.12. Métrica RMSD obtenidas tras el análisis FEL.....	127
Tabla V.13. Métrica IDDT obtenidas tras el análisis FEL.....	128

ÍNDICE DE ANEXOS

ANEXO 1. Producción científica y calidad de las revistas 167
ANEXO 2. Participaciones en Congresos 173
ANEXO 3. Software y licencias 177

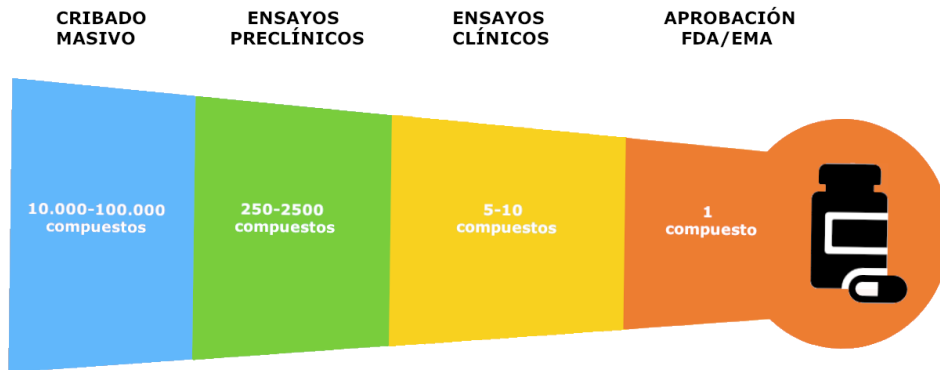
I – INTRODUCCIÓN

I - INTRODUCCIÓN

1.1. DESCUBRIMIENTO DE FÁRMACOS

A pesar de los avances médicos y el aumento de las condiciones de salubridad e higiene, siguen existiendo enfermedades que afectan de forma significativa a la calidad de vida. Las más perjudiciales en este contexto son las enfermedades crónicas, como la diabetes o el cáncer (1). Estas afecciones se agravan con la senectud, dificultando la vida de los enfermos y de sus familiares y cuidadores (2). Además de afectar a la calidad de vida, estas patologías reducen la esperanza de vida en países desarrollados, donde las principales causas de muerte son el cáncer, accidentes cerebrovasculares, trastornos cardíacos y respiratorios (3,4). Esta problemática se encuentra agravada en países en desarrollo debido a que además de las patologías mencionadas previamente existen diversas enfermedades infecciosas no erradicadas, como el síndrome de la inmunodeficiencia adquirida (SIDA), la tuberculosis o la malaria(5).

Estas afecciones no solo deterioran el bienestar de los afectados, sino que también generan un elevado coste económico. Por un lado, la extensa duración de las consultas médicas y, por otro, las ausencias laborales implican un desembolso financiero significativo al necesitar sustituir a los empleados ausentes (6). Asimismo, los propios estados gastan más dinero al tener los sistemas de salud sobrecargados y teniendo que hacer grandes inversiones en salud pública. Durante los últimos años, se ha visto cómo la pandemia provocada por el SARS-CoV-2 provocó problemas de desabastecimiento en los productos básicos, pérdidas de puestos de trabajo; teniendo más impacto en el sector de la hostelería y complicaciones en la educación, desde el nivel primario hasta el superior, debido al cierre de colegios, institutos y universidades (7). Es importante subrayar que esta crisis sanitaria exacerbó los problemas previamente señalados tales como la saturación de los servicios médicos.

Figura I.1. Fases del descubrimiento y desarrollo del descubrimiento de fármacos.

1.2. ENFERMEDADES ACTUALES

Entre las numerosas enfermedades presentes en la actualidad, se encuentran muchas que carecen de tratamiento y presentan una alta tasa de mortalidad. Una gran proporción de estas enfermedades son de origen vírico, como el síndrome respiratorio agudo grave (SARS), síndrome respiratorio de Oriente Medio (MERS), el dengue y el ébola (8–10). Sin embargo, también existen enfermedades no infecciosas que necesitan una mayor investigación debido a su naturaleza multifactorial y a la desconocida etiología de alguna de ellas. Entre las enfermedades no infecciosas más relevantes se encuentran la enfermedad de Alzheimer, la enfermedad de Parkinson, las patologías cardíacas y las enfermedades autoinmunes, además de condiciones sistémicas como la obesidad (11). El desafío principal con estas afecciones radica en la ausencia de fármacos efectivos; en muchos casos, aunque se conoce el mecanismo o la proteína implicada en la patología, no existe un tratamiento que la cure completamente. Otro obstáculo es la aparición de complicaciones durante los ensayos clínicos, como la toxicidad o la insuficiente absorción del fármaco por el organismo. Por consiguiente, es imperativo avanzar en la mejora de los medicamentos existentes y en la búsqueda de nuevas terapias para enfermedades que aún no cuentan con un tratamiento adecuado (12).

Asimismo, se ha registrado un incremento en el número de pacientes con enfermedades crónicas que, aunque cuentan con tratamientos establecidos, han experimentado un aumento reciente de casos; ejemplos notables incluyen el cáncer, la diabetes y la obesidad (13). Si bien estas afecciones no son de origen infeccioso, se ha detectado cierta correlación entre infecciones virales y enfermedades crónicas como el cáncer y la esclerosis múltiple (14,15). Estas patologías requieren tratamientos prolongados o incluso vitalicios, por lo que es esencial desarrollar medicamentos altamente efectivos y con mínimos efectos secundarios. El crecimiento en la prevalencia de las enfermedades crónicas está relacionado principalmente al incremento de la esperanza de vida, debido a que condiciones como el cáncer y los trastornos neurodegenerativos suelen asociarse al envejecimiento (16,17). Además, factores como un mayor sedentarismo y una alimentación deficiente en la sociedad contemporánea incrementan los riesgos de padecimientos como la obesidad y exacerban las complicaciones de enfermedades como la diabetes(18). Por último, el aumento de factores estresantes y la contaminación, especialmente en los países desarrollados, han contribuido al incremento de afecciones como el cáncer, el asma, las alergias y la obesidad (19,20).

Tabla I.1. Prevalencia de enfermedades crónicas en el mundo según la OMS (Organización Mundial de la Salud).

Enfermedad	Prevalencia según la OMS
Diabetes	422 millones de casos en 2014
Cáncer	10 millones de fallecidos en 2020
Enfermedades Cardiovasculares	17,9 millones de fallecidos cada año
Alzheimer	6-7 millones de nuevos diagnósticos cada año
Parkinson	8.5 millones de casos en 2019
Obesidad	890 millones de casos en 2022

Datos extraídos de la web oficial de la OMS.

Junto a las enfermedades crónicas, se ha observado un incremento en la incidencia de enfermedades que durante el siglo anterior habían experimentado una notable disminución. Este fenómeno se atribuye principalmente a la insuficiente cobertura de vacunación contra ciertos patógenos que ya estaban prácticamente erradicados, como el virus del sarampión (21). También el uso indebido de los antibióticos ha resultado en el incremento de microorganismos multirresistentes como el *Staphylococcus aureus* resistente a la meticilina (MRSA) (22). Además, se debe considerar que la proliferación de estos microorganismos esté vinculada al incremento de antibióticos encontrados en la carne de ganado y en el agua, resultado de la contaminación por antibióticos de uso veterinario (23,24). Por tanto, es crucial identificar nuevas fuentes de antibióticos que puedan reemplazar o complementar los medicamentos ya existentes. Un ejemplo destacado es la combinación de ácido clavulánico con amoxicilina, debido a que muchos microorganismos son resistentes a los betalactámicos y esta resistencia se neutraliza con el ácido clavulánico. Sin embargo, se ha observado un incremento de la resistencia al propio ácido clavulánico (25).

1.3. DESAFÍOS PARA RESOLVER EL PROBLEMA

1.3.1. Complejidad de mecanismos

Uno de los principales problemas a la hora de encontrar nuevos medicamentos se debe a que hay muchas enfermedades que se producen en rutas de señalización complejas. Estas rutas tienen múltiples puntos de regulación para evitar descompensaciones dado que son rutas clave para el funcionamiento del organismo (26,27). Es importante destacar que la regulación de la velocidad a la que funcionan estas rutas suele estar definida por mecanismos de regulación alostéricos, teniendo muchas veces el producto final o intermediarios como reguladores inhibidores de proteínas al inicio de la ruta (28). Esto sumado al hecho de que hay muchas patologías multisistémicas o, dentro de las patologías infecciosas, hay microorganismos que pueden infectar diferentes tipos celulares (29). Este problema se ve agravado por la falta de caracterización de algunas proteínas, sobre todo las que se encuentran en células del sistema nervioso. Estas proteínas son clave para el tratamiento de enfermedades como el Parkinson, la

enfermedad de Alzheimer o la esclerosis lateral amiotrófica (ELA) (30). Esta problemática no solamente se da en cuestiones médicas sino que también afecta gravemente al ámbito de la ganadería y la agricultura, dado que hay patógenos que pueden asolar granjas y cultivos (31,32). Por lo que se deben seguir investigando la función y regulación de múltiples proteínas en diferentes campos.

1.3.2. Desigualdades en el acceso a medicamentos

Otro de los principales problemas que hay es el hecho de que muchas enfermedades que tienen una alta incidencia en los países en vías de desarrollo se debe a que el coste de los medicamentos es demasiado caro, por lo que es difícil que los enfermos que viven en estos países tengan acceso a estas medicinas (33). Esto sumado al hecho de que en regiones en vías de desarrollo siguen existiendo enfermedades ya erradicadas en los países desarrollados (cólera o malaria) debido a la falta de higiene y salubridad (falta de agua potable y condiciones sanitarias deficientes). Asimismo, la falta de tratamiento adecuado en los alimentos para evitar la presencia de patógenos hace que sean más comunes las parasitosis, como la enfermedad de Chagas en Sudamérica (36).

Dichas enfermedades están desatendidas debido a que investigar su tratamiento requiere de una gran inversión de tiempo y recursos que está fuera del alcance de los países en vías de desarrollo y, en la mayoría de casos, los países desarrollados se niegan a financiar (37). La falta de inversión económica se debe a que la cantidad de dinero que pueden ofrecer los enfermos de estas enfermedades es menor a la cantidad de dinero invertida en el desarrollo, síntesis y purificación de los fármacos. Esto, sumado al hecho de que muchas veces se necesita una investigación para desarrollar técnicas y protocolos que permitan diagnosticar la patología que tienen cada uno de los enfermos, hace que sea aún más difícil encontrar cura para estas infecciones (38).

Tabla 1.2. Prevalencia de enfermedades transmitidas por vectores según la OPS (Organización Panamericana de la Salud).

Enfermedad	Prevalencia según la OPS	Vector
Dengue	2,3 millones de casos en 2013 (> 1000 fallecidos)	Mosquito
Chagas	30. 000 infectados al año y 12.000 fallecidos	Vinchuca/Chinche
Leishmaniasis	12 millones de infectados	Mosquito
Fiebre Amarilla	200. 000 infectados al año y 30.000 fallecidos	Mosquito
Malaria	480 000 infectados en 2022	Mosquito

Datos extraídos de la web oficial de la OPS.

Al igual que ocurre con las enfermedades de los países en vías de desarrollo, existen muchas enfermedades raras, aquellas que se presentan en menos de 5 pacientes por cada 10 mil habitantes, cuyo tratamiento e investigación no se realiza debido a la falta de financiación por el bajo número de personas que padecen estas patologías. Los medicamentos que se desarrollan para tratar de curar estas afecciones son los llamados medicamentos huérfanos (39). La investigación y el desarrollo de los medicamentos huérfanos suelen estar subvencionados por organizaciones gubernamentales de forma que se incentiva a la industria farmacéutica a su desarrollo. Un ejemplo de medicamento huérfano es el ivacaftor, el cual se usa para el tratamiento de la fibrosis quística, este medicamento no cura completamente la enfermedad pero ayuda a paliar los síntomas de mayor gravedad (40,41).

1.3.3. Limitaciones de los tratamientos actuales

Es necesario tener en cuenta que no solamente hay que investigar las enfermedades que carecen de tratamiento en la actualidad, sino también hay que mejorar los actuales debido a que muchas veces su eficacia es limitada. Estas medicinas tienen efectos secundarios adversos por ser poco específicos o requerir una vía de administración compleja, como por ejemplo intravenosa o intramuscular (42). Por todo ello, es necesario seguir mejorando los fármacos actuales para por un lado mejorar la eficacia del compuesto original; un ejemplo es la amoxicilina la cual es un antibiótico con una efectividad más amplia que el compuesto original, la penicilina (43). Respecto a la reducción de efectos adversos, un claro ejemplo es el uso de lidocaína como anestésico local, eliminando efectos adversos como los que producía la cocaína (44,45). Por último, se están desarrollando alternativas a las inyecciones intramusculares e intravenosas mediante medicamentos en formato de píldoras, que facilitan la toma de la medicación por parte de los pacientes de forma autónoma (46).

Sin embargo, muchas veces este tipo de investigaciones no se llevan a cabo debido al alto coste que requiere descubrir, sintetizar, purificar y testar los compuestos producidos por la industria farmacéutica. La investigación desde que se identifica la proteína que produce la enfermedad hasta que se obtiene un compuesto válido que pueda ser probado en ensayos clínicos requiere de años y una gran inversión de dinero (47). Los procesos que más tiempo y dinero cuestan involucran la síntesis y el proceso de purificación química de los compuestos, así como los ensayos *in vitro* e *in vivo* siendo necesario además un tedioso proceso burocrático hasta obtener el compuesto final (48,49).

La última etapa en el descubrimiento de fármacos son los ensayos clínicos, los cuales incluyen varias fases y procedimientos para controlar toxicidad, efectividad y dosificación. Este proceso requiere de un control riguroso para asegurarse que los estudios clínicos son fiables y eliminan compuestos tóxicos que puedan ser perjudiciales para la salud de los participantes de estos ensayos (50,51). Además, es necesario una gran inversión económica para la realización de los ensayos clínicos que pueden llevar años para obtener un compuesto fiable en el

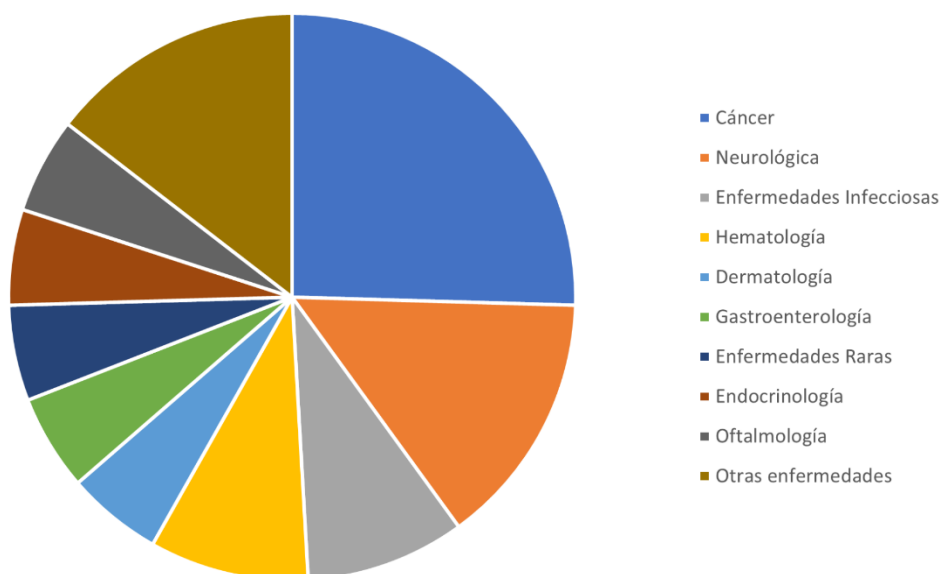
mercado farmacéutico. Hay que destacar que incluso después de salir al mercado estos compuestos siguen siendo vigilados para controlar posibles efectos secundarios adversos (52).

1.3.4. Problemas éticos y regulatorios

Hay que tener en cuenta que todos estos compuestos suelen ser testados en animales por lo que es necesario tener unos protocolos adecuados para tratar con el máximo respeto y minimizando el sufrimiento del animal. Estos animales son criados y cuidados en un ambiente específico que suele ser lo más aséptico posible para evitar cualquier tipo de interferencia con el medio. Todo ello además encarece considerablemente estas pruebas (53). En cuanto a la parte ética y regulatoria, el uso de animales en investigación se basa en la regla de las “tres Rs”: reemplazo, reducción y refinamiento (54). El reemplazo dicta que siempre que sea posible utilizar otra técnica como métodos computacionales o técnicas *in vitro*, en lugar de usar animales se ha de utilizar esta técnica alternativa (55). Y dentro del uso de animales, intentar usar animales sin sistemas nerviosos complejos como los invertebrados en preferencia sobre los vertebrados. La reducción se basa en la premisa de que hay que planificar el diseño de los experimentos con animales para poder obtener la mayor cantidad de información con el número mínimo de animales (56). La última regla es la del refinamiento, la cual tiene que regular los procesos a los que son sometidos los animales para minimizar el dolor, sufrimiento y estrés intentando utilizar siempre técnicas menos invasivas para los animales (57,58).

Figura I.2. Porcentaje de fármacos aprobados por la FDA en el año 2023 según la patología. El total de compuestos aprobados fueron 55.

Fármacos Aprobados FDA en 2023



Pero el uso de fármacos no solo está regulado por la forma de testar estos compuestos en los animales, también es necesario que haya organizaciones que regulen su uso y administración en seres humanos. Una de las más conocidas es la Administración de Alimentos y Medicamentos de EE.UU., conocida como FDA (<https://www.fda.gov/>) por sus siglas en inglés, actualmente hay más de 4000 compuestos aprobados por la FDA para ser utilizados como tratamiento de diversas patologías (59). Además de la FDA en la Unión Europea existe la Agencia Europea del Medicamento (conocida como EMA, <https://www.ema.europa.eu/>). Este organismo es el responsable de la validación de medicamentos dentro de Europa, en España actualmente hay más de 15.000 medicamentos autorizados para su uso en medicina (60).

1.3.5. Integración de tecnologías y ciencias de datos en la medicina

En los últimos años se ha producido un gran avance en el ámbito de la tecnología que ha permitido obtener muchísima información y procesarla de una manera rápida y efectiva; con los avances en *Machine Learning* se han encontrado relaciones entre variables y actividad de compuestos que no pueden ser deducidas a simple vista. Una de las aplicaciones del *Machine Learning* es el descubrimiento de nuevos biomarcadores que permitan detectar de forma temprana enfermedades como el cáncer, actualmente se hacen estudios para prevenir y detectar estadios precoces de cáncer como por ejemplo el cáncer de próstata, detectando niveles elevados del antígeno prostático específico (PSA) (61,62). Este tipo de estrategias junto con la gran cantidad de datos que hay en la actualidad permite descubrir nuevos biomarcadores que permitan detectar precozmente las enfermedades y así tratarlas antes de que tengan mayor gravedad y mortalidad (63,64).

Gracias al avance de las nuevas tecnologías y el descubrimiento de estos biomarcadores se ha empezado a utilizar el concepto de medicina de precisión, en la cual se tiene en cuenta tanto la genética como las condiciones ambientales individuales de cada paciente para hacer una terapia individualizada y con una mayor tasa de éxito (65). Estas técnicas consiguen mejorar la precisión en el diagnóstico y tratamiento de las patologías padecidas por cada paciente, reduciendo a su vez las pruebas realizadas y los efectos secundarios de la medicación, permitiendo de esta manera tener una terapia individual y más adecuada para cada paciente (66).

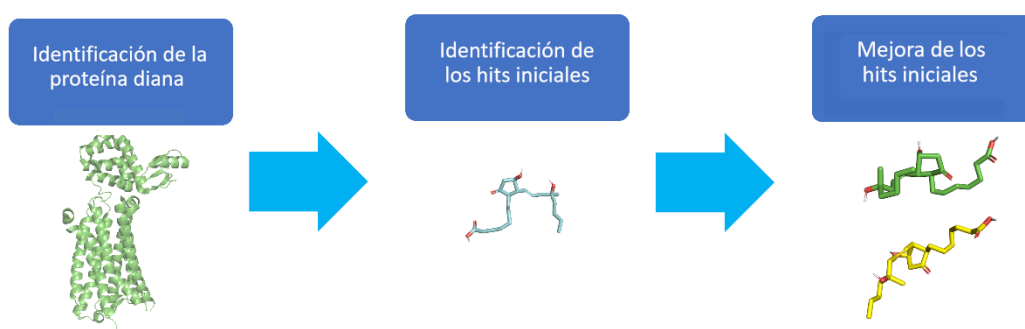
1.4. DISEÑO DE FÁRMACOS

1.4.1. Diseño racional de fármacos

Para tratar todas estas patologías se están diseñando compuestos siguiendo un esquema lógico y con una metodología basada en las propiedades bioquímicas de la proteína cuya actividad se quiere modular. De esta manera se pueden obtener fármacos que actúen como ligandos específicos no solamente de una proteína, sino también específicos de una región concreta de dicha proteína (67). Uno de los

modelos más utilizados en el diseño racional de fármacos es la teoría de la llave y cerradura de Fischer (68). Este modelo sugiere que la especificidad entre sustrato y enzima debe ser alta para que se produzca la reacción, de la misma manera que existe una relación única entre una llave (sustrato) y una cerradura (enzima). Esta teoría permite optimizar la búsqueda de moduladores de diferentes enzimas. Una forma de aplicar este tipo de metodología es la siguiente: *i*) Identificación de la proteína diana, *ii*) Hacer un cribado masivo de una librería de compuestos con una gran variedad química, ya sea mediante técnicas de laboratorio o técnicas computacionales, *iii*) Obtención de un compuesto líder con una actividad moduladora, *iv*) Diseño de nuevos compuestos basados en la optimización de la molécula líder, y *v*) Realización de nuevos cribados y optimizaciones hasta obtener un compuesto válido para ser llevado a los ensayos clínicos. Una de las principales ventajas de esta metodología es el ahorro de tiempo y dinero debido a que con este enfoque se optimiza enormemente el desarrollo de fármacos (69,70).

Figura I.3. Diagrama de flujo del diseño racional de fármacos.



1.4.2. Química computacional

1.4.2.1. Orígenes de la química computacional

Entre los años 1950 y 1970 se empezaron a utilizar ordenadores para aplicar los principios de la mecánica cuántica a pequeños sistemas químicos (71). Estos cálculos, aunque eran bastante simples y tenían muchas limitaciones permitieron establecer una base del uso de la computación para el desarrollo de fármacos, la

computación sigue siendo clave en el diseño de medicamentos en la actualidad. En esta época se desarrollaron modelos como la aproximación de Hartree-Fock (72) o el inicio de la teoría del funcional de la densidad (DFT), esta técnica se continuó desarrollando mostrando su eficacia en los metales conductores y semiconductores, y por último en las décadas posteriores, se empezó a utilizar para calcular las estructuras moleculares y las energías de reacción (73,74).

En la década de los 70, la química computacional experimentó un gran desarrollo debido a la mejora en los microprocesadores y el aumento de la memoria RAM de los ordenadores; y a la mayor accesibilidad gracias a que la industria informática (Apple, Commodore, etc.) desarrolló equipos informáticos que fueron utilizados por la sociedad científica. No solo se incrementó la accesibilidad, sino que también se empezó a modelar moléculas más complejas e incluso reacciones químicas (75), lo que permitió un gran desarrollo interdisciplinar entre biólogos y químicos con científicos de la computación. En esta década, algunos de los programas más utilizados por la comunidad científica fueron Gaussian (76), MOPAC (77), GAMESS (78) y AMBER (79). Gaussian y AMBER son dos software que permiten realizar cálculos de mecánica cuántica y molecular de biomoléculas y, a pesar de haber sido desarrollados durante la década de los años 70, se siguen utilizando hoy en día (aunque las versiones actuales permiten hacer cálculos más complejos y con mayor eficiencia) (80,81). Otro avance importante fue la creación de una base de datos con información sobre estructuras en 3D de proteínas y sus ligandos obtenidas por cristalografía de rayos X o resonancia magnética nuclear (RMN), esta base de datos se llama Protein Data Bank (PDB, <https://www.rcsb.org/>) (82).

Durante los últimos años del siglo XX (década de los 80 y 90), se produjo una consolidación de las aplicaciones y fundamentos de la química computacional. La Teoría del Funcional de la Densidad (DFT: *Density Functional Theory*), desarrollada en la década de los años 60, adquirió una importancia clave junto con el desarrollo de los métodos *ab initio* que, aunque necesitaban un gran coste computacional, eran mucho más precisos debido a que permitía manejar estructuras con cientos de átomos, lo que permitió lograr grandes avances en las áreas de la reactividad química y la espectroscopía (83). Otra de las técnicas que más auge tuvo durante

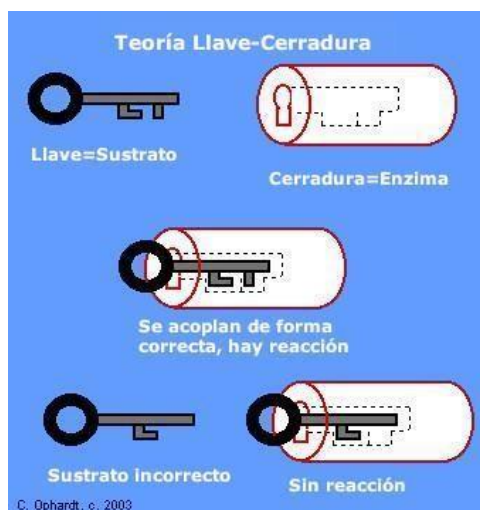
esta etapa fue la simulación de dinámica molecular, la cual permite simular la evolución del sistema proteína-ligando en el medio biológico durante un periodo de tiempo corto (menor al milisegundo). Esta simulación permite evaluar el potencial modulador del ligando para la proteína de interés. Para ello, una vez ha finalizado la simulación se analizan las interacciones producidas entre ambos y los cambios conformacionales producidos en ambas estructuras. Esta técnica, basada en las leyes de Newton, permite obtener información crucial sobre los mecanismos de acción enzimática, transporte de moléculas y el plegamiento proteico (84,85). Todas estas técnicas tuvieron una importancia clave en ámbito farmacéutico dado que permitió mejorar el diseño de fármacos, incrementando de forma considerable el método del diseño racional de fármacos, debido a que gracias a ellas se produjo un incremento en la precisión de las propiedades fisicoquímicas de los compuestos a testar sin necesidad de sintetizar y purificar dichos compuestos. De esta manera se logró un ahorro considerable de tiempo y costes económicos (86).

1.4.2.2. *Desarrollo de la computación de alto rendimiento*

Al final del siglo XX aumentó de forma considerable la capacidad de los supercomputadores para realizar cálculos. Esto permitió el uso simultáneo de ordenadores de gran capacidad creando los llamados clústeres de supercomputación (*High Performance Computing: HPC*), por lo que desde entonces se han podido realizar cálculos en paralelo de forma masiva. Esto ha provocado que sea posible realizar cálculos contra librerías de compuestos de millones de fármacos en un tiempo viable, horas o, como máximo, días (87). Por lo que ahora es posible no solo encontrar compuestos ya existentes en la naturaleza, sino también crear librerías virtuales con compuestos que únicamente se encuentran en dicha librería virtual para poder ser sintetizados si los ensayos computacionales sugieren que es un buen candidato para su testeo en el laboratorio. Además, las unidades de procesamiento gráfico (GPU), que se han adaptado para ser utilizadas en cálculos científicos, en lugar de su uso original en la industria de los videojuegos. Estas GPUs permiten acelerar cálculos complejos como las simulaciones de dinámica molecular o ser utilizadas en el campo del *Machine Learning* para hacer cálculos con una gran rapidez (88,89).

Las técnicas computacionales actuales se dividen principalmente en dos grandes bloques: las basadas en la estructura de las proteínas y las basadas en los ligandos (moléculas pequeñas, que modulan la actividad de la proteína), ambas están muy relacionadas con el modelo llave-cerradura de Fischer (68). En el primer bloque se encuentran técnicas como el *Docking* molecular, las simulaciones de dinámica molecular y el modelado por homología (90). El modelado por homología es una técnica utilizada para predecir la estructura tridimensional de proteínas que no han podido caracterizarse completamente, debido al alto coste computacional que requiere obtener la estructura proteica de forma precisa. Además, existen proteínas desordenadas o proteínas cuya cristalización no es posible, lo que dificulta su caracterización. Para ello, este tipo de programas comparan la secuencia de aminoácidos de la proteína diana con la de proteínas similares cuya estructura sí es conocida y con todas estas estructuras es capaz de inferir la estructura final de la proteína diana (91,92). Los software más utilizados son MODELLER (93), SWISS-MODEL (92) y ROSETTA (94), estos software se pueden utilizar tanto para proteínas cuya estructura no se conoce completamente como para predecir cambios conformacionales provocados por una mutación de una estructura ya conocida. Por último, desde la primera década del siglo XXI se han utilizado herramientas de *Machine Learning* que permiten predecir la conformación tridimensional de la proteínas partiendo únicamente de la secuencia aminoacídica, un ejemplo de programa desarrollado para predecir la estructura de las proteínas es AlphaFold (95).

Figura I.4. Modelo llave cerradura de Fischer. Adaptado de C. Ophardt 2003.

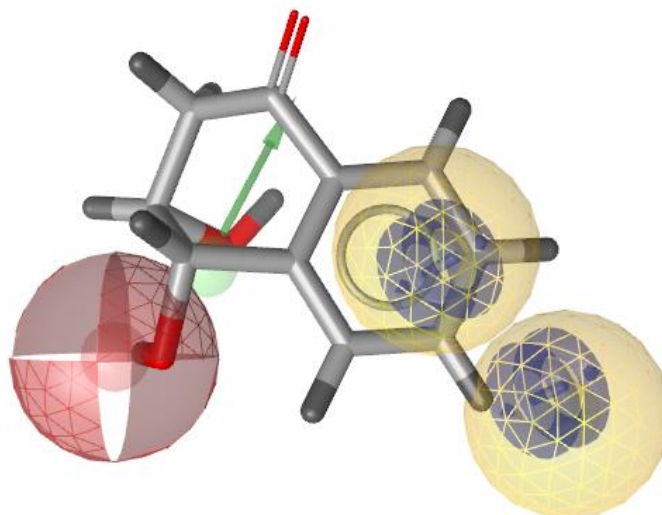


Otra de las técnicas basadas en la estructura es el *Docking* molecular: esta herramienta es capaz de predecir el sitio de unión y la conformación más probable de un ligando frente a una proteína (96). El *Docking* simplifica notablemente las condiciones del medio biológico, utilizando modelos implícitos de solvatación, las macromoléculas son estáticas, y no tiene en cuenta otros componentes del medio como la membrana o los iones. De esta manera, los cálculos de *Docking* se realizan de una forma rápida. Sin embargo, a pesar de ser un modelo simplificado su efectividad es muy alta en la predicción de la pose que adquiere un ligando al interactuar con una determinada proteína, los estudios muestran como los software actuales son capaces de obtener una predicción con una media de RMSD menor a 2.0 Å (97). Los software más utilizados para realizar cálculos de *Docking* molecular son: AutoDock Vina (98), Lead Finder (99), METADOCK (100), Glide (101) and Gnina (102). Además, el *Docking* no tiene porqué ser dirigido, también existe una técnica llamada *Blind Docking* que permite hacer un cribado por toda la superficie de la proteína, obteniendo las zonas de unión más probables, así como la conformación que el ligando adquiere en cada una de las poses (103). Lamentablemente, debido a que el *Docking* parte de un modelo muy simple; necesita ser usado junto con otras técnicas para ser capaz de predecir la actividad moduladora de un compuesto frente a una determinada proteína (104). La técnica más utilizada junto con el *Docking* es la simulación de dinámica molecular; está

técnica basada en las leyes de fuerza de Newton permite predecir la evolución de un sistema proteína-ligando a lo largo de un periodo corto de tiempo, teniendo en cuenta no solo las interacciones del ligando con la proteína sino también las interacciones con el medio (agua e iones) (84). De esta manera, se puede predecir si hay un cambio conformacional en la proteína o el desplazamiento del ligando a lo largo de la simulación. Actualmente hay muchos software de dinámica molecular, los más utilizados son GROMACS (105), Desmond (106), AMBER (107) y CHARMM (108).

Respecto al segundo bloque de técnicas, están las técnicas basadas en ligandos, dentro de las cuales se encuentran: el modelado farmacofórico, los estudios QSAR y los cálculos de similaridad electrostática, tamaño y forma (109,110). El modelado farmacofórico es una técnica que asigna características farmacofóricas como la hidrofobicidad, la capacidad de formar enlaces de hidrógeno o la densidad de carga positiva o negativa que tiene una determinada región de una molécula. El análisis de estas características permite encontrar compuestos similares y descubrir las propiedades farmacofóricas únicas que pueden ser clave para la actividad de un grupo específico de ligandos. Uno de los softwares más utilizados en modelado farmacofórico es Ligand Scout (111). La similaridad electrostática permite analizar los ligandos de forma similar al modelado farmacofórico pero de una manera más simple, centrándose solamente en las características electrostáticas de la molécula, y así comparar una librería de compuestos contra un ligando diana, encontrando fármacos cuyas semejanzas no son evidentes. El programa EON es uno de los software con licencia más utilizados para realizar cálculos de similaridad electrostática (112). Sin embargo, hay otros software como ROCS que permiten hacer cálculos más simples basados únicamente en la forma y el tamaño de los compuestos, de esta manera se encuentran compuestos que puedan acoplarse al sitio de unión de la proteína diana (113).

Figura I.5. Resultado obtenido tras lanzar un cálculo de Virtual Screening basado en modelos farmacofóricos realizado por el software LigandScout.



Matching Features	Pharmacophore-Fit Score	Mol. Index	Active/Decoy	smiles
	47.03	24	active	O=C1C[C@@H]...

Por último, está la técnica QSAR (relaciones cuantitativas estructura-actividad), dicha técnica está basada en descriptores moleculares de cada una de las moléculas. Los descriptores moleculares son parámetros numéricos que cuantifican las propiedades de los ligandos que se analizan. De esta manera se puede estimar que propiedades son las más importantes para la actividad de una librería de compuestos y cuánto aporta cada una de las propiedades calculadas para modular la actividad de dicho compuesto, así como si la característica favorece o no la actividad del ligando. Uno de los programas más utilizados para calcular descriptores QSAR es DRAGON, el cual permite calcular más de 4000 descriptores (114).

1.4.2.3. *Machine Learning: Desafíos de la química computacional*

Además de las técnicas QSAR existen otros descriptores que pueden ser utilizados actualmente como los descriptores circulares o de Morgan. Estos descriptores permiten encontrar patrones utilizando técnicas de *Machine Learning* (115,116). Esta técnica utiliza múltiples algoritmos (el vecino más cercano, *Random Forest* o árboles de decisiones) para obtener información sobre los patrones que existen en un determinado conjunto de datos y ser capaz de predecir propiedades como la actividad de un compuesto frente a una proteína diana de forma completamente automática (117). Entre todas las clases de *Machine Learning* que existen, una de las más utilizadas en el ámbito del descubrimiento de fármacos es el *Machine Learning* supervisado. En este caso, cada uno de los datos utilizados para entrenar y generar el modelo de *Machine Learning* debe incluir una etiqueta con la salida que corresponde cada dato, se puede utilizar tanto para clasificar datos (activos/inactivos, agonista/antagonista) como para predecir un valor numérico (basada en IC50/EC50) (118). Los modelos de *Machine Learning* supervisado se pueden utilizar no solo para predecir actividad moduladora de los ligandos, sino también para predecir otras propiedades como la toxicidad del compuesto, la dosis efectiva o las propiedades ADME (Absorción, Distribución, Metabolismo y Excreción) (119).

Por último, el *Machine Learning* no solamente se utiliza en el diseño de fármacos; también se está utilizando para la predicción de la estructura tridimensional de proteínas con el desarrollo de software como AlphaFold 2 (95). Esta herramienta desarrollada por DeepMind permite predecir la estructura tridimensional de una proteína partiendo únicamente de su secuencia aminoacídica. Esta herramienta ha obtenido resultados significativamente mejores en sus predicciones en comparación con otras técnicas de predicción basadas en la homología de proteínas. En el prestigioso concurso CASP 14 se observó como AlphaFold 2 obtenía un valor sumatorio de Z-score de 244, una puntuación mucho mayor que la obtenida por otras técnicas (la segunda mejor herramienta obtuvo un valor sumatorio de Z-score de 90) (120).

II – JUSTIFICACIÓN

II - JUSTIFICACIÓN

En la actualidad, existen un gran número de patologías que necesitan ser tratadas, cuyo tratamiento todavía no es efectivo o produce demasiados efectos secundarios. Por ello, es necesario continuar con las investigaciones en la industria farmacéutica y descubriendo nuevos medicamentos que permitan tratar estas enfermedades. Esta investigación es necesaria debido a: *i)* el aumento de la esperanza de vida, lo cual ha propiciado un incremento en el número de pacientes con enfermedades crónicas, *ii)* la existencia de múltiples enfermedades neurodegenerativas que carecen aún de tratamiento, *iii)* el aumento de microorganismos resistentes a antibióticos, *iv)* las pandemias y epidemias víricas cuyo tratamiento no existe o no ha llegado a todos los pacientes, y *v)* la falta de caracterización de muchas proteínas hace que sea imposible encontrar un tratamiento contra enfermedades producidas debido a estas proteínas.

Para poder encontrar solución a estos problemas es importante diseñar una metodología que permita identificar la proteína causante de la patología y obtener moléculas candidatas a modular su actividad, de esta forma se puede tratar la patología de interés. Los métodos actuales se basan en el diseño racional de fármacos de manera que tras identificar la proteína se realizan rondas sucesivas de cribado masivo, ya sea virtual o en el laboratorio. De esta forma se produce un incremento de la precisión en cada una de las etapas de cribado. Para ello, se seleccionan los mejores candidatos en cada paso y se optimizan los compuestos predichos, obteniendo finalmente un compuesto candidato cuya actividad potencial contra la proteína es suficiente para poder ser evaluado en los ensayos clínicos.

La química computacional es clave en esta etapa de selección de candidatos. Debido a los grandes avances obtenidos en los últimos años en el área de la computación de altas prestaciones, se pueden realizar múltiples cálculos de cribado virtual en paralelo y evaluar así miles de compuestos candidatos de una manera más rápida y barata que utilizando de forma exclusiva técnicas *in vitro* e *in vivo*. Sin

embargo, la precisión de las técnicas *in silico* sigue siendo inferior a las de las mencionadas anteriormente, por lo que sigue siendo necesario seguir testando los compuestos candidatos en laboratorios.

Las técnicas computacionales más clásicas suelen ser: el acoplamiento molecular, conocido como *Molecular Docking* en inglés, que permite predecir la interacción inicial entre una proteína y un ligando determinado siendo capaz de calcular la pose tridimensional que adquiere el ligando y dando una estimación relativa de la energía de unión. Además, existen técnicas basadas en los ligandos que permiten encontrar ligandos con mayor afinidad a la proteína que el ligando natural. Una de las técnicas *in silico* más utilizadas es el modelado farmacofórico, basado en la identificación de las características farmacofóricas (hidrofobicidad, carga positiva o negativa, etc.) asociadas a los grupos funcionales presentes en los compuestos a analizar, de esta forma se puede comparar los distintos modelos farmacofóricos obtenidos del ligando de interés contra las de una librería de ligandos. Además de la comparación de grupos farmacofóricos, hay otras técnicas que permiten realizar cálculos de similaridad electrostática y/o cálculos basados en la forma y tamaño del ligando original. Todas estas técnicas se pueden utilizar tanto de forma individual como de forma conjunta para intentar obtener un compuesto que sea modulador adecuado de la proteína de interés.

Sin embargo, algunas proteínas presentan una gran flexibilidad, lo que requiere técnicas con mayor precisión y consecuentemente con un mayor coste computacional que el *Docking* o el modelado farmacofórico; dado que estas técnicas al ser estáticas tienen una baja precisión al modelar proteínas con una gran movilidad. El desarrollo de las unidades de procesamiento gráfico (GPUs) ha permitido desarrollar técnicas computacionales que permiten realizar complejas operaciones matemáticas. Dentro de las técnicas *in silico* más utilizadas que necesitan GPU se encuentran las simulaciones de Dinámica Molecular y el *Machine Learning*. La primera de estas técnicas está basada en las leyes de física descritas por Newton y permite predecir y analizar la evolución de un sistema proteína-ligando a lo largo de un breve intervalo de tiempo (< 1 segundo). De esta manera se pueden predecir cambios conformacionales que pueda sufrir la proteína, así como evoluciona el ligando en el sistema.

Por otro lado, el *Machine Learning* es una rama de la inteligencia artificial que al usarse en el ámbito de la química permite encontrar patrones moleculares de una manera rápida y además puede descubrir relaciones entre propiedades o características. Por ejemplo, si un determinado grupo funcional es el responsable de la actividad del compuesto, además de predecir, la actividad moduladora de compuestos que sería imposible de obtener sin estas técnicas. La capacidad para descubrir estos patrones se debe no solamente al *Machine Learning*, sino también a que actualmente hay una gran cantidad de datos que se pueden utilizar como dataset de entrada para generar modelos de *Machine Learning*. Cabe destacar el uso de AlphaFold 2 (<https://alphafold.ebi.ac.uk/>) como software de *Machine Learning* para predecir el plegamiento proteico.

Por lo tanto, esta tesis doctoral se centra en aprovechar los métodos descritos para las tecnologías mencionadas anteriormente: las simulaciones de Dinámica Molecular y el *Machine Learning* en el desarrollo y descubrimiento de fármacos para tratar diversas patologías. En esta tesis se explora el uso combinado e individual de ambas técnicas; junto con otras técnicas de la química computacional con un coste y complejidad menor, en diversos ámbitos como el cáncer, la diabetes y otras enfermedades.

III – OBJETIVOS

III - OBJETIVOS

Esta tesis doctoral tiene dos objetivos principales:

1. El primero de ellos es la mejora del rendimiento de la herramienta de dinámica molecular Maestro-Desmond, la cual es usada en el descubrimiento de nuevos compuestos utilizados para tratar enfermedades. Para ello, se ha propuesto la implementación de un protocolo de dinámica molecular masiva. Desarrollando primero dicho protocolo y, posteriormente se implementa el protocolo en una herramienta que se liberará a la comunidad científica.

2. El segundo objetivo principal es la aplicación de métodos de computación de alto rendimiento (CAR) en diversas patologías. De esta manera se pretende obtener nuevos compuestos en diferentes contextos como:
 - 2.1. Cáncer.
 - 2.2. Alzheimer.
 - 2.3. Retinitis pigmentosa.
 - 2.4. Enfermedades relacionadas con la cicatrización.
 - 2.5. Diabetes.
 - 2.6. Resistencia a antibióticos.

Por último, también se fijaron como objetivos secundarios la evaluación y mejora de software ya existentes:

- 3.1. Mejora de las herramientas utilizadas para lanzar cálculos de *Docking* y su posterior análisis.
- 3.2. Evaluación de técnicas de *Machine Learning* en diferentes enfermedades

IV - MATERIAL Y MÉTODO

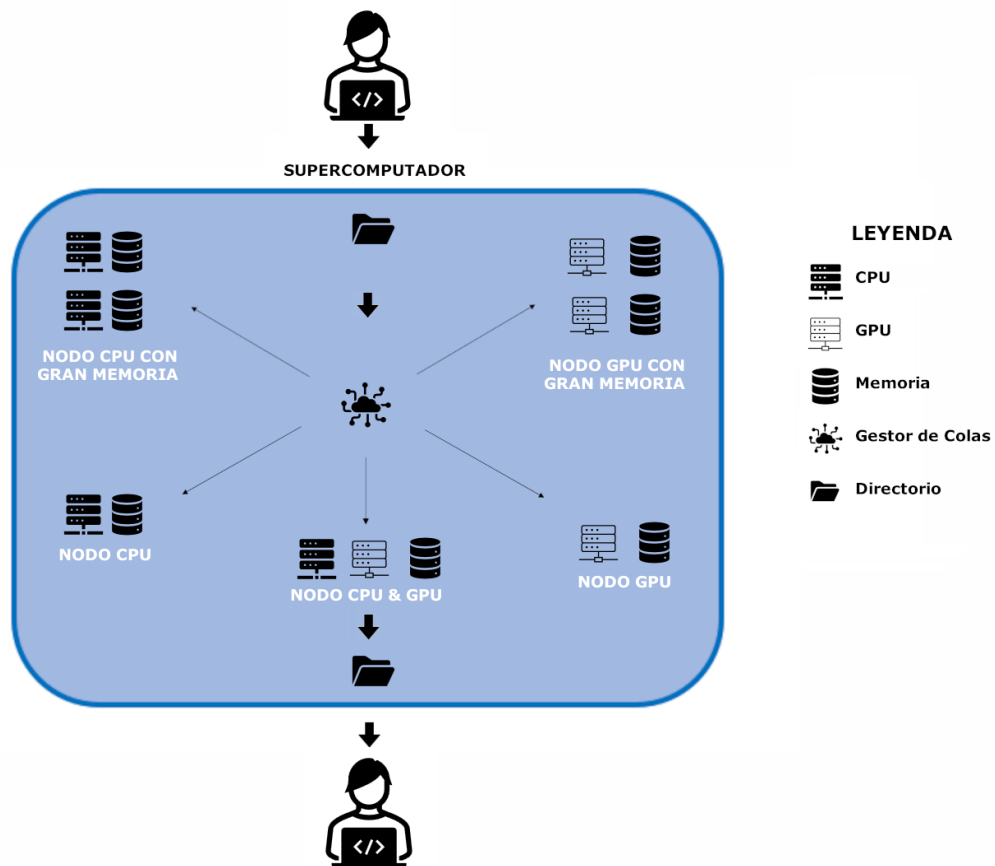
IV -MATERIAL Y MÉTODO

4.1. HIGH PERFORMANCE COMPUTING

La computación de altas prestaciones (HPC) se ha utilizado para ejecutar trabajos complejos que requieren muchos recursos. Para ello, se han utilizado recursos tanto de CPUs (unidad de procesamiento central) como de GPUs (unidades de procesamiento gráfico) de diferentes supercomputadores como son Turgalium de CIEMAT (<https://www.ceta-ciemat.es/>), Picasso del SCBI (Centro de Supercomputación y Bioinnovación) de la Universidad de Málaga (<https://www.scbi.uma.es/web/es/inicio/>), Saga de Sigma en Noruega (https://documentation.sigma2.no/hpc_machines/saga.html), Eagle del PSNC (Poznan Supercomputing and Networking Center) en Polonia (<https://pcss.plcloud.pl>) y Leftraru del NLHPC (Laboratorio Nacional de Computación de Alto Rendimiento) en Chile (<https://www.nlhpc.cl/infraestructura/>). Los recursos de CPU han sido necesarios para realizar *Virtual Screening* utilizando técnicas como el *Docking* molecular o el modelado farmacofórico, en este caso el uso de CPUs se debe a que los algoritmos de *Docking* utilizados por nuestro grupo solamente pueden ejecutarse utilizando una arquitectura de CPU. Mientras que los recursos de GPU se han utilizado principalmente para ejecutar modelos de *Machine Learning* y realizar simulaciones de Dinámica Molecular, en este caso debido principalmente a que el software Maestro-Desmond ha sido diseñado para ser utilizado mediante arquitecturas GPU. Este tipo de cálculos utilizan la programación paralela para optimizar el uso de los recursos de los supercomputadores mencionados previamente. Los supercomputadores dividen sus capacidades en particiones según las características específica de cada tipo de recursos (memoria máxima a utilizar, número de GPUs, número de CPUs, etc.). Cada partición está dividida en nodos, estos nodos son las máquinas donde se realizan los cálculos individuales que ejecuta el supercomputador. Para ejecutar de forma óptima cada trabajo , se utilizan gestores de colas tales como SLURM (121) o SGE (122), los cuales asignan los recursos de forma que utilizan el nodo más óptimo disponible para lanzar un cálculo específico, todo ello está representado en la Figura IV.1. Además, dado que los cálculos se realizan en sistemas operativos basados en Linux, instalar la versión

adecuada de los recursos informáticos específicos requeridos (librerías) para ejecutar cada software es complejo, y es necesario utilizar contenedores, los cuales son archivos que tienen instalado un sistema operativo concreto con las librerías mínimas necesarias para funcionar un determinado programa de forma que permiten ejecutar los software sin necesidad de instalar ninguna dependencia adicional. Un ejemplo de estos programas es Singularity (<https://sylabs.io/docs/>) el cual está instalado en la mayoría de los supercomputadores mencionados anteriormente, y ha sido utilizado en la mayoría de los programas y cálculos desarrollados en esta tesis.

Figura IV.1 Diagrama de trabajo en HPC. Se trata de un diagrama de trabajo de la ejecución de un trabajo mediante el uso de un supercomputador, a través del uso de gestores de colas.



4.2. MODELADO MOLECULAR

El modelado molecular es un conjunto de técnicas las cuales tratan de simular el comportamiento de las moléculas en el medio biológico mediante la generación de modelos computacionales. Nuestro grupo trabaja con modelos de mecánica clásica, modelado farmacofórico y *Machine Learning*. Estos modelos permiten predecir la evolución de proteínas unidas a ligandos reduciendo el número de experimentos a realizar en células y animales. El uso del modelado molecular permite no solo ahorrar tiempo y dinero sino también optimizar el diseño de los fármacos y testar compuestos cuya síntesis y purificación es compleja. Se pueden clasificar las moléculas a modelar en base a su tamaño y complejidad. Se considera que una molécula es pequeña si tiene las siguientes características: máximo 50 átomos pesados, menos de 20 enlaces rotables y un peso atómico inferior a 500 Dalton. Estas moléculas suelen ser ligandos o pequeños péptidos y se modelan de forma diferentes a las proteínas o a los péptidos con un gran número de aminoácidos (más de 10 aminoácidos).

4.2.1. Preparación de Estructuras en Formato mol2

Para realizar los modelos moleculares es necesario transformar las moléculas en formatos que puedan ser procesados por ordenadores. Uno de los formatos más importantes utilizados es el formato mol2. Este formato da información tanto del orden de los enlaces como de las cargas y la hibridación que tiene cada átomo de la molécula. El formato mol2 se utiliza para hacer cálculos de *Docking* molecular usando el software Lead Finder (<https://www.cresset-group.com/software/leadfinder/>) (99) o para hacer simulaciones de Dinámica Molecular. Las proteínas se descargan originalmente de la base de datos Protein Data Bank (<https://www.rcsb.org/>) en formato pdb. Posteriormente, se abre dicho archivo pdb con la suite de Maestro y se limpia la proteína (se eliminan las moléculas de agua, iones y ligandos cocrystalizados). Una vez se ha dejado la proteína sin ningún elemento cocrystalizado, se procede al uso de la herramienta Protein Preparation Wizard para corregir posibles errores en el orden de los enlaces. Por último, con la proteína ya preprocesada se utiliza la herramienta System Builder para asignar el campo de fuerza adecuado (OPLS2005 (123), OPLS3e (124) u OPLS4 (125)). La proteína ya cargada se exporta en formato mol2.

Los ligandos y péptidos pequeños se obtienen en formato pdb, sdf o SMILES (126). Si se obtiene en formato SMILES es necesario transformar dicho formato a sdf o pdb utilizando la herramienta structconvert de Maestro vía línea de comandos. También se ha de revisar el compuesto para observar que no hay problemas en el orden de los enlaces o en las cargas del compuesto. Para evitar este tipo de errores, se debe o bien descargar el ligando en sdf o bien utilizar la suite de Maestro para asignar orden de los enlaces. Una vez se tiene el compuesto con su conformación en 2D correcta, se procede a usar la herramienta LigPrep que permite ionizar, tautomerizar y asignar el campo de fuerza adecuado a la molécula. Por último, al igual que ocurre con las proteínas se exporta el ligando en formato mol2.

4.2.2. Preparación de Estructuras en Formato pdbqt

Otro de los formatos utilizados en el modelado molecular es el formato pdbqt, el cual tiene datos sobre el tipo de átomo y su carga parcial, aunque carece de información sobre el orden de los enlaces. Este formato es necesario para hacer cálculos de *Docking* molecular usando el software AutoDock Vina (98). El método más adecuado para procesar las proteínas y convertirlas al formato pdbqt consiste partir de la molécula en formato mol2, obtenida a través Maestro-Desmond, y utilizar los programas AutoDockTools o MGLTools para realizar la conversión a pdbqt. Para ello, se carga la proteína en formato mol2 y se hace en el siguiente orden: asignar el tipo de átomo a AD4, asignar las cargas del campo de fuerza Gasteiger (127) y se exporta la proteína en formato pdbqt (seleccionando exclusivamente los átomos).

Sin embargo, los ligandos o péptidos pequeños no se procesan con un programa con GUI (Interfaz Gráfica de Usuario), sino que se usan un script en pythonsh llamado *prepare_ligand4.py*, con este script se puede convertir pequeñas moléculas en formato pdb o mol2 a pdbqt asignando el campo de fuerza Gasteiger. Un ejemplo de uso de esta línea de comandos es la siguiente:

```
pythonsh prepare_ligand4.py -l ligand.mol2 -o ligand.pdbqt -A
`hydrogens' -U ` ` `
```

4.2.3. Preparación de Estructuras en Formato pmz/pml

Las moléculas en formato pmz o pml se usan con el software Ligand Scout para ser utilizadas para generar modelos farmacofóricos que posteriormente permiten hacer cribado virtual con ellas. Debido a que se puede hacer modelado farmacofórico tanto basado en los ligandos como en la estructura, es necesario procesar las moléculas de forma diferente.

Si es un modelo farmacofórico basado en los ligandos es necesario abrir el software Ligand Scout (111) y cargar el ligando en formato sdf o mol2. Si está en formato sdf, es necesario minimizar la energía usando el campo de fuerza MMFF94 (128). Una vez se tiene el compuesto con su conformación más favorable se procede a generar el modelado farmacofórico (se puede generar de forma automática y refinar las características farmacofóricas de forma manual). Por último, se añaden las esferas de exclusión y se exporta el modelo farmacofórico en formato pmz/pml.

Mientras que, si se trata de un modelo farmacofórico basado en la estructura, es necesario descargar el complejo proteína-ligando de la base de datos PDB o cargar el complejo en formato pdb. En este caso es importante no minimizar la energía. Se selecciona la zona de interés donde se encuentra el ligando y se genera automáticamente tanto las características farmacofóricas como las esferas de exclusión. Por último, al igual que ocurre con el modelado basado en ligandos se exporta el modelo farmacofórico en formato pmz/pml.

4.2.4. Librerías de compuestos

Los cálculos masivos de *Virtual Screening* utilizan librerías virtuales de compuestos. Estas librerías son colecciones digitales de moléculas y permiten simular las propiedades que tiene cada molécula para ser utilizada en técnicas *in silico*. Las principales librerías utilizadas en esta tesis son DrugBank (<https://go.drugbank.com>), FoodBank (<https://foodb.ca>) y Enamine

(<https://new.enaminestore.com>). Dichas tienen 10000, 12000 y más de 2 millones de compuestos respectivamente. Estas librerías de compuestos se encuentran en formato sdf, mol2, pdbqt y ldb (formato utilizado por Ligand Scout).

La librería en formato ldb se genera utilizando el software Ligand Scout (<https://docs.inteligand.com/ligandscout/>) en la pestaña de *Screening*. Se crea una librería a partir de un archivo sdf y permite automatizar la creación de la librería en formato ldb. La librería ldb se usa para obtener moléculas candidatas mediante modelado farmacofórico.

Es importante destacar que no solamente se pueden utilizar librerías ya creadas como DrugBank y FoodBank. Si se quieren realizar cálculos con moléculas personalizadas se pueden crear librerías en sdf de bases de datos como PubChem (<https://pubchem.ncbi.nlm.nih.gov>) y ZINC (<https://zinc15.docking.org>). De esta manera se pueden filtrar de forma rápida los millones de moléculas que hay en estas bases de datos, y obtener librerías con menor número de moléculas de forma que se puede procesar utilizando menos recursos y en menor tiempo.

4.3. PREDICCIÓN DE ESTRUCTURAS DE PROTEÍNAS

En la sección anterior se encuentra el procedimiento para preparar las proteínas que van utilizadas realizar cálculos que permitan hacer los modelos computacionales utilizados en las técnicas *in silico* y así ser capaz de predecir como interaccionan la proteína con el ligando de interés. La mayoría de estas proteínas son procesadas directamente desde la base de datos PDB. Sin embargo, a pesar de que nuestro grupo está centrado principalmente en el descubrimiento de fármacos, también se pueden utilizar técnicas para predecir la estructura tridimensional de las proteínas; así como el efecto de mutaciones puntuales en proteínas que no se encuentran depositadas en la PDB. Esto nos puede permitir modelar estas proteínas y encontrar nuevos compuestos que permitan tratar patologías cuya estructura tridimensional aún es desconocida. Para ello, se han utilizado herramientas como AlphaFold 2 (<https://alphafold.ebi.ac.uk/>) (95) y Rosetta (91). AlphaFold 2 es una herramienta de *Machine Learning* que permite predecir la

estructura tridimensional basándose únicamente en la secuencia de aminoácidos de la proteína, gracias a esto y tal como se han mencionado previamente, se pueden obtener nuevos compuestos que permitan corregir la función de proteínas mutantes. Las herramientas de *Machine Learning* necesitan de una base de datos que utilizar para obtener los modelos tridimensionales, en este caso se ha utilizado AlphaFold 2 con la base de datos predefinida. Esta base de datos contiene todas las proteínas presentes en la PDB, hay que tener en cuenta que, debido a la diversidad de las proteínas presentes en esta base de datos, es posible que el modelo generado por AlphaFold 2 fuerce la secuencia de entrada a una estructura tridimensional ya existente en la base de datos predefinida (*overfitting*). Mientras que la herramienta de Rosetta se ha utilizado para predecir cómo de favorable o desfavorable es una determinada mutación, de esta manera se puede por un lado predecir como afectaría la mutación en la función de una proteína, además del compuesto de interés para modular la proteína mutada. La herramienta Rosetta se puede utilizar mediante un servidor web, ROSIE2 (<https://r2.graylab.jhu.edu/>), para predecir mutaciones puntuales, es decir, cambios en un solo aminoácido.

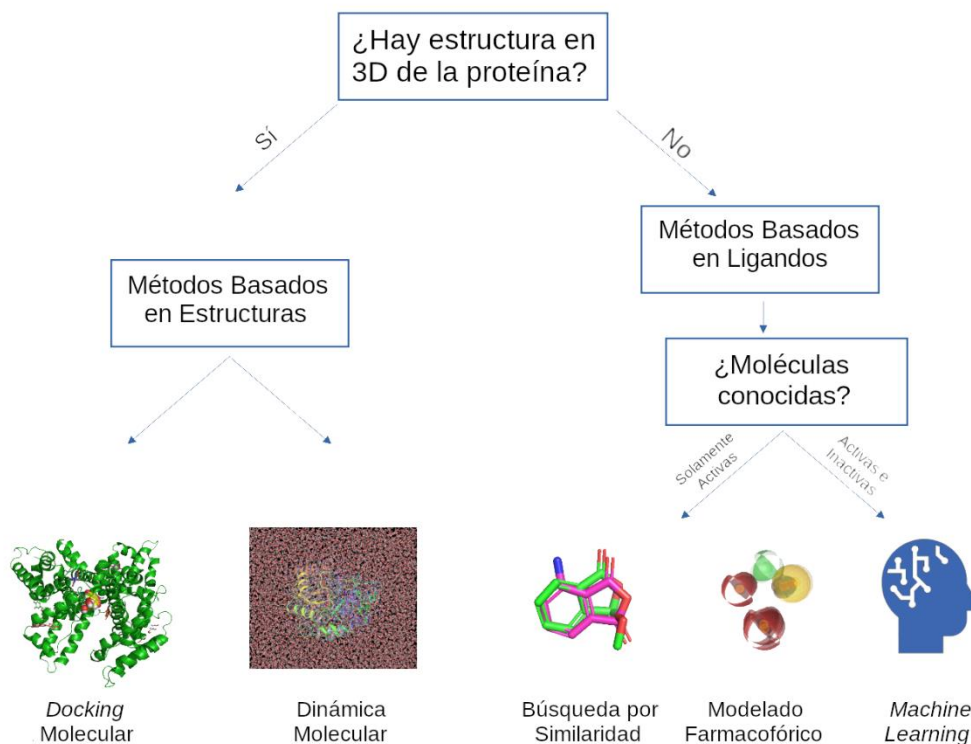
4.4. CRIBADO VIRTUAL

Los modelados de proteínas y compuestos mencionados en las secciones anteriores se utilizan para realizar un cribado virtual. Este cálculo de cribado virtual pretende simular las condiciones biológicas mediante cálculos y algoritmos realizados por ordenador. De esta manera se pueden identificar potenciales candidatos que serán testados en el laboratorio mediante el uso de técnicas *in vitro* e *in vivo*.

En la actualidad existen librerías con millones de compuestos, potenciales candidatos para el tratamiento de diversas enfermedades. Debido al gran número de compuestos es necesario utilizar las técnicas computacionales para filtrar este gran número de compuestos de una forma rápida y con una inversión económica menor. Estas técnicas están basadas en las propiedades fisicoquímicas tanto del complejo proteína-ligando como los de las moléculas del medio biológico (agua, iones, etc.). Dependiendo del tipo de información sobre la que se tiene acceso se pueden realizar dos tipos de técnicas: las basadas en ligandos, cuando no se tiene

acceso a la estructura tridimensional de la proteína y las basadas en la estructura, cuando la proteína sí está caracterizada.

Figura IV.2. Diagrama de flujo de métodos de diseño de fármacos. El diagrama explica que método se puede utilizar según la disponibilidad de la estructura 3D de la proteína y la información de moléculas conocidas.



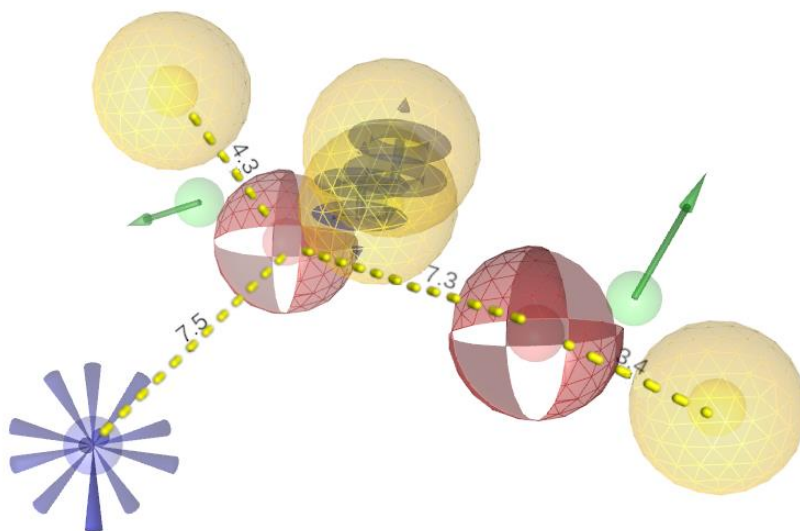
4.5. TÉCNICAS DE CRIBADO VIRTUAL BASADAS EN LIGANDOS

Las técnicas basadas en los ligandos permiten encontrar compuestos similares al ligando o fármaco a estudiar. En nuestro caso nos vamos a centrar en dos tipos de técnicas: el modelado farmacofórico y el modelado basado en *fingerprints*.

4.5.1. Modelado farmacofórico

El modelado farmacofórico utilizado por nuestro grupo se basa en el uso del software Ligand Scout (111). Este software se puede utilizar mediante su interfaz gráfica. Para ello, primero se prepara el fichero pmz/pml que se usará como *target*. Este fichero se prepara en la pestaña “*Structure-Based*” de Ligand Scout y se pueden añadir las características farmacofóricas de forma automática o manual. Por otro lado, en la pestaña de “*Screening*” se genera las librerías en formato ldb partiendo de las moléculas en formato sdf. Una vez se tienen tanto los archivos pmz como el ldb, se realiza el cribado virtual en la pestaña de “*Screening*”. La interfaz gráfica se puede utilizar si el número de modelos farmacofóricos presentes en el archivo ldb es bajo (<100 modelos). Sin embargo, para la mayoría de los casos el número de moléculas a analizar es del orden de miles de compuestos, tal y como se ha mencionado en el apartado “IV.2.4. Librerías de compuestos”. El uso de estas librerías permite encontrar un número mayor de candidatos y de esta manera aumentar la posibilidad de que una de las moléculas candidatas pueda ser un modulador de la proteína de interés. Debido a este alto número de compuestos es necesario recurrir a la computación paralela para realizar los cálculos, sin la computación de alto rendimiento sería inviable realizar en un tiempo razonable estos cálculos. Además, para facilitar en gran medida el cálculo masivo realizado por el software Ligand Scout, los cálculos se han realizado utilizando un software desarrollado por nuestro grupo, llamado MetaScreener (<https://github.com/bio-hpc/metascreeener>); de esta manera se consigue que el usuario tenga que lanzar una sola línea de comandos en lugar de lanzar todos los cálculos de forma manual. Esta herramienta permite ejecutar cálculos de *Virtual Screening* en supercomputadores explotando al máximo los recursos de cada clúster. Al igual que ocurre en los cálculos realizados mediante la interfaz gráfica, el cálculo realizado con MetaScreener necesita dos ficheros de entrada: el primer fichero, es el *target*, el cual es un archivo pmz o pml, generado mediante la interfaz gráfica. El archivo pmz contiene las características farmacofóricas que tiene el modelo generado por el software LigandScout del fichero *target*. El modelo contiene información no solo sobre qué características farmacofóricas tiene el modelo, sino además de su disposición tridimensional (distancias y ángulos que existen entre cada característica farmacofórica), un ejemplo de este modelo se ve en la Figura IV.3.

Figura IV.3 Ejemplo de modelo farmacofórico obtenido por Ligand Scout.



El otro fichero de entrada es la librería en formato ldb (tal y como se ha descrito en el apartado IV.3.4 “Librerías de compuestos”). Una vez han finalizado los cálculos se obtiene un *ranking* con los mejores resultados obtenidos realizados en este cribado virtual, teniendo en cuenta tanto el tamaño y forma como el número de características farmacofóricas comunes entre el *target* y los compuestos de la librería. De esta manera se obtiene un valor de similaridad farmacofórica cuyo rango se encuentra en cero y uno; siendo uno un valor de similaridad perfecta y siendo cero un valor de similaridad nula. Por lo que los compuestos que tenga un valor cercano a uno son los compuestos con mayor similaridad respecto al compuesto *target*, estos compuestos serán candidatos a ser utilizados en los experimentos posteriores. Si se tiene varios compuestos *targets* de interés se puede realizar un *Virtual Screening* consenso utilizando MetaScreener para encontrar moléculas similares a los *targets* de interés. Además, si se tienen varias librerías a analizar se pueden clasificar de forma conjunta realizando este *Virtual Screening* consenso. Un ejemplo de este consenso se observa en la Figura IV.4.

La siguiente línea de comandos muestra un ejemplo de ejecución de un cálculo de modelado farmacofórico con la herramienta MetaScreener:

```
./ms.sh -t target.pmz -q query.ldb -j 100 -o VS -s LS -qu nodo1  
-hi y -a 0
```

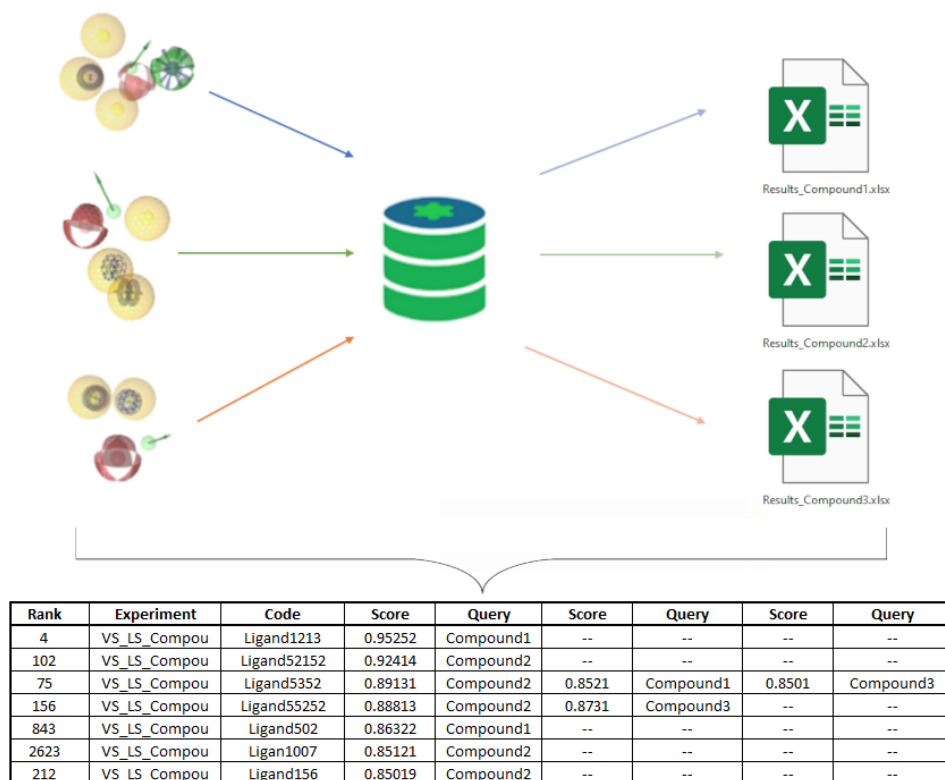
En este caso además de los parámetros de *target* (-t) y *query* (-q), se ha indicado que el número de trabajos (cada uno de los bloques de cálculos en los que se divide el total de cálculos) es 100 (-j), el tipo (MetaScreener admite dos tipos: *Blind Docking* y *Virtual Screening*) de trabajo a realizar (-o), el software utilizado (-s), la partición (parte del clúster donde se encuentra el nodo utilizado) del clúster donde se va a realizar el cálculo (-qu), el análisis final, *ranking* con los resultados agrupados de los trabajos individuales, (-hi) que se va a realizar (opción y) y, dado que es un cálculo de Ligand Scout se necesita indicar si se puede omitir alguna característica, para ello, se utiliza el parámetro -a; por defecto este valor es dos pero es recomendable poner un valor de -a igual a cero. Hay que recalcar que el tiempo de cálculo crece de forma exponencial con respecto al valor de a. Por ello, no es recomendable poner un valor de a superior a dos para evitar que el tiempo de cálculo sea inviable.

Por último, si se quiere realizar un cálculo de consenso se utiliza el script consensus de la siguiente manera:

```
./MetaScreener/extra_metascreener/results/consensus.sh -o LS -  
f VS_LS_target_ -out LS_Consensus_target_query
```

En este caso se indica el tipo de trabajo con el parámetro -o, el inicio de los directorios que serán incluidos en el consenso (-f) y el nombre del archivo final (-out).

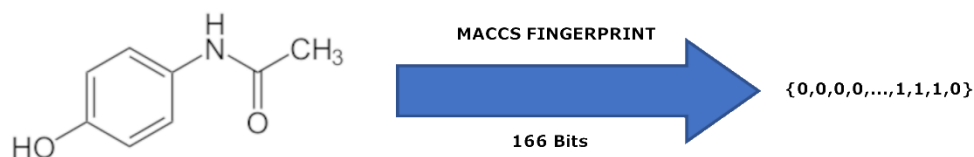
Figura IV.4. Diagrama de flujo de cribado basado en modelado farmacofórico utilizando un Virtual Screening consenso.



4.5.2. Modelado basado en fingerprints

El modelado basado en *fingerprints* permite desglosar los compuestos en diferentes características que tiene la molécula, tales como patrones moleculares (vectores binarios que representan la ausencia o presencia de un determinado grupo químico). Para ello, nuestro grupo ha desarrollado un programa capaz de transformar el código SMILES de un ligando en diferentes *fingerprints* moleculares. Estas fingerprints se están utilizando para generar modelos de *Machine Learning*.

Figura IV.5. Representación en MACCS fingerprints del paracetamol.



4.6. TÉCNICAS DE CRIBADO VIRTUAL BASADAS EN LA ESTRUCTURA

Las técnicas basadas en la estructura se fundamentan en procesar la información correspondiente a las interacciones proteína-ligando o la información estructural tridimensional que tiene la proteína diana, de esta forma se pueden encontrar compuestos que interactúan con la proteína diana. En este tipo de técnicas es imprescindible disponer de la proteína con su estructura tridimensional para poder realizar los cálculos necesarios para descubrir nuevos compuestos.

4.6.1. Docking Molecular

El *Docking* Molecular es una técnica basada en la estructura que permite predecir cuál es la pose, que es la conformación específica que adquiere el ligando si es capaz de interactuar con la proteína de interés. Además, el *Docking* Molecular permite dar un valor estimado de cómo de favorable es la interacción dado que proporciona un valor de *Docking score*, el cual está directamente relacionado con la energía de Gibbs. Este tipo de técnicas suelen tener un bajo coste computacional y, aunque, predicen con gran precisión la pose con mayor probabilidad que va a adquirir el ligando son menos precisas a la hora de evaluar la energía que tiene el complejo proteína-ligando generado. En la mayoría de los casos se realiza el cálculo de *Virtual Screening* contra la zona de interés, ya sea el sitio activo o una zona alostérica.

4.6.1.1. *AutoDock Vina*

Uno de los software más utilizados para realizar los cálculos de *Docking* es AutoDock Vina (98). Este software utiliza los archivos en formato pdbqt tal y como se ha descrito en el paso IV.3.2 “Moléculas en formato pdbqt”. Al igual que se ha descrito en el paso anterior con el software Ligand Scout, se ha utilizado el software MetaScreener. En este caso, el cálculo se realiza de la siguiente manera:

```
./ms.sh -t Protein.pdbqt -q Ligand.pdbqt -j 100 -o VS -s AD -  
qu nodol -hi y -x -9 -y 1.2 -z -8
```

En este caso, además de los parámetros de target (-t) y query (-q), se ha indicado que el número de trabajos es 100 (-j), el tipo de trabajo a realizar (-o), el software utilizado (-s), las coordenadas donde se va a realizar el *Docking* (-x, -y, -z), la partición donde se va a realizar el cálculo (-qu) y el análisis final (-hi) que se va a realizar (opción y).

4.6.1.2. *Lead Finder*

Otro de los software utilizados para realizar este tipo de cálculos es el software Lead Finder (99). Este software utiliza los archivos en formato mol2 generados según los protocolos vistos en el paso IV.3.1. “Moléculas en formato mol2”. Este software se utiliza mediante la herramienta MetaScreener, un ejemplo de la línea de comandos a realizar se encuentra en la siguiente línea:

```
./ms.sh -t Protein.mol2 -q Ligand.mol2 -j 100 -o VS -s LF -qu  
nodol -hi y -x -9 -y 1.2 -z -8
```

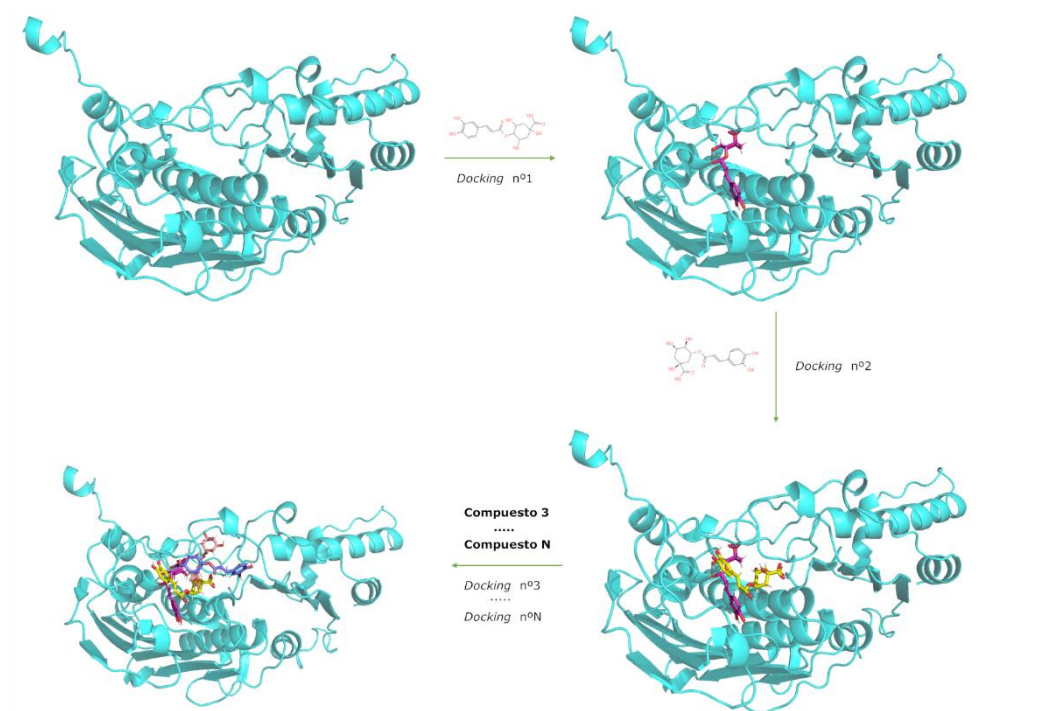
En este caso además de los parámetros de target (-t) y query (-q), se ha indicado que el número de trabajos es 100 (-j), el tipo de trabajo a realizar (-o), el software utilizado (-s), las coordenadas donde se va a realizar el *Docking* (-x, -y, -z), la partición donde se va a realizar el cálculo (-qu) y el análisis final (-hi) que se va a realizar (opción y).

4.6.1.3. *Docking Sucesivos*

También se pueden realizar varios cálculos de *Dockings* sucesivos en la zona de interés. De esta forma se puede ver cómo interaccionan diferentes ligandos de forma simultánea en la misma zona de interés. Se puede utilizar tanto el software AutoDock Vina como Lead Finder. La forma de realizar los *Dockings* sucesivos es bastante simple, se realiza de forma paralela los cálculos de *Docking* individuales. Posteriormente, la pose del ligando obtenida en el cálculo se une en formato pdb (si viene de AutoDock Vina) o en mol2 (si viene de Lead Finder). Después, se usa el protocolo visto en la sección IV.3. “Moléculas y formatos” para generar cada target que será la proteína unida a cada ligando. Una vez se ha obtenido el complejo proteína-ligando, se realizará un nuevo cálculo de *Docking* individual, tal y como se ve en la Figura IV.6.

Por último, se selecciona la secuencia basándose en los *Docking score* individuales obtenidos en cada paso. Evitando secuencias que tengan un paso con una puntuación que implique una interacción desfavorable.

Figura IV.6. Esquema de cálculos de *Dockings* sucesivos.



4.6.1.4. *Docking Consenso*

De la misma manera que se ha visto con el uso del software Ligand Scout, se puede realizar un cálculo de consenso frente a varios cálculos de *Virtual Screening*. Hay dos tipos de consenso: el primero se trata de usando el mismo tipo de software (AutoDock Vina o Lead Finder) hacer un consenso de varias librerías de compuestos, para ver qué compuestos son los mejores. El otro tipo de consenso es teniendo una sola librería comparar los resultados de *Docking* obtenidos usando los dos softwares de forma que se obtiene un ranking promedio con los mejores resultados obtenidos en ambos software.

Un ejemplo de cálculo de *Docking* consenso es:

```
./MetaScreener/extra_metascreener/results/consensus.sh -o VS -  
f VS_*_Protein_Library_2024_*/ -out VSConensusProteinLibrary -r  
Protein.mol2
```

En este caso se indica el tipo de trabajo con el parámetro -o, el inicio de los directorios que serán incluidos en el consenso (-), el nombre del archivo final (-out) y la proteína (-r).

4.6.1.5. *Blind Docking*

Además del *Virtual Screening*, muchas veces se quiere hacer un barrido a la superficie de la proteína utilizando un ligando a modo de sonda para saber qué zonas de la proteína son las que tienen mayor probabilidad de interactuar con el ligando de interés. Esta técnica se conoce como *Docking* a ciegas o *Blind Docking* en inglés. Gracias al *Blind Docking* se puede escanear toda la superficie de la proteína, permitiendo encontrar el sitio de unión de ligandos cuya actividad contra una determinada proteína es conocida pero la zona de interacción contra dicha proteína es todavía desconocida. Al igual que en el resto de los cálculos es necesario el uso de la herramienta MetaScreener. Un ejemplo de uso es el siguiente:

```
./ms.sh -t Protein.mol2 -q Ligand.mol2 -j 100 -o BD -s LF -qu  
nodol -hi y
```

En este caso se indican los parámetros de target (-t), query (-q), se ha indicado que el número de trabajos es 100 (-j), el tipo de trabajo a realizar (-o), el software utilizado (-s), la partición donde se va a realizar el cálculo (-qu) y el análisis final (-hi) que se va a realizar (opción y).

Al igual que se ha visto en la sección anterior es posible realizar cálculos de *Blind Docking* consenso. Este consenso se puede utilizar para hacer consenso de dos software, para hacer un consenso de varios ligandos o una combinación de las dos opciones previas. En este caso, el cálculo a realizar se haría de la siguiente forma:

```
./MetaScreener/extra_metascreener/results/consensus.sh -o BD -  
f BD_*_Protein_Compound_2024_*/ -out BDConsensusProteinCompound -r  
Protein.mol2
```

En este caso se indica el tipo de trabajo con el parámetro -o, el inicio de los directorios que serán incluidos en el consenso (-), el nombre del archivo final (-out) y la proteína (-r).

4.6.2. Dinámica Molecular

Las simulaciones de Dinámica Molecular son una de las técnicas basadas en la estructura. Esta herramienta requiere de unos costes computacionales mucho mayores que los utilizados en el *Docking*, además la precisión de esta técnica es mayor. Dentro de todos los motores de Dinámica Molecular el que se ha utilizado en esta tesis es Desmond, utilizado a través del software Maestro-Desmond desarrollado por la empresa Schrodinger. Para realizar este tipo de cálculos se puede utilizar la interfaz gráfica de Maestro-Desmond. Además de realizar la simulación, Maestro-Desmond permite analizar la trayectoria de una forma fácil e intuitiva. Las trayectorias generadas por Maestro-Desmond permiten hacer análisis

externos como el MMGBSA (*Molecular Mechanics Generalized Born Surface Area*) y el PSN (*Protein Structure Network*).

4.6.2.1. *MMGBSA*

El análisis del cálculo de MMGBSA (129) se utiliza para obtener la energía libre que hay en un complejo proteína-ligando y las diferencias que hay entre el sistema acompañado y sin ligando. Este análisis es altamente utilizado en química computacional y bioinformática estructural. Es importante tener en cuenta que, para hacer el cálculo de MMGBSA, es necesario tener la licencia de la herramienta Prime (130,131).

4.6.2.2. *PSN*

El cálculo de PSN permite obtener cómo se forman clústeres entre los diversos residuos de una proteína. De esta manera al aplicarlo a una trayectoria de Dinámica Molecular permite que se pueda calcular qué grupo de residuos interactúan durante más tiempo y evaluar de esta manera la estabilidad de la proteína a nivel local. Este análisis está implementado mediante el uso de herramientas de PSNtools (132).

4.7. MACHINE LEARNING

Otra de las técnicas utilizadas en el descubrimiento de fármacos es el *Machine Learning* interpretable. Nuestro grupo ha desarrollado una herramienta que permite lanzar cálculos de *Machine Learning* (ML) interpretable y que puede ser combinada con otras técnicas como los modelos basados en ligandos (diferentes *fingerprints* de moléculas) u otros modelos basados en estructura (simulaciones de Dinámica Molecular). Esta herramienta llamada SIBILA (<https://github.com/bio-hpc/sibila>) permite realizar modelos de ML interpretable y además hacer inferencia. Utiliza diferentes modelos como ANN, DT, KNN, etc. y usa las librerías SHAP (133) y LIME (134) para realizar el cálculo de interpretabilidad.

4.8. EXPERIMENTACIÓN

Tal y como se ha visto en capítulos anteriores, a pesar de la rapidez y bajo coste de las técnicas computacionales actuales; el principal problema de estas técnicas se debe a que no permiten predecir con la máxima precisión el comportamiento de un fármaco. Debido a ello, sigue siendo necesario utilizar técnicas experimentales clásicas del laboratorio. Hay dos tipos de técnicas principales: las técnicas *in vitro* y las técnicas *in vivo*. Ambas técnicas permiten analizar con exactitud el comportamiento del compuesto a estudiar contra la proteína o el organismo diana. Estas técnicas han sido realizadas por colaboradores para testar las predicciones obtenidas mediante técnicas computacionales.

V – RESULTADOS Y DISCUSIÓN

V - RESULTADOS Y DISCUSIÓN

En este capítulo se describen los resultados obtenidos durante la tesis, así como los objetivos planteados en cada sección. En el primer apartado, se desarrolla la implementación de un software para mejorar la eficiencia de la ejecución de las simulaciones de Dinámica Molecular, este software llamado TOLEDO ha sido desarrollado y testado por nuestro grupo. En los apartados que se incluyen entre los puntos 5.2 hasta el 5.7 (ambos incluidos), se va a cumplir el objetivo 2 de la tesis: “La aplicación de métodos de computación de alto rendimiento (CAR) en diversas patologías”. Para el cumplimiento de este objetivo se van a utilizar diversas técnicas incluyendo técnicas *in silico* basadas tanto en la estructura como en los ligandos. Las patologías que se van a tratar son: Cáncer (5.2), Alzheimer (5.3), Retinitis pigmentosa (5.4), Cicatrización (5.5), Diabetes (5.6) y Resistencia a los antibióticos (5.7). En el último apartado, se van a tratar los objetivos secundarios los cuales son la valoración del uso de herramientas de ML, en este caso AlphaFold 2, y de las simulaciones de Dinámica Molecular en el ámbito de la trombosis (5.8.1). También, en este último apartado se han explicado las colaboraciones realizadas con miembros del grupo para la mejora y testeo de la herramienta MetaScreener.

5.1. IMPLEMENTACIÓN DE SOFTWARE PARA DINÁMICA MOLECULAR

Para la consecución del primer objetivo de esta tesis, nuestro grupo ha desarrollado un software basado en Maestro-Desmond llamado TOLEDO (Throughput Optimization of Ligand-Protein systems Exploration through Dynamics simulation in Optimized HPC systems), este software permite explotar las características de Maestro-Desmond. Gracias a TOLEDO es posible ejecutar de forma masiva simulaciones de Dinámica Molecular utilizando una sola línea de comandos, sin la necesidad de tener que preparar de forma individual cada uno de los complejos proteína-ligando. TOLEDO es una herramienta basada en los lenguajes de programación Bash y Python; además de usar contenedores como Singularity para mayor facilidad en su ejecución e instalación. TOLEDO utiliza Desmond como motor de Dinámica Molecular, permitiendo además ejecutar

dinámicas del tipo proteína-proteína, proteína-ligando con membrana o proteínas sin ligando.

TOLEDO ha sido diseñado para además de realizar el análisis básico del comportamiento de la proteína y del ligando, obtener un informe con el análisis completo de MMGBSA y PSN.

El desarrollo de TOLEDO implicó la creación de una serie de scripts que se ejecutan secuencialmente. Teniendo los siguientes scripts: *i)* Script lanzador; *ii)* Script creador de los complejos; *iii)* Script elongador de dinámicas; *iv)* Script para unir las dinámicas; y *v)* Scripts de análisis de las dinámicas.

5.1.1. Script Lanzador

Este programa es el que es ejecutado por el usuario; permite seleccionar las características de la dinámica a ejecutar: presencia/ausencia de membrana, concentración salina del complejo y cambios en los parámetros de configuración de la simulación (temperatura, presión, semilla, etc.). En total, TOLEDO contempla seis modos de simulación distintos:

1. *Virtual Screening* clásico sin membrana.
2. *Virtual Screening* clásico con membrana.
3. *Virtual Screening* sin membrana, pero con modificaciones en la creación del complejo.
4. *Virtual Screening* sin membrana, pero con modificaciones en los parámetros de configuración de la simulación.
5. *Virtual Screening* sin membrana, pero con modificaciones en los parámetros de configuración de la simulación y en la creación del complejo.

6. *Virtual Screening* con membrana y modificaciones en los parámetros de configuración de la simulación.

Posteriormente, dependiendo de las características del tipo de simulación elegida se ejecuta de forma diferente el protocolo de creación de los complejos.

5.1.2. Script Creador de Complejos

Este script crea los complejos de todos los modos de simulación excepto de aquellos que ya tienen membrana (los modos 2 y 6), dado que, en este caso, los complejos están ya creados.

Una vez se han creado los complejos se procede a hacer una dinámica corta (de solo un nanosegundo) para calcular la velocidad y eficiencia de la simulación para ejecutar el resto de los scripts. Una vez ha finalizado esta simulación, se procede a ejecutar el script para elongar las dinámicas.

5.1.3. Script Elongador de Dinámicas

Una vez se ha ejecutado la primera simulación de un nanosegundo se procede a ejecutar la elongación de las dinámicas. Para ello, este script tiene en cuenta la velocidad media de simulación de la primera simulación y el rendimiento del clúster (este parámetro proporciona información sobre el número máximo de horas que un cálculo puede ejecutarse de forma continua en la GPU). Gracias a este cálculo TOLEDO permite optimizar la ejecución de las simulaciones. TOLEDO elonga la simulación hasta el número de nanosegundos calculado previamente, de esta forma se obtiene la simulación aprovechando el máximo de recursos disponibles en el supercomputador (para evitar cancelaciones del trabajo por fallos en el rendimiento se hace una aproximación del 75% del rendimiento máximo). Una vez ha terminado la simulación se procede a realizar otra vez este script de nuevo. En este caso, se vuelve a calcular la velocidad media y el rendimiento; y se da un nuevo tiempo de simulación. Este tiempo puede ser igual o superior al solicitado por el usuario, de ser así se procede a igualar el tiempo de simulación al solicitado por el usuario, y al finalizar la simulación se procede a realizar el script

de unión de cada simulación, así como a realizar un muestreo de la trayectoria. Sin embargo, si el tiempo es inferior al solicitado por el usuario, se procede a volver a ejecutar el script con un nuevo tiempo de simulación final, repitiendo este ciclo hasta que se iguale o supere el tiempo de simulación solicitado por el usuario. Una vez se ha terminado la simulación, se procede a la unión y análisis de las dinámicas.

5.1.4. Script de Unión de Dinámicas

Este script se utiliza al final de la fase de elongación; una vez se ha alcanzado el tiempo de simulación solicitado por el usuario. El script de unión de las dinámicas realiza dos procedimientos: une todas las trayectorias generadas por los scripts anteriores y hace un muestreo para reducir el número de *frames* de la trayectoria final. El motivo de este muestreo es reducir el tiempo de cálculo y tamaño de los ficheros generados. Finalmente, se obtiene una sola trayectoria con el muestreo y la unión ya realizados. Esta trayectoria será utilizada para su posterior análisis por el último script de TOLEDO. Existen dos scripts diferentes para realizar el análisis dependiendo del número de ligandos que tenga el complejo.

5.1.5. Script de Análisis de Dinámicas

Dependiendo del número de ligandos que tenga el complejo, se ejecutará un script diferente de los dos existentes. Si hay un solo ligando, se ejecutará un script que realiza un análisis completo de la proteína y ligando, así como el análisis MMGBSA y PSN. Sin embargo, si hay varios ligandos se ejecutará el segundo script tantas veces como ligandos haya, haciendo un análisis completo proteína-ligando y MMGBSA para cada ligando, pero solo se realizará el análisis PSN en la primera ejecución del script, dado que este análisis es exclusivo de la proteína y no tiene en cuenta el ligando. De esta forma se optimiza el análisis de dinámicas con múltiples ligandos.

5.1.6. Casos de Estudio

Dado que las simulaciones de Dinámica Molecular se pueden utilizar en diferentes contextos, se seleccionaron tres tipos diferentes de casos de estudio para comprobar tanto la eficacia como la aplicabilidad de TOLEDO en los escenarios de mayor importancia. Estos casos de estudio pretenden por un lado evaluar el rendimiento de TOLEDO frente al cálculo individual de cada dinámica de forma independiente. Y, por otro lado, comprobar que los resultados obtenidos en cada enfoque permiten llegar a la misma conclusión, siendo exactamente iguales o presentando diferencias mínimas y despreciables.

Estos casos de estudio comprenden tres clases distintas de cálculos:

1. *Virtual Screening* contra la proteasa del SARS-CoV-2.
2. *Inverse Virtual Screening* de proteínas relacionadas con la diabetes.
3. *Blind Docking* de la luteolina contra la seroalbúmina humana.

El primer caso es el más común de todos, y consiste en la ejecución masiva de múltiples ligandos contra la misma proteína. Este escenario es el más habitual dado que es un procedimiento complementario al *Docking*, en este caso se seleccionan los mejores resultados del *Docking* y se realiza una simulación corta de dinámica molecular (100 ns) para filtrar los resultados obtenidos por el *Docking*. De esta manera, los resultados obtenidos por ambas técnicas son mucho más precisos que los obtenidos por cada técnica de forma individual.

El segundo caso aplicado consiste en realizar simulaciones de Dinámica Molecular de un solo ligando contra una librería de proteínas. Esta técnica llamada *Inverse Virtual Screening* se utiliza de forma masiva en la industria farmacéutica para encontrar posibles proteínas diana y de esta manera encontrar nuevos usos terapéuticos del ligando de interés, esta estrategia se conoce como *target fishing*.

Por último, el tercer caso es importante porque permite evaluar si TOLEDO es capaz de ejecutar y analizar simulaciones de Dinámica Molecular con varios

ligandos. Esto es importante dado que hay proteínas que tienen múltiples sitios activos o se regulan de forma alostérica.

Por ello, realizando estos tres casos de estudio se abordan la mayoría de los usos de las simulaciones de Dinámica Molecular que utilizan complejos proteína-ligando.

5.1.6.1. *Virtual Screening*

Para este caso se han replicado los cálculos realizados por Abel et al. (135), en este estudio se buscaba encontrar compuestos eficaces contra la principal proteasa del SARS-CoV-2 (M^{pro}). Esta proteasa es clave para la replicación viral, por ello, fue utilizada como diana en el estudio previamente mencionado. En el estudio original se realizó una serie de simulaciones de Dinámica Molecular utilizando la GUI de la suite de Maestro-Desmond con la versión 2020-04. Debido a esto, y para tratar de obtener los resultados con la mayor reproducibilidad, se utilizó esta versión de Maestro-Desmond en TOLEDO. Los ligandos seleccionados de este estudio fueron: Naldemedine, Pseudostellarin C, Saralasin, SN00019468, Eledoisin y SN00017653.

Para evaluar las similitudes y diferencias obtenidas entre ambos trabajos se han evaluado las interacciones entre proteína y ligando, el valor de RMSD (*Root Mean Square Deviation*) y de RMSF (*Root Mean Square Fluctuation*) tanto de la proteína como de cada uno de los ligandos en cada complejo.

Con los resultados obtenidos y comparados entre ambos artículos se concluye que a pesar de existir ligeras diferencias (provenientes del uso de muestras al utilizar TOLEDO), los resultados obtenidos muestran cómo TOLEDO ha sido capaz de replicar los resultados obtenidos por Abel et al, en términos de interacciones con residuos comunes en ambos trabajos; así como valores del mismo orden de magnitud tanto en RMSD como en RMSF (135).

Tabla V.1. Valores máximos de RMSD y RMSF obtenidos tras el análisis del Virtual Screening. Los análisis fueron realizados tras las simulaciones de dinámica molecular de la principal proteasa del SARS-CoV-2 contra los ligandos de interés.

Ligando	Proteína		Ligando	
	Valor máximo de RMSD (Å)	Valor máximo de RMSF (Å)	Valor máximo de RMSD (Å)	Valor máximo de RMSF (Å)
Naldemedine	3,1	6,3	6,4	3,5
Pseudostrallarín C	2,8	8,8	10,5	4,4
Saralasin	2,8	4,7	6,4	6,2
SN00019468	2,8	6,1	10,5	4,1
Eledoisin	2,7	4,4	9,0	8,6
SN00017653	2,7	6,5	6,4	4,7

5.1.6.2. *Inverse Virtual Screening*

En este caso, se llevarán a cabo una serie de cálculos para evaluar la aplicabilidad de TOLEDO en una nueva estrategia de diseño de fármacos llamada polifarmacología que forma parte de las estrategias incluidas en el *Inverse Virtual Screening*. Esta estrategia se utiliza para predecir el comportamiento de un ligando de interés frente a un grupo de proteínas involucradas en la misma enfermedad. La enfermedad seleccionada ha sido la diabetes; de la cual se han seleccionado tres proteínas con una gran relevancia como son la dipeptidil peptidasa 4 (DPP4), la aldocetoreductasa ARK1B1 y el receptor activado por proliferador de peroxisomas δ (PPAR δ). El ligando de interés es el teaflavina-3'3'-digalato; este ligando ya ha mostrado tener potencial para interactuar con las proteínas relevantes en un estudio de *Docking* previo (136).

Para este trabajo se realizó una serie de simulaciones de Dinámica Molecular de una duración de 100 ns y se evaluaron tanto las interacciones clave como los valores de RMSD y RMSF tanto de cada proteína como del ligando.

Tabla V.2. Valores máximos de RMSD y RMSF obtenidos tras el análisis del Inverse Virtual Screening. Los análisis fueron realizados tras las simulaciones de dinámica molecular del teaflavina-3'3-digalato contra las proteínas de interés.

Proteína	Proteína		Ligando	
	Valor máximo de RMSD (Å)	Valor máximo de RMSF (Å)	Valor máximo de RMSD (Å)	Valor máximo de RMSF (Å)
DPP4	3,8	4,0	4,5	2,4
ARK1B1	2,2	3,9	4,3	2,6
PPARδ	10,8	6,4	8,9	3,0

Los resultados mostraron que el teaflavina-3'3-digalato podía ser un compuesto con posibilidades de modular la actividad de la dipeptidil peptidasa 4 y de la aldocetoreductasa ARK1B1. Sin embargo, el receptor activado por proliferador de peroxisomas δ mostraba una gran falta de estabilidad alcanzando valores de RMSD cercanos a 11 Å y de RMSF mayores de 6 Å.

5.1.6.3. *Blind Docking combinado con simulaciones de Dinámica Molecular*

Para la realización de este caso de estudio fue necesario un paso previo: la realización de un cálculo de *Blind Docking* del ligando de interés (en este caso la luteolina) frente a la proteína diana (la seroalbúmina humana), en este caso el *Docking* fue realizado usando el algoritmo AutoDock Vina. Se seleccionaron las ocho poses con mejor *Docking score*, dentro de las cuales se encontraba una zona de interacción (pose predicha en el clúster nº3) previamente documentada en la

bibliografía (137). La duración de la simulación fue la misma que en el resto de los casos de estudio (100 ns). También se evaluaron tanto las interacciones clave de cada una de las poses del ligando como los valores de RMSD y RMSF de la proteína y del ligando en cada una de las poses.

Tabla V.3. Valores máximos de RMSD y RMSF obtenidos tras los cálculos de Blind Docking y la simulación de dinámica molecular. La simulación fue realizada por el complejo formado por la luteonina, en sus ocho mejores clusters predichos por el Blind Docking, y la seroalbúmina humana. Además, se incluye el valor de MMGBSA promedio en kcal/mol.

nº Clúster	Valor máximo de RMSD (Å)	Valor máximo de RMSF (Å)	MMGBSA medio (kcal/mol)
1º	8,7	5,3	-33,0
2º	9,1	3,2	-39,9
3º	6,0	1,9	-45,2
4º	7,2	2,2	706,0
5º	24,9	8,4	-21,0
6º	8,0	2,2	-30,8
7º	10,2	6,4	934,7
8º	9,6	4,0	-36,4

Los resultados mostraron que la proteína tuvo bastante estabilidad a lo largo de toda la simulación, los valores máximos de RMSD y RMSF fueron de 4,5 y 5 Å respectivamente. Además, los valores de energía libre predichos por el cálculo de MMGBSA predijeron que el ligando en la pose nº 3 era el más estable. Esto fue confirmado con el análisis de RMSD y RMSF, máximos de 5,6 y 1,9 Å

respectivamente. Por otro lado, el ligando mostró poca afinidad en las poses n^o4 y n^o7 (según los valores de energía libre calculados por el análisis MMGBSA); esto sugiere que estos clústeres no son sitios de unión a luteolina.

Además de estos casos de estudio, TOLEDO es una herramienta en constante crecimiento, por lo que se espera que en próximas actualizaciones pueda analizar sistemas proteína-proteína y mejorar el análisis actual. TOLEDO se encuentra actualmente liberado a la comunidad científica a través de la plataforma GitHub (<https://github.com/bio-hpc/TOLEDO>). GitHub es una plataforma utilizada para la liberación y colaboración en el desarrollo de software, esta plataforma utiliza su propio sistema de control de versiones (git), lo cual permite añadir nuevas características o restaurar versiones previas del software de interés.

5.2. CÁNCER

Para lograr el segundo objetivo de esta tesis, se han realizado cálculos en diversos ámbitos patológicos para predecir el posible uso de compuestos como posibles tratamientos. Por ello, en este apartado de la tesis se abordará el objetivo 2.1. “Aplicación de métodos de computación de alto rendimiento en el cáncer”. Este apartado incluye el desarrollo de todos los trabajos realizados durante la tesis cuya finalidad ha sido el tratamiento de dicha enfermedad.

5.2.1. Casos Especies Reactivas de Oxígeno

La Xantina Oxidasa (XO) es una enzima clave en el metabolismo de las purinas, obteniendo como producto final ácido úrico y generando especies reactivas de oxígeno (ROS). Diversos estudios muestran como un aumento en la actividad de la Xantina Oxidasa está correlacionado con enfermedades como el estrés oxidativo, diabetes o cáncer. Estas enfermedades se producen debido al aumento de las especies reactivas de oxígeno, las cuales producen daños a nivel celular, incluyendo las enzimas implicadas en la replicación celular, favoreciendo de esta manera el desarrollo de diversos cánceres. Diversos compuestos naturales y sintéticos como el alopurinol, la quercetina (138) o el febuxostat (139) ya están

siendo investigados para inhibir la xantina oxidasa en el contexto de la gota. Por ello, se han realizado dos colaboraciones, ambas con el Dr. Housseem Boulebd de la Universidad de Constantina en Argelia, para buscar nuevos inhibidores de esta enzima.

La primera colaboración tiene como ligando de interés a la neobavaisoflavona, la cual es un flavonoide natural que se encuentra en múltiples plantas. Estudios previos muestran como este flavonoide tiene propiedades antiinflamatorias, inhibiendo la producción de citoquinas inflamatorias y reduciendo la infiltración de células inmunitarias (140,141). Respecto a los radicales libres, se ha visto cómo este compuesto reduce la concentración del radical libre DPPH (2,2-difenil-1-picrilhidracilo), por lo que es una candidato con propiedades anticancerígenas (142). En este trabajo se ha visto como la neobavisoflavona puede regular la actividad de la xantina oxidasa (XO).

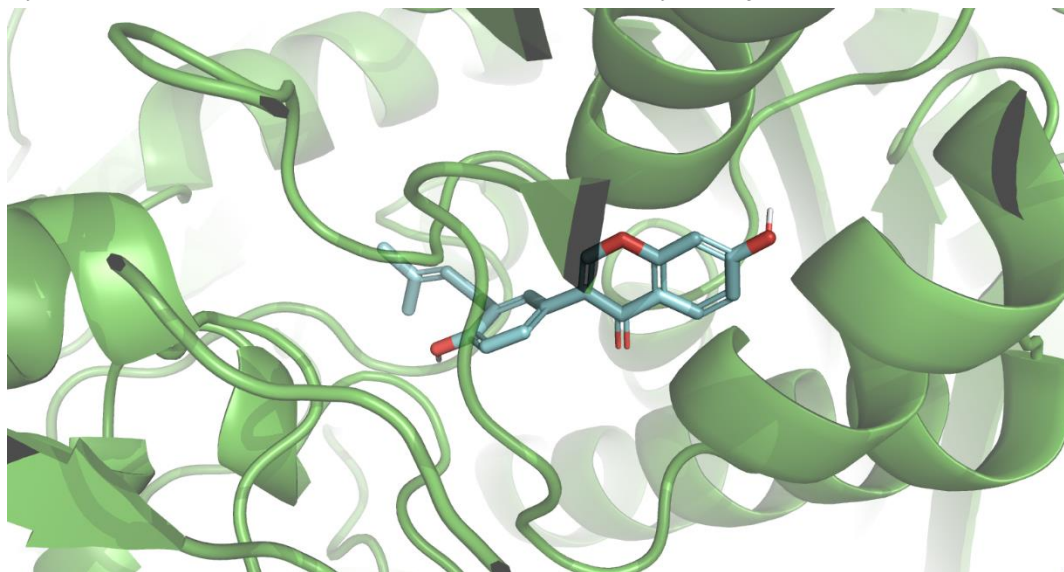
En este estudio se realizaron dos tipos de cálculos: primero se realizó un cálculo de *Blind Docking* del ligando frente a la proteína diana utilizando tanto el algoritmo de AutoDock Vina y Lead Finder (Figura V.1). Y posteriormente, se hizo un consenso de ambos algoritmos. La pose de mayor relevancia fue utilizada para realizar una simulación de 100 ns de Dinámica Molecular utilizando la mejor pose predicha por el *Blind Docking*.

Los resultados obtenidos por el *Blind Docking* consenso mostraron que había tres clústeres que tenían un *Docking score* similar, entre -11,92 y -11,18 kcal/mol. Se seleccionó el tercer clúster debido a que al buscar en la bibliografía se observó que este clúster era el sitio activo de la proteína. En esta zona se producía la interacción con la quercetina (143). El ligando además tenía interacciones predichas con residuos clave de la proteína (la fenilalanina 914 y el ácido glutámico 1261).

Los resultados de la simulación de Dinámica Molecular mostraron que la proteína alcanzaba una conformación estable, teniendo un valor máximo de RMSD de 2.7 Å. El ligando a su vez tenía un valor bajo de RMSD (3.6 Å), y es capaz de interactuar con la fenilalanina 798 y la arginina 912. Todo ello, es una evidencia

de que la estabilidad de la xantina oxidasa con la neobavaisoflavona es alta, por lo que la neobavaisoflavona podría ser un inhibidor de esta proteína.

Figura V.1. Blind Docking de la neobavaisoflavona contra la Xantina Oxidasa. Se representa la unión de la neobavaisoflavona, en color azul, junto con la Xantina Oxidasa, en color verde; unión predicha mediante el software Lead Finder. La visualización se ha obtenido mediante el software Pymol.



Los resultados obtenidos por el grupo del Dr Boulebd mostraron como el mecanismo de transferencia más favorable para la neobavaisoflavona era el FHT (*formal hydrogen transfer*) tanto para el radical hidroperoxilo ($\text{HOO}\cdot$) como para la molécula de dióxido de nitrógeno (NO_2).

Además de la neobavaisoflavona también se ha testado el uso de los ácidos 5-cafeoilquínico y 5-feruloilquínico. Al igual que en el caso anterior estos compuestos se encuentran en las plantas, principalmente en el café. Ambos ácidos tienen propiedades antioxidantes y al igual que el resto de los compuestos fenólicos contienen múltiples grupos hidroxilos que permiten eliminar los radicales libres. Además, ya hay estudios que muestran como ambos ácidos tienen propiedades antioxidantes y antiinflamatorias (144–146). Respecto al potencial de ambos compuestos para inhibir a la Xantina Oxidasa, el ácido 5-cafeoilquínico fue testado mediante técnicas *in vitro* junto con otros siete compuestos fenólicos; otros estudios mostraron como diversos compuestos fenólicos eran capaces de inhibir a esta

enzima (147,148). Por todo ello, en este trabajo se ha estudiado como ambos ácidos puedan regular la actividad de la xantina oxidasa (XO).

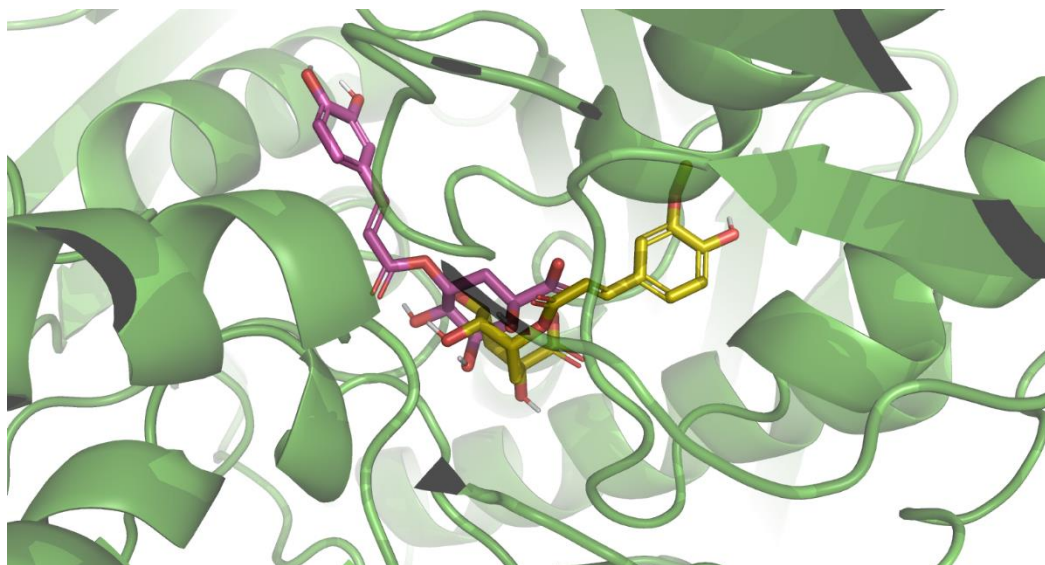
Los cálculos realizados para estos dos compuestos fueron similares a los del caso anterior. Se realizaron dos cálculos de *Blind Docking* contra la Xantina Oxidasa utilizando tanto los software AutoDock Vina y Lead Finder usando el software MetaScreener tanto para los cálculos de *Docking* individuales como para el cálculo del consenso. Por último, la mejor pose predicha por cada uno de los compuestos fue utilizada para realizar un cálculo de 100 ns de simulación de Dinámica Molecular utilizando Desmond como motor de dinámica.

Los resultados del *Blind Docking* consenso mostraron que el *Docking score* de ambos compuestos fue similar, en concreto cada compuesto tenía tres clústeres con un *Docking score* sin diferencias significativas. El *Docking score* del ácido 5-cafeoilquínico estaba entre -11,21 y -9,94 kcal/mol y en el caso del ácido 5-feruloilquínico su *score* variaba entre -10,12 a -9,47 kcal/mol. El primer clúster de los tres seleccionados es el sitio activo de la Xantina Oxidasa, es el sitio donde se encuentra la quercetina de forma natural (Figura V.2.). La interacción con la enzima más importante era la que se producía entre los ligandos y el ácido glutámico en la posición número 1261. Esta interacción se producía tanto con ambos compuestos como con la quercetina.

Los resultados de la simulación de Dinámica Molecular mostraron cómo la proteína alcanzaba rápidamente la estabilidad, mostrando un valor máximo de RMSD de 2,7 Å. Ambos ligandos tenían una gran estabilidad a lo largo de la trayectoria con valores máximo de 4,5 Å (ácido 5-cafeoilquínico) y 3,6 Å (ácido 5-feruloilquínico). Respecto a las interacciones que ocurrían entre el ácido 5-cafeoilquínico y la Xantina Oxidasa, estas eran con el ácido glutámico 1261 mediante interacciones iónicas, con las fenilalaninas 914 y 1009, la valina 1011 y la alanina 1078 mediante interacciones hidrofóbicas y con las glutaminas 767 y 1040, la treonina 1010, la alanina 1078 y las serinas 1080 y 1082 mediante interacciones polares. Mientras que con el ácido 5-feruloilquínico las interacciones iónicas se produjeron con la arginina 912, la lisina 1045 y los ácidos aspárticos 1191 y

glutámico 1261, y las interacciones polares con la treonina 1077 y las serinas 1080 y 1082. Estas evidencias sugieren la posibilidad de que los ácidos 5-cafeoilquínico y 5-feruloilquínico podrían ser inhibidores de la Xantina Oxidasa.

Figura V.2. Blind Docking de los ácidos del café contra la Xantina Oxidasa. Se representa la unión del ácido 5-cafeoilquínico en color magenta y del ácido 5-feruloilquínico en color amarillo, junto con la Xantina Oxidasa, en color verde; unión predicha mediante el software Lead Finder. La visualización se ha obtenido mediante el software Pymol.



En este trabajo el resultado experimental del grupo colaborador realizado por el Dr. Boulebd consistió en evaluar las constantes de velocidad del radical hidroperoxilo. Estos experimentos mostraron como el ácido 5-cafeoilquínico producía la estabilización mediante el mecanismo HAT (*hydrogen atom transfer*) mientras que el ácido 5-feruloilquínico lo hacía mediante el mecanismo SET (*single electron transfer*).

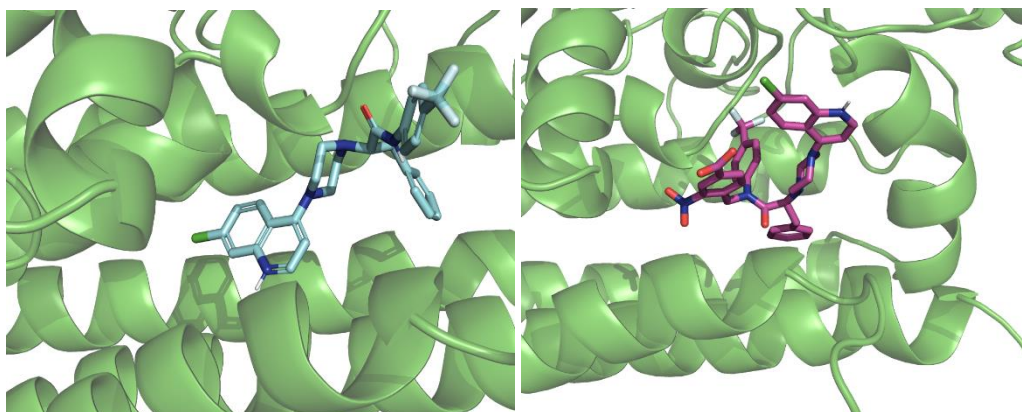
5.2.2. Quimioresistencia a tumores

La indolamina 2,3-dioxigenasa 1 (IDO1) es una enzima que participa en la degradación del triptófano mediante la vía de la quinurenina. Esta enzima además aumenta su expresión en procesos inflamatorios, y en tipos celulares como los macrófagos y las células dendríticas. Además, se ha observado que existe una sobreexpresión de esta enzima en algunos tipo de cáncer, diversos estudios

muestran como en estos cánceres se produce una disminución del número de células inmunitarias (149), principalmente NK (*natural killer*) y linfocitos T, debido a la sobreexpresión de IDO1 (150). En los últimos años, se han investigado diversos inhibidores de esta enzima como el epacadostat (151), el linrodostat (152) y el indoximod (153). Sin embargo, se ha observado que la mayoría de IDO1 se encuentra en su forma libre del grupo hemo (apo-IDO1) (154). Por ello, inhibir esta enzima, tanto en su forma de apoproteína como en su forma normal, es una de las estrategias para el tratamiento de esta enfermedad.

El compuesto elegido en este estudio para ver su actividad inhibitoria es la 4,7-dicloroquinona. Para ello, primero se realizó un cálculo de *Blind Docking* utilizando tanto el algoritmo de AutoDock Vina como el de Lead Finder (Figura V.3.), y posteriormente, se realizó un cálculo de consenso entre ambos algoritmos. Por último, se ha realizado una simulación de Dinámica Molecular con una duración de 100 ns.

Figura V.3. *Blind Docking* de la 4,7-dicloroquinona contra la indolamina 2,3-dioxigenasa 1. Se muestra la forma activa de la 4,7-dicloroquinona (color azul) y su forma preactiva (color magenta), la indolamina 2,3-dioxigenasa 1 está en color verde, el software utilizado fue Lead Finder. La visualización se ha obtenido mediante el software Pymol.



Los resultados obtenidos en este trabajo mostraron que el primer clúster es el sitio activo de la proteína IDO1. Las principales interacciones son producidas por puentes de hidrógeno o por interacciones de tipo hidrofóbicas. Por otro lado, los resultados de la dinámica mostraron cómo la proteína tenía una estabilidad media,

alcanzando valores de RMSD de 7,2 Å, mientras que el ligando si alcanzó una estabilidad mayor con valores de RMSD cercanos a 4 Å. Al analizar la estabilidad local mediante el RMSF, se observó que la zona con menos estabilidad se encontraba cerca a la parte del carbono terminal de la proteína. El compuesto interacciona con la proteína mediante los siguientes residuos: tirosina 113, alanina 198, fenilalanina 202, isoleucina 205, fenilalanina 214, fenilalanina 258 e histidina 334. Gracias a todos estos resultados se puede concluir que el 4,7-dicloroquinona podría ser un buen candidato para inhibir la indolamina 2,3-dioxigenasa 1.

Los resultados obtenidos por el grupo colaborador del Dr. Manna realizó estudios de actividad enzimática, obteniéndose un valor de Kd de 6 µM con la apoproteína; aunque tenía una baja afinidad por la proteína sola (Kd= 291 µM). También, se midió la capacidad inhibitoria de la 4,7-dicloroquinona y se obtuvo un valor de IC50 del orden de 100 nM. Todo esto hace que la 4,7-dicloroquinona sea un buen candidato para seguir con ensayos *in vivo*.

Otro grupo de proteínas que son importantes en la resistencia a los fármacos son las topoisomerasas. Estas enzimas se encargan de separar las hebras de ADN durante el proceso de replicación celular (155). Dentro de todas las topoisomerasas se encuentra la Topoisomerasa II α , la cual esta enzima es clave en procesos celulares como la transcripción y la recombinación (156,157). Lamentablemente, la desregulación de esta enzima puede producir enfermedades, un claro ejemplo es su sobreexpresión en células tumorales (158,159).

El café tiene una gran cantidad de compuestos polifenólicos, además de la cafeína, y se ha reportado que estos compuestos tienen propiedades anticancerígenas (160,161). Dentro de estos compuestos polifenólicos se encuentran los ácidos cafeoilquínicos, en este trabajo se utilizaron los ácidos ferúlicos, cafeico, 3-O-cafeoilquínico, 4-O-cafeoilquínico, 5-O-cafeoilquínico, 3,4-di-O-cafeoilquínico; 3,5-di-O-cafeoilquínico;4,5-di-O-cafeoilquínico; así como la cafeína.

En este caso se realizó una colaboración con el grupo de la Grażyna Budryn de la universidad de Lodz de Polonia (<https://p.lodz.pl/en/research/most-prominent-scientists/prof-grazyna-budryn>). Para ello, se realizaron unos *Dockings* individuales de cada uno de los compuestos previamente mencionados, y una serie de *Dockings* sucesivos de los ácidos cafeoilquínicos. El software utilizado fue Lead Finder y la proteína utilizada fue obtenida de PDB con el código 1ZXM y procesada con la suite de Maestro en formato mol2. Las interacciones de los *Dockings* individuales fueron evaluadas mediante el software PoseView.

Respecto a los *Dockings* sucesivos se realizaron dos combinaciones de 3 compuestos cada una. La primera con los derivados monosustituidos del ácido quínico (3-O-cafeoilquínico, 4-O-cafeoilquínico y 5-O-cafeoilquínico) y la segunda con los derivados disustituidos del ácido quínico (3,4-di-O-cafeoilquínico; 3,5-di-O-cafeoilquínico y 4,5-di-O-cafeoilquínico). En este caso se observó que no había diferencias notables entre los compuestos monosustituidos (con valores finales de *Docking score* que variaba entre -47,61 a -44,52 kJ/mol; Figura V.4.), mientras que sí había una diferencia entre los compuestos disustituidos (siendo más desfavorable la interacción inicial del ácido 3,5-di-O-cafeoilquínico con -48,70 kJ/mol respecto a los -54,64 kJ/mol del 3,4-di-O-cafeoilquínico y los -56,53 kJ/mol del 4,5-di-O-cafeoilquínico; Figura V.5.).

Respecto a las interacciones establecidas por cada compuesto, los resultados de *Docking* mostraron como el ácido 4-O-cafeoilquínico era el que más interacciones establecía, mientras que el ácido 3,5-di-O-cafeoilquínico era el que menos interaccionaba con la proteína; todas las interacciones están resumidas en la tabla V. 4. Destacan la asparagina 120, la tirosina 165 y la glicina 166 que establecían interacciones con 5 de los 6 compuestos.

Tabla V.4. Resumen de las interacciones entre los derivados del ácido cafeoilquínico y la Topoisomerasa II α . Estas interacciones han sido calculadas mediante el software PLIP tras la pose obtenida por el algoritmo de Lead Finder.

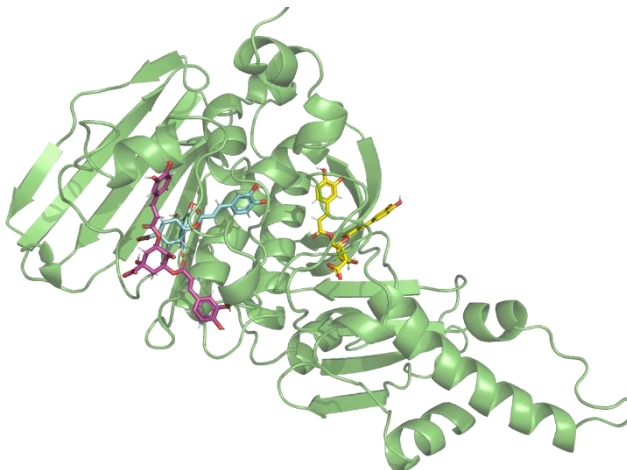
Compuesto	Residuos	Docking score (kJ/mol)
Ácido 3-O-cafeoilquínico	Asn94, Asn120, Asn150, Arg162, Gly164, Tyr165, Gly166 y Lys168	-43,96
Ácido 4-O-cafeoilquínico	Asn91, Asn120, Ile141, Ser148, Arg162, Gly164, Tyr165, Gly166, Ala167 y Lys168	-44,38
Ácido 5-O-cafeoilquínico	Asn94, Asn120, Asn150, Arg162, Gly164, Tyr165, Gly166, Lys168 y Thr215	-42,70
Ácido 3,4-di-O-cafeoilquínico	Asn91, Asn120, Ile141, Ser149, Asn150, Tyr165 y Gly166	-51,33
Ácido 3,5-di-O-cafeoilquínico	Asn95, Asn120, Asp126 y Asp147	-49,40
Ácido 4,5-di-O-cafeoilquínico	Glu87, Asn91, Tyr165, Gly166 y Gln376	-57,35

Los resultados obtenidos por el grupo colaborador de la Dra. Budryn, utilizando la calorimetría isoterma de titulación, mostraron como la digestión del café en el aparato digestivo disminuía de forma considerable la afinidad de los componentes fenólicos del café por la topoisomerasa II.

Figura V.4. Dockings sucesivos de los derivados monosustituidos del ácido cafeoilquínico contra la Topoisomerasa II α . La unión de los ácidos 3-O-cafeoilquínico (azul), 4-O-cafeoilquínico (amarillo) y 5-O-cafeoilquínico (magenta), junto con la Topoisomerasa II α , en color verde, predicho por el Docking mediante el software Lead Finder. La visualización se ha obtenido mediante el software Pymol. El orden de los Docking fue 3-O-cafeoilquínico, 4-O-cafeoilquínico y 5-O-cafeoilquínico.



Figura V.5. Dockings sucesivos de los derivados disustituidos del ácido cafeoilquínico contra la Topoisomerasa II α . La unión de los ácidos 3,4-di-O-cafeoilquínico (amarillo), 3,5-di-O-cafeoilquínico (magenta) y 4,5-di-O-cafeoilquínico (azul), junto con la Topoisomerasa II α , en color verde, predicho por el Docking mediante el software Lead Finder. La visualización se ha obtenido mediante el software Pymol. El orden de los Docking fue 4,5-di-O-cafeoilquínico; 3,4-di-O-cafeoilquínico y 3,5-di-O-cafeoilquínico.



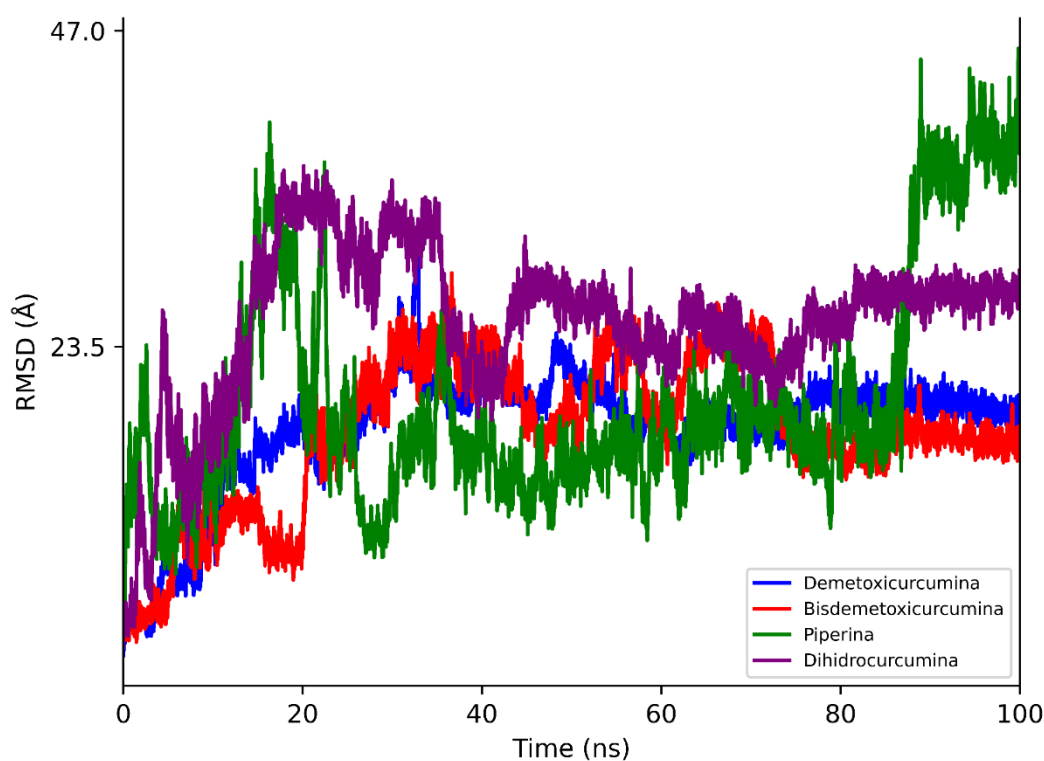
5.2.3. Metástasis y angiogénesis

Además de las especies reactivas de oxígeno y los problemas ocasionados por la resistencia tumoral, otra de las vías de actuación para combatir el cáncer es solucionar los problemas ocasionados tanto por la metástasis tumoral, como por la angiogénesis en las zonas cercanas al tumor. Uno de los grupos proteicos más importantes en metástasis es la superfamilia de lectinas, estas proteínas están involucradas en múltiples vías de señalización. Dentro de esta superfamilia de lectinas se encuentra la lectina CLEC-2 o CELC-1b, la cual está involucrada tanto en cáncer como en trombosis, aunque solamente se sabe su función detallada en el caso del cáncer (162). Además, se ha observado como esta lectina está relacionada con la metástasis de algunos tumores, concretamente en el carcinoma hepatocelular y en el cáncer de colon (163). La lectina CLEC-2 se une a la podoplanina (PDPN), esta unión provoca un aumento en la progresión de los tumores, así como en la metástasis y en la angiogénesis (164). Estas evidencias, junto con los estudios que muestran como inhibir la CLEC-2 en ratones aumenta la supervivencia de los animales, sugieren que esta proteína es una buena candidata para tratar la metástasis tumoral (165).

En este caso los cálculos de *Docking* fueron realizados por el grupo colaborador del Dr. Ramaswamy Krishnaraj de la Universidad Dambi Dollo de Etiopía. Los ligandos utilizados fueron derivados de la curcumina y se realizó un cálculo de *Docking* usando AutoDock. Nuestro grupo realizó las simulaciones de Dinámica Molecular utilizando el motor Desmond mediante la interfaz gráfica de Maestro, el tiempo de trayectoria simulado fue de 100 ns. Debido a la baja afinidad de los curcuminoídes con esta lectina solamente 4 compuestos fueron seleccionados para los cálculos de Dinámica Molecular. Los resultados de las simulaciones mostraron como la lectina tenía una gran estabilidad a lo largo de las 4 simulaciones (el valor máximo de RMSD nunca fue superior a los 3 Å). Sin embargo, los ligandos tenían una alta inestabilidad llegando a alcanzar valores de RMSD de 45 Å (Figura V.6.). Los ligandos interaccionan de forma constante con la arginina 107, la glicina 118 y la tirosina 213. Mientras que interaccionan de forma irregular con la arginina 118, la lisina 150, la isoleucina 156, la histidina 199, la asparagina 210 y la tirosina

213. Con todos estos resultados se concluyó que los curcuminoides que mejor interaccionan con la lectina eran la bisdemetoxicurcumina y la demetoxicurcumina.

Figura V.6. Evolución de RMSD de los compuestos de interés frente a la lectina CLEC-2. Se muestran los cuatro ligandos estudiados durante la simulación a lo largo de los 100 ns junto con la lectina CLEC-2. La leyenda muestra el color asignado a cada ligando: azul (demetoxicurcumina), rojo (bisdemetoxicurcumina), verde (piperina) y morado (dihidrocurcumina).



Una de las principales vías de metástasis se produce cuando el tumor produce angiogénesis y consigue ser capaz de desplazarse recorriendo el resto del organismo. Por ello, inhibir la angiogénesis cercana a la zona tumoral podría ser una vía para inhibir la progresión del cáncer (166). Tal y como se ha mencionado previamente en esta tesis, los compuestos polifenólicos son prometedores candidatos en el tratamiento del cáncer. La cúrcuma es una planta que tiene altos niveles de compuestos polifenólicos tales como la curcumina, la demetoxicurcumina y la bisdemetoxicurcumina (167). Recientes estudios muestran como el uso en ensayos *in vitro* hechos en ratones de estos tres compuestos mostraban una disminución en la proliferación y migración de células endoteliales mediante con el receptor del factor de crecimiento endotelial vascular-2 (VEGFR2)

(168). En este estudio realizado en colaboración con investigadores del CEBAS del grupo del Dr. Juan Antonio Giménez-Bastida (http://www.cebas.csic.es/dep_spain/alimentos/calidad/aliment_lineas.html), se realizó un análisis de la interacción entre estos compuestos curcuminoides con VEGFR2 a concentraciones fisiológicas para evaluar sus propiedades antiangiogénicas.

En este estudio se realizó un *Blind Docking* de estos tres curcuminoides: la curcumina, la demetoxicurcumina y la bisdemetoxicurcumina. La estructura de estos compuestos fue obtenida en formato SDF de la base de datos PubChem y la proteína fue obtenida de la base de datos PDB con el código 6GQO. Se realizaron cálculos de *Docking* mediante los software AutoDock Vina y Lead Finder mediante el uso de la herramienta MetaScreener. Posteriormente, se realizó un cálculo de *Blind Docking* consenso y se obtuvo que el metaclúster con mejor *Docking score* era el mismo que en el que se encontraba un inhibidor ya conocido de la proteína (Figura V.7.). Además, las interacciones obtenidas por cada *Docking* han sido resumidas en la Tabla V.5. Estos resultados tenían coherencia con los obtenidos por el grupo colaborador, dado que los tres curcuminoides inhibían la migración celular de forma directamente proporcional a la cantidad de ligando añadida al cultivo celular.

Figura V.7. *Docking* de los curcuminoides frente a VEGFR. Se representa la unión de la curcumina (Izq: Blanco, Dcha: Naranja), la demetoxicurcumina (Izq: Azul, Dcha: Morado) y la bisdemetoxicurcumina (Izq: Verde, Dcha: Amarillo), junto con el receptor del factor de crecimiento endotelial vascular-2, en color rosa, predicho por el *Docking* mediante los software AutoDock Vina (A) y Lead Finder (B). La visualización se ha obtenido mediante el software Pymol.

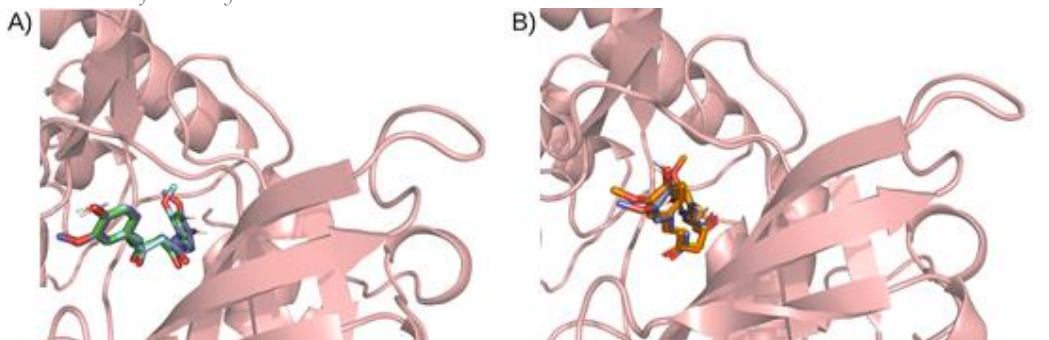


Tabla V.5. Resumen de las interacciones entre los curcuminoides y VEGFR2. Estas interacciones han sido calculadas mediante el software PLIP tras la pose obtenida por los algoritmos de AutoDock Vina y Lead Finder.

Compuesto	Residuos
Curcumina	Val848, Ala866, Lys868, Glu885, Leu889, Val899, Pro918, Cys919, Leu1035 y Asp1046
Demetoxicurcumina	Leu840, Val848, Ala866, Lys868, Glu885, Leu889, Val899, Val916, Pro918, Cys919, Leu1035, Ile1044, Asp1046 y Pro1047
Bisdemetoxicurcumina	Leu840, Val848, Ala866, Lys868, Glu885, Val899, Val916, Pro918, Cys919, His1026, Leu1035, Ile1044, Asp1046 y Pro1047

Los resultados obtenidos por el grupo colaborador del Dr. Giménez-Bastida mostraron cómo se producía tanto una inhibición de la formación de las estructuras tubulares, como de la migración celular, siendo la demetoxicurcumina y la bisdemetoxicurcumina las que mayor actividad mostraban. Aunque todavía se tiene que seguir investigando en profundidad el uso de estos compuestos y su implicación en el cáncer.

5.3. ALZHEIMER

En este apartado se aborda el objetivo 2.2. “Aplicación de métodos de computación de alto rendimiento en el Alzheimer”. Esta patología es una enfermedad neurodegenerativa producida por la acumulación de las placas beta amiloides. Esta enfermedad produce daños a nivel cerebral provocando demencia. Para tratar esta enfermedad todavía no hay cura ni tratamiento, por lo que es importante seguir investigando esta patología.

5.3.1. Acetilcolinesterasa

Una de las proteínas involucradas en la enfermedad de Alzheimer es la acetilcolinesterasa (AChE). Esta proteína es la encargada de la degradación de la acetilcolina en acetato y colina. La acetilcolina es un neurotransmisor que está involucrado en la demencia severa. Por ello, una de las líneas de investigación para el tratamiento de la enfermedad de Alzheimer es la inhibición de la acetilcolinesterasa con fármacos como la galantamina o el donepezilo. Sin embargo, estos medicamentos tienen muchos efectos secundarios. Debido a este problema, se están buscando nuevos inhibidores de origen natural que tengan menos efectos adversos, estos compuestos son flavonoides, ácidos fenólicos y carotenoides (169). Se ha observado también que la cafeína y sus derivados ayudan a inhibir la hidrólisis de la acetilcolina (170). Otros estudios mostraron que las lactonas sesquiterpénicas tenían un alto potencial para ser inhibidores de la AChE (171). Dentro de este grupo de compuestos, se encuentran dos provenientes de la achicoria: la lactucina y la lactupicrina.

Por ello, desde el grupo de la doctora Grazyna Budryn en la Universidad de Lodz en Polonia, se estableció una colaboración para buscar cómo estas dos lactonas podrían ser potenciales inhibidores de la acetilcolinesterasa. Para este estudio se obtuvo de la PDB la estructura obtenida por rayos X de AChE con el código 1EVE. Y se utilizó el software Lead Finder a través de MetaScreener para predecir la actividad de la lactucina y la lactupicrina.

Los resultados de *Docking* mostraron cómo ambos compuestos tenían un valor de *Docking score* siendo -47,90 kJ/mol con la lactupicrina y -37,64 kJ/mol con la lactucina (Figura V.8., Figura V.9.). Además, las interacciones más importantes entre la proteína y cada ligando se han resumido en la Tabla V.7.

Figura V.8. Docking sucesivo de la lactucina y la lactupicrina frente a la AChE. Se representa la unión de la lactupicrina (morado) y la lactucina (verde) junto con la Acetilcolinesterasa, en color naranja, predicho por el Docking mediante el software Lead Finder. La visualización se ha obtenido mediante el software Pymol. El orden de los Docking fue lactucina y lactupicrina.

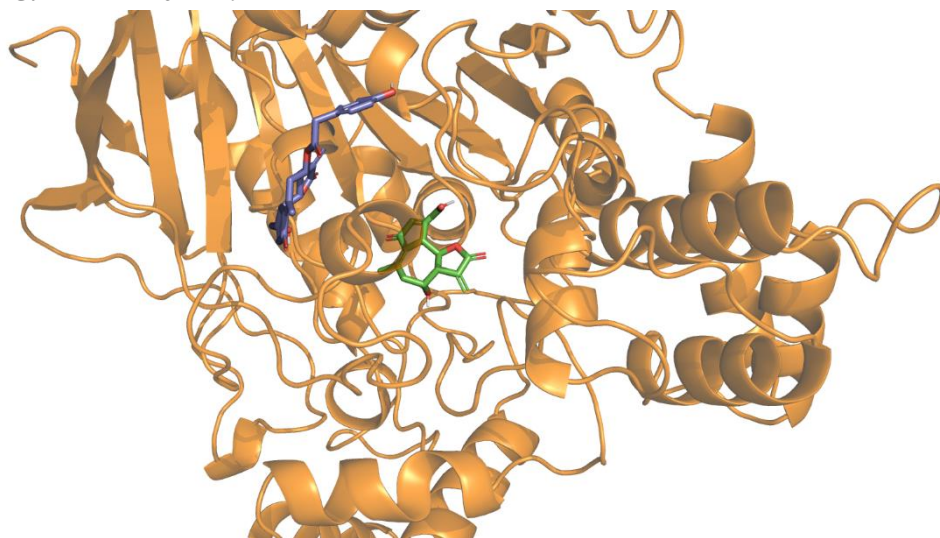
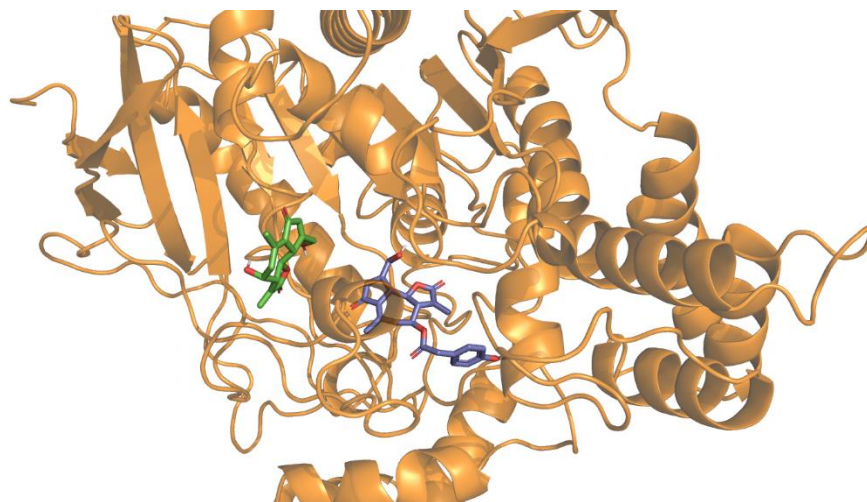


Figura V.9. Docking sucesivo de la lactupicrina y la lactucina frente a la AChE. Se representa de la lactupicrina (morado) y la lactucina (verde) junto con la Acetilcolinesterasa, en color naranja, predicho por el Docking mediante el software Lead Finder. La visualización se ha obtenido mediante el software Pymol. El orden de los Docking fue lactupicrina y lactucina.



Estos resultados junto con los resultados experimentales obtenidos por el grupo de la Dra. Budryn muestran como ambas moléculas tienen un potencial inhibitorio para la AChE, teniendo un valor de IC₅₀ de 1,74 μ M para la lactucopirina y de 1,32 μ M para la lactucina. Estos resultados sugieren que tanto la lactucopirina como la lactucina podrían ser candidatos para ser futuros tratamientos para paliar la enfermedad de Alzheimer.

Tabla V.6. Resumen de las interacciones entre la lactucopirina y la lactucina contra la AChE. Estas interacciones han sido calculadas mediante el software PLIP tras la pose obtenida por el algoritmo de Lead Finder.

Residuo	Orden de los compuestos	Compuesto (s) con el (los) que interacciona
MET83	Lactupirina - Lactucina	Lactucina
	Lactucina-Lactupirina	Lactupirina
TRP 84	Lactupirina - Lactucina	Lactupirina
	Lactucina-Lactupirina	Lactucina y Lactupirina
SER128	Lactupirina - Lactucina	Lactupirina
	Lactucina-Lactupirina	Lactupirina
ASP128	Lactupirina - Lactucina	Lactucina
	Lactucina-Lactupirina	Lactupirina
VAL129	Lactupirina - Lactucina	Lactucina
	Lactucina-Lactupirina	Lactupirina

5.3.2. Butirilcolinesterasa

Otra de las enzimas importantes en la hidrólisis de la acetilcolina es la butirilcolinesterasa (BChE), esta enzima produce la degradación de la acetilcolina cuando no hay suficientes niveles de acetilcolinesterasa. Al igual que en el trabajo anterior está enzima es clave en la enfermedad de Alzheimer. Por ello, inhibirla es una posible vía para tratar esta enfermedad. Ya hay estudios que muestran la búsqueda de compuestos naturales que carezcan de toxicidad y puedan ser candidatos para tratar esta patología (172,173). Estudios recientes sugieren que el consumo moderado de café tiene una correlación con la ralentización del deterioro cognitivo (174,175). Se ha observado que hay componentes en el café, nutracéuticos, que pueden ser inhibidores de la BChE. Al igual que en el trabajo anterior, el grupo de la Dra Grazyna Budryn de la Universidad de Lodz en Polonia fue el que realizó la parte experimental del trabajo. Además, este grupo de investigadores ya estudió como el extracto del café podía tener efectos beneficiosos, por lo que es vital estudiar de forma aislada cada uno de los componentes de este extracto.

En todos los casos se ha visto que el primer compuesto es el que más afinidad tiene con la proteína. Mientras que el resto de los compuestos adquiere una conformación con menor afinidad. Además, tal y como se puede observar en las Figuras V.10 y V.12, los seis compuestos eran capaces de interactuar con múltiples residuos de la BChE; produciéndose de esta manera una gran estabilidad entre la proteína y los ligandos. Por todo ello, el estudio concluyó que estos compuestos eran candidatos con posibilidades para paliar los síntomas producidos por la enfermedad de Alzheimer, aunque es necesario realizar estudios de cultivos celulares en tejidos humanos para confirmar la hipótesis.

Figura V.10. Dockings sucesivos de los derivados monosustituidos del ácido cafeoilquínico frente a la BChE. Se representan las interacciones producidas entre cada ligando y la Butirilcolinesterasa en la secuencia de Docking ácido 5-O-cafeoilquínico (A) -> ácido 4-O-cafeoilquínico (B) -> ácido 3-O-cafeoilquínico (C). La visualización se ha obtenido mediante el software Maestro.

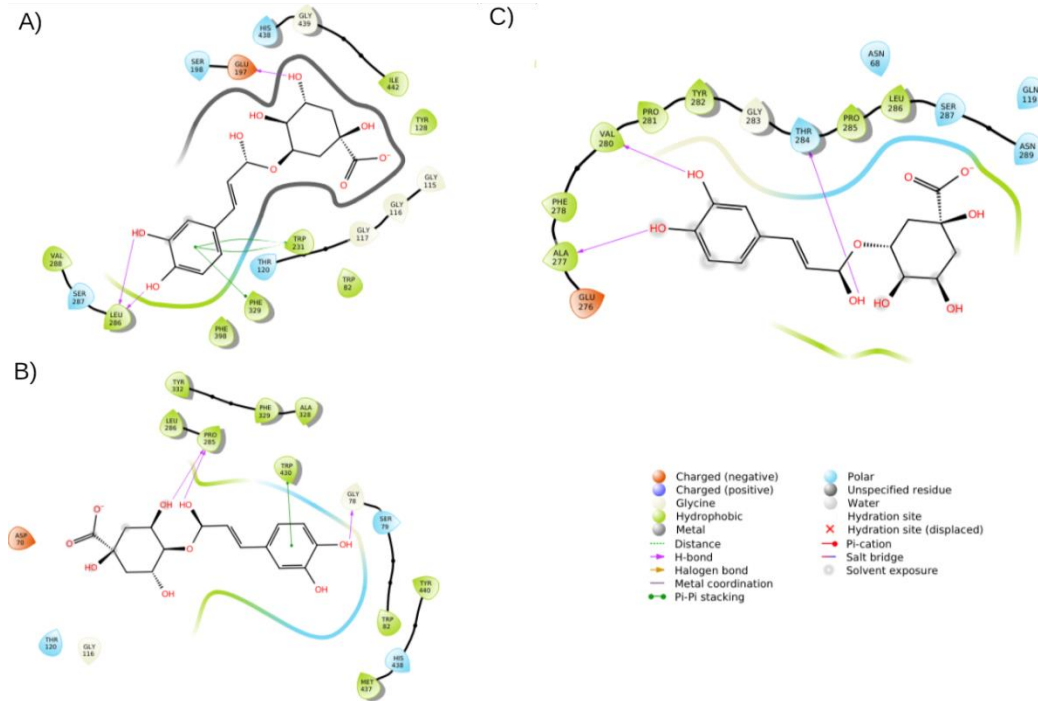


Figura V.11. Dockings sucesivos de los derivados monosustituidos del ácido cafeoilquínico frente a la BChE. Se representa la unión de los ácidos 3-O-cafeoilquínico (azul), 4-O-cafeoilquínico (amarillo) y 5-O-cafeoilquínico (magenta), junto con la Butirilcolinesterasa, en color rosa, predicho por el Docking mediante el software Lead Finder. La visualización se ha obtenido mediante el software Pymol. El orden de los Docking fue 5-O-cafeoilquínico, 4-O-cafeoilquínico y 3-O-cafeoilquínico.

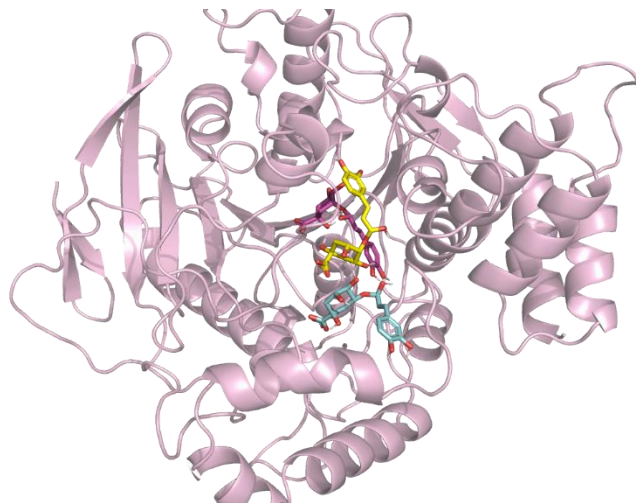


Figura V.12. Dockings sucesivos de los derivados disustituidos del ácido cafeoilquínico frente a la BChE. Se representan las interacciones producidas entre cada ligando y la Butirilcolinesterasa en la secuencia de Docking ácido 4,5-di-O-cafeoilquínico (A) -> 3,4-di-O-cafeoilquínico (B) -> 3,5-di-O-cafeoilquínico (C). La visualización se ha obtenido mediante el software Maestro.

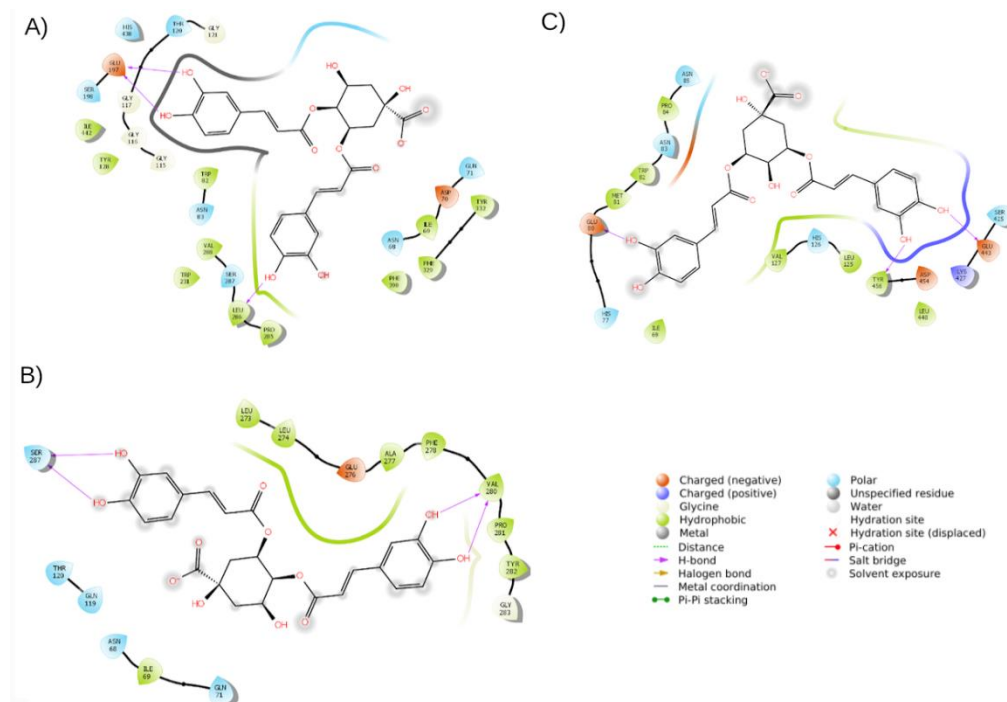
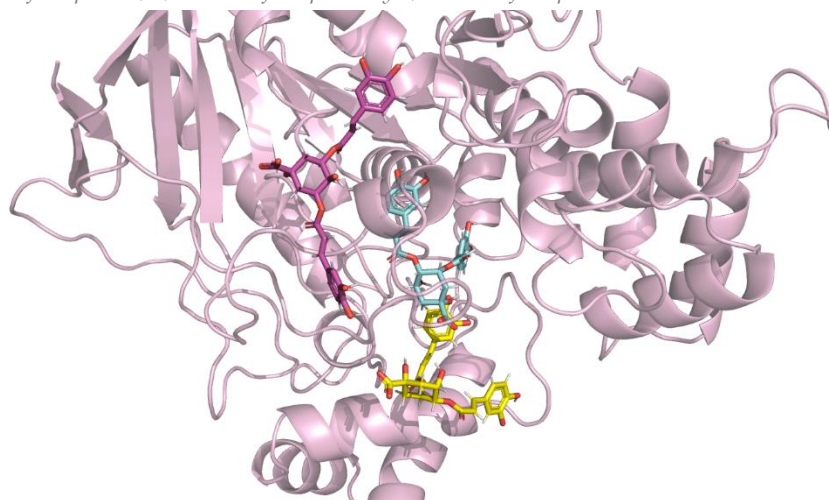


Figura V.13. Dockings sucesivos de los derivados disustituidos del ácido cafeoilquínico frente a la BChE. Se representa la unión de los ácidos 3,4-di-O-cafeoilquínico (amarillo), 3,5-di-O-cafeoilquínico (magenta) y 4,5-di-O-cafeoilquínico (azul), junto con la Butirilcolinesterasa, en color rosa, predicho por el Docking mediante el software Lead Finder. La visualización se ha obtenido mediante el software Pymol. El orden de los Docking fue 4,5-di-O-cafeoilquínico; 3,4-di-O-cafeoilquínico y 3,5-di-O-cafeoilquínico.



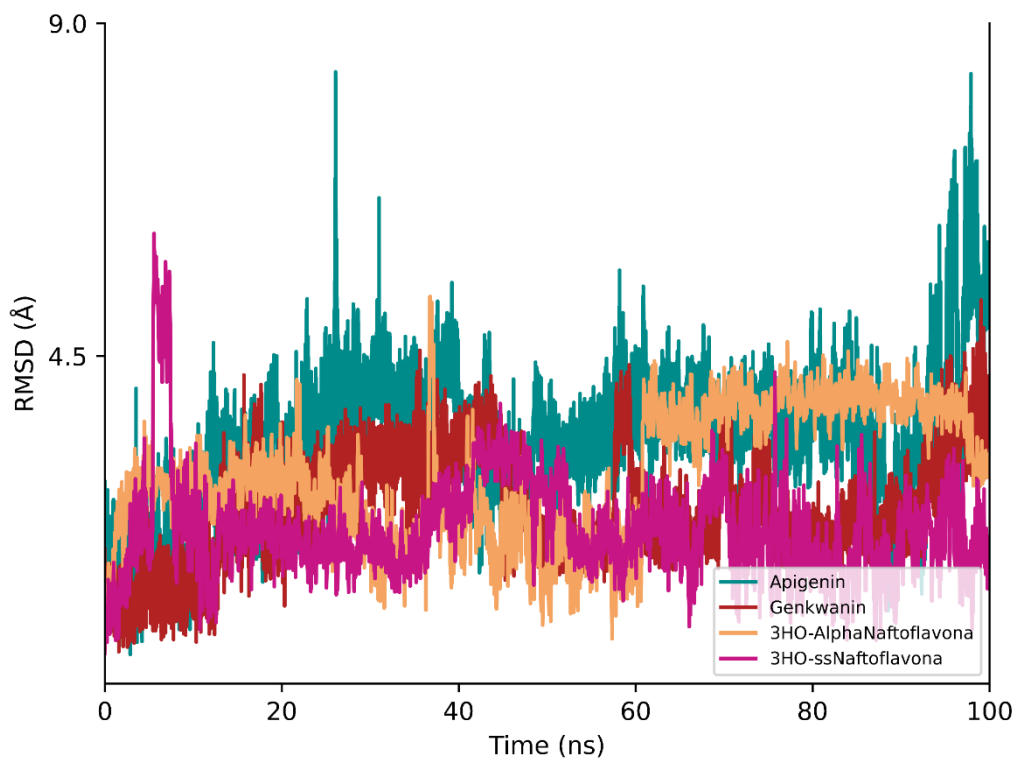
5.3.3. Monooxigenasas

Otra enzima importante involucrada en la enfermedad de Alzheimer es la quinurenina 3-monooxigenasa (KMO). Esta enzima participa en la degradación del triptófano mediante la vía de la quinurenina y está implicada en numerosas enfermedades (176,177). Un exceso de actividad de la enzima KMO genera un exceso de radicales libres que causan la muerte de las neuronas y provocan la enfermedad de Alzheimer (178). Sin embargo, todavía ningún estudio ha conseguido mostrar como los inhibidores de la KMO son efectivos para estas patologías, por ello es necesario seguir investigando esta enzima y sus inhibidores (179). En este trabajo se va a testar la actividad inhibitoria de determinados flavonoides contra esta enzima. Los flavonoides son conocidos por sus propiedades antimicrobianas, antivirales y antiinflamatorias (180,181). Por ello, son excelentes candidatos para inhibir la enzima KMO y tratar así la enfermedad de Alzheimer.

Para ello, se han realizado tanto ensayos enzimáticos como técnicas computacionales. En este caso, los cálculos de *Docking* fueron realizados por el grupo colaborador liderado por el Dr Redouane Rebai de la Universidad Mohamed Khider de Argelia (<https://univ-biskra.dz/index.php/fr/enseignants-par-facult%C3%A9-2/1271-enseignants-facult%C3%A9-des-sciences-exactes-et-des-sciences-de-la-nature-et-de-la-vie>). La proteína fue obtenida de la base de datos PDB con el código 4J31 y se usó el software AutoDock Vina para realizar el *Docking*. El grupo colaborador realizó un cálculo de *Virtual Screening* seleccionando finalmente cuatro compuestos para la realización del cálculo de Dinámica Molecular a 100 ns.

Los resultados de las simulaciones mostraron como la apigenina era el compuesto que menos estabilidad tenía con la enzima. Mientras que la 3-hidroxi-alfa-naftoflavona alcanzaba una conformación estable a los 60 ns de simulación, la 3-hidroxi-ss-naftoflavona tenía una conformación estable en toda la simulación, excepto durante los 40 y 60 ns y la genkwanina tenía una estabilidad a partir de los 10 primeros ns de la simulación (Figura V.14). Los residuos que mayor inestabilidad presentaban están entre la glicina 81 y la tirosina 96.

Figura V.14. Evolución de RMSD de los compuestos de interés frente a la KMO. Se muestran los valores de RMSD de los cuatro ligandos estudiados durante la simulación a lo largo de los 100 ns con la quinurenina 3-monooxigenasa. La leyenda muestra el color asignado a cada ligando: verde (apigenina), rojo (genkwantina), naranja (3-hidroxi-alfa-naftoflavona) y violeta (3-hidroxi-ss-naftoflavona).



Los resultados obtenidos por el grupo experimental del Dr. Rebai mostraron que los dos derivados de la naftoflavona eran inhibidores competitivos de la KMO, con unos valores de IC₅₀ de 16 μM (3-hidroxi-alfa-naftoflavona) y 19 μM (3-hidroxi-ss-naftoflavona). Mientras que la apigenina y la genkwantina eran inhibidores no competitivos, con valores de IC₅₀ de 22 μM (genkwantina) y 24 μM (apigenina). Sin embargo, aunque estos resultados sugieren que estos compuestos pueden ser una nueva vía de inicio para tratar la enfermedad de Alzheimer, se debe seguir investigando dado que sus valores de IC₅₀ siguen siendo demasiado elevados en comparación con el compuesto control Ro 61-8048 (5 μM).

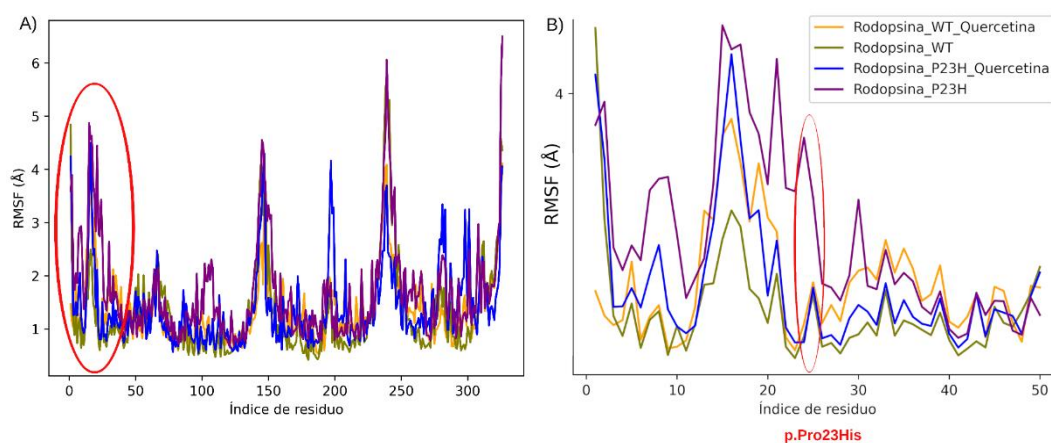
5.4. RETINITIS PIGMENTOSA

En este apartado se abordará el objetivo 2.3. “Aplicación de métodos de computación de alto rendimiento en la retinitis pigmentosa”. Esta patología es una enfermedad neurodegenerativa que provoca la pérdida de los fotorreceptores, siendo una de las causas más prevalentes de la ceguera en el ser humano (182,183). Esta enfermedad se produce por mutaciones en la rodopsina, lo que provoca una unión incorrecta entre esta proteína y el 11-cis-retinal, causando una degeneración de los fotorreceptores. La mutación de clase 2 más frecuente en la rodopsina que causa esta patología es la sustitución de la prolina 23 por una histidina (P23H). Esta mutación provoca una inestabilidad en los aminoácidos de la proteína (184). Esto produce una disminución del grosor de una de las capas de la retina, provocando ceguera. Por ello, evitar la inestabilidad en esta proteína podría ser clave para reducir la gravedad de la patología. Diversos estudios *in vitro* muestran cómo algunos flavonoides son capaces de mejorar el funcionamiento de la proteína con la mutación P23H; sin embargo, es necesario seguir investigando estos compuestos en modelos animales (185).

Respecto al uso de los flavonoides en ensayos *in vivo*, se ha observado que son capaces de estabilizar las vías de estrés del retículo endoplasmático (186) y tener efectos neuroprotectores en ratones que padecen retinitis pigmentosa (187). También se ha observado que estos compuestos evitan la degeneración de la retina, evitando de esta manera la muerte de los fotorreceptores (188). Por ello, se ha realizado este estudio en colaboración con la Dra. Beata Jastrzebska de la Universidad Case Western Reserve (<https://case.edu/medicine/pharmacology/people/beata-jastrzebska>) en Cleveland (Estados Unidos de América). En este estudio se observó como la quercetina (uno de los flavonoides más conocidos) era capaz de mitigar los efectos producidos por la mutación P23H, aumentando el grosor de la capa de la retina afectada por la mutación en el modelo murino utilizado por los colaboradores. Además, los resultados experimentales mostraron como la quercetina mejoraba la supervivencia tanto de los conos como de los bastones respecto al tratamiento con el vector y con la mirecitina.

Respecto a los experimentos computacionales realizados por nuestro grupo, se utilizó la proteína obtenida en la PDB cuyo código es 3CAP. La mutación en la proteína cristalizada fue generada por los colaboradores usando el software DUET (189) y validada usando PROCHECK (190). Se realizó un cálculo de *Blind Docking* utilizando Auto Dock Vina como software. Se seleccionó la mejor pose para realizar un cálculo de Dinámica Molecular de 100 ns insertando además la proteína en una membrana de POPC, la localización de la membrana fue obtenida de la base de datos UniProt (<https://www.uniprot.org/>). Se realizaron cuatro simulaciones: WT (*Wild type*) sin ligando, p.Pro23His sin ligando, p.Pro23His con ligando y WT con ligando. Posteriormente, se realizó un análisis de RMSD, RMSF y PSN.

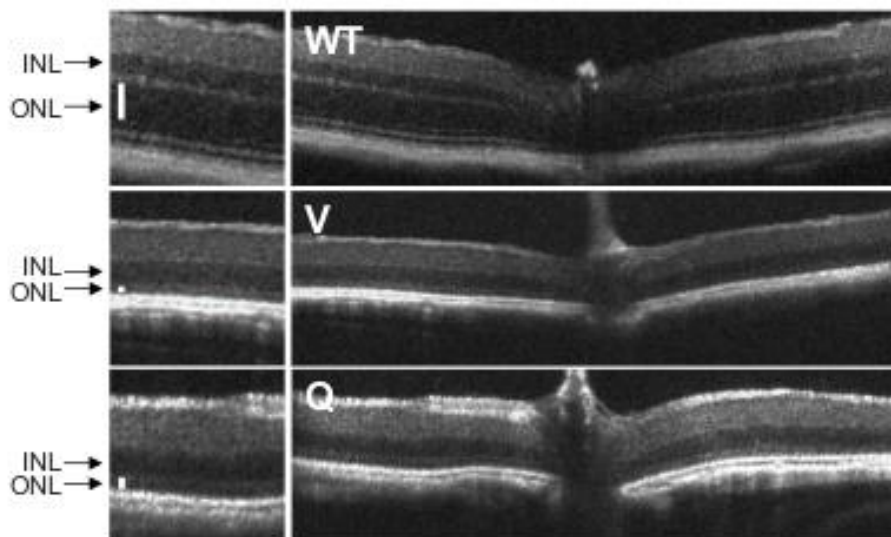
Figura V.15. Evolución de RMSF durante las simulaciones de la rodopsina. A) RMSF de la rodopsina WT tratada con quercetin (naranja) y sin ella (verde), y de la mutación p.Pro23His con quercetina (azul) y sin ella (morado). B) Vista ampliada del RMSF de los primeros 50 residuos, destacando la región mutada p.Pro23His.



Los resultados mostraron cómo la quercetina era capaz de recuperar la actividad de la proteína en el intervalo de tiempo simulado. Además, la quercetina no afectaba a las proteínas que no presentaban la mutación. Concretamente, la quercetina estabilizaba la zona de la mutación reduciendo el valor original de RMSF del residuo mutado de 4 a 1 Å, obteniendo un valor similar al obtenido en la proteína WT, tal y como se puede ver en la Figura V.15. Además, el análisis de PSN reveló como el principal clúster de residuos de la proteína WT se fragmentaba en la proteína con la mutación, pero se recuperaban muchos de estos residuos al producirse la unión de la quercetina con la rodopsina mutante.

Los resultados experimentales mostraron como el tratamiento con quercetina en las células NIH-3T3 con la mutación tenían una menor concentración de AMPcíclico, esto indicaba que había una mejora en la señalización del receptor. Además, los ratones que tenían la mutación y eran tratados con quercetina tenían un aumento significativo del grosor de la retina (Figura V.16.). Esto sugiere que la quercetina es un candidato como tratamiento para paliar la sintomatología provocada por la retinitis pigmentosa causada por esta mutación.

Figura V.16. Imágenes de la retina de ratón tomada mediante tomografía de coherencia óptica (OCT). Se muestran la capa nuclear interna (INL) y la capa nuclear externa (ONL) están señaladas mediante flechas. Las imágenes muestran los distintos tipos de ratones (WT: ratones sin mutación, V: ratones con la mutación p.Pro23His tratados con el vehículo y Q: ratones con la mutación p.Pro23His tratados con quercetina). La imagen está modificada de Ortega et. al, 2022.



5.5. CICATRIZACIÓN

En este apartado se abordará el objetivo 2.4., “Aplicación de métodos de computación de alto rendimiento en enfermedades relacionadas con la cicatrización”. Uno de los procesos clave para la cicatrización es la activación plaquetaria, si hay fallos en dicha activación se producen patologías como son las trombocitopenias o la enfermedad de von Willebrand. En la década de los 90 se descubrieron las oxipilinas vegetales, las cuales tienen una gran similitud estructural con las prostaglandinas. Estas oxipilinas han sido capaces de modular la respuesta inflamatoria asociadas a la disfunción endotelial (191). Las prostaglandinas tienen diversas funciones en el cuerpo humano, incluyendo la función plaquetaria y vascular. Algunas de las más relevantes son la prostaglandina D2 (PGD2) y la prostaglandina E2 (PGE2) (192). En este estudio nos centraremos en la PGE2, la cual tiene interacción con cuatro subtipos de los receptores EP: EP1, EP2, EP3 y EP4, que desencadenan vías de señalización acopladas a proteínas G (193). Los receptores EP3 y EP4 están presentes en la superficie de las plaquetas, el receptor EP3 tiene funciones protrombóticas y el EP4 funciones antitrombóticas (194). Además, los fitoprostano son isómeros geométricos de las prostaglandinas humanas, por lo que son buenos candidatos para regular la activación plaquetaria mediante el receptor EP3 (195).

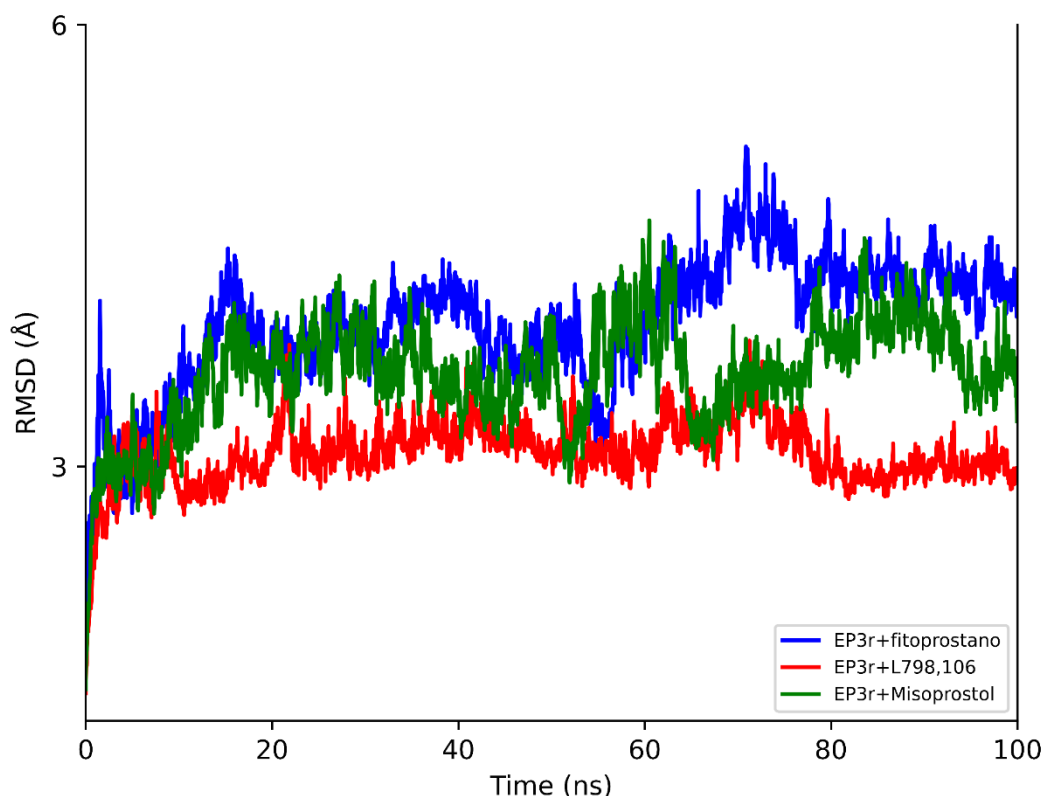
Por todo ello, se ha realizado un estudio para medir el efecto de un fitoprostano de interés en la activación plaquetaria. Este estudio se ha realizado en colaboración con el grupo liderado por la Dra Silvia Montoro García (<https://investigacion.ucam.edu/grupos/grupo/investigacion-preclinica-compuestos-bioactivos-farmacos/>) de la Universidad Católica de Murcia. En este estudio se comprobó los efectos de un fitoprostano, observando cómo se produce la activación plaquetaria en células vasculares humanas, además de predecir la interacción entre este fitoprostano y el receptor de EP3.

Los ensayos *in silico* se realizaron usando el receptor EP3 obtenido de la PDB cuyo código es 6M9T, además se utilizaron tanto antagonistas (L-798,106) como agonistas (misoprostol). Para ello, se realizó un cálculo de *Blind Docking* de los tres compuestos (los dos compuestos control y el fitoprostano a estudiar). Posteriormente, se realizaron una serie de cálculos de dinámica molecular cuya duración fue de 100 ns. Respecto a estas dinámicas es importante destacar que se realizaron simulando una membrana de POPC, la información sobre la proteína transmembrana fue obtenida de la base de datos UniProt.

Los resultados obtenidos por el *Docking* mostraron como los ligandos interaccionaban en la misma zona la cual era la zona donde se encontraba cristalizado el misoprostol en la PDB. Esta zona fue seleccionada para realizar las simulaciones. Los resultados de las simulaciones mostraron como tanto el misoprostol como el fitoprostano interaccionaban con residuos clave como la tirosina 114 y la arginina 333, estas interacciones se establecían con el grupo carboxilo de ambas moléculas. Sin embargo, el compuesto L-798,106 no era capaz de interaccionar con estos residuos y tenía unas interacciones más débiles; interaccionando principalmente con la treonina 106. Estos resultados junto con los resultados experimentales mostraron como el fitoprostano ayuda a estimular y a activar la agregación plaquetaria sin comprometer la viabilidad celular. La Figura V.17. muestra como los compuestos modificaban de forma distinta la conformación del receptor. Por todo ello, se concluye que este fitoprostano tiene potencial para ser un fármaco con actividad trombótica que pueda ser utilizado para enfermedades cardiovasculares.

Los resultados obtenidos por el grupo colaborador de la Dr. Montoro García mostraron como el fitoprostano 9-D1t-PhytoP, a una concentración de 100 nM, era capaz de aumentar los niveles de la P-selectina y la glucoproteína IV, de forma que promovía la activación plaquetaria. Respecto a la adhesión de las plaquetas a los leucocitos, se observó como este fitoprostano era capaz de activar esta adhesión al combinarse con sulprostona, aunque no era capaz de activar dicha adhesión por sí solo o en combinación con otros compuestos testados. Por último, también se observó que este fitoprostano, a una concentración de 300 nM, era capaz de activar la migración de las plaquetas de forma similar a la prostaglandina E2 (PGE2).

Figura V.17. Evolución de RMSD de los compuestos de interés frente al receptor de EP3. Se muestran los valores de RMSD del receptor de EP3 con los tres ligandos estudiados. La leyenda muestra el color asignado a cada complejo: azul (unido al fitoprostano a estudiar), rojo (unido a L-798,106), y verde (unido al misoprostol).



5.6. DIABETES

En este apartado se abordará el objetivo 2.5, “Aplicación de métodos de computación de alto rendimiento en la diabetes”. Esta patología es una enfermedad crónica que afecta a millones de personas en todo el planeta, en 2021 tenía una prevalencia del 6.1% (196). En la actualidad, el principal tratamiento de la diabetes es la inyección de insulina, sin embargo, recientemente se han empezado a aprobar otros fármacos con mayor especificidad por proteínas clave en esta enfermedad como el receptor del péptido-1 similar al glucagón (GLP-1), la dipeptidil peptidasa-4 (DPP4) y la α -glucosidasa, aunque muchos de estos compuestos siguen teniendo efectos secundarios adversos (197). Un ejemplo de este tipo de fármacos es la metformina, la cual se está utilizando para tratar la diabetes mellitus tipo II, aunque su mecanismo de acción aún es desconocido (198,199). Por todo ello, es necesario seguir trabajando en la búsqueda de nuevos compuestos que ayuden a tratar esta

patología. Uno de los compuestos que se está investigando es el carvacrol. El cual es un fenol monoterpeneo que se encuentra en hierbas aromáticas como el orégano y el tomillo (200). Este compuesto ya ha demostrado tener propiedades antibacterianas y anticancerígenas, gracias a la presencia de grupos hidroxilo en su estructura (201,202). Esto junto con los estudios que muestran como extractos de orégano (con un alto contenido en carvacrol) son capaces de inhibir la enzima DPP4, sugieren que el carvacrol es un buen candidato para tratar la diabetes (203,204).

Para la realización de este trabajo se ha colaborado con el Dr. Nail Besli de la Universidad de Ciencias de la Salud en Estambul (Turquía). El grupo colaborador utilizó la herramienta DIA-DB diseñada por nuestro grupo en 2016 (136), para localizar compuestos con potencial antidiabético. El uso de esta herramienta permitió encontrar el carvacrol como un compuesto con potencial para tratar la diabetes mellitus. La herramienta DIA-DB es una herramienta que permite predecir el potencial antidiabético de una molécula basándose en un cálculo de *Docking* y similaridad.

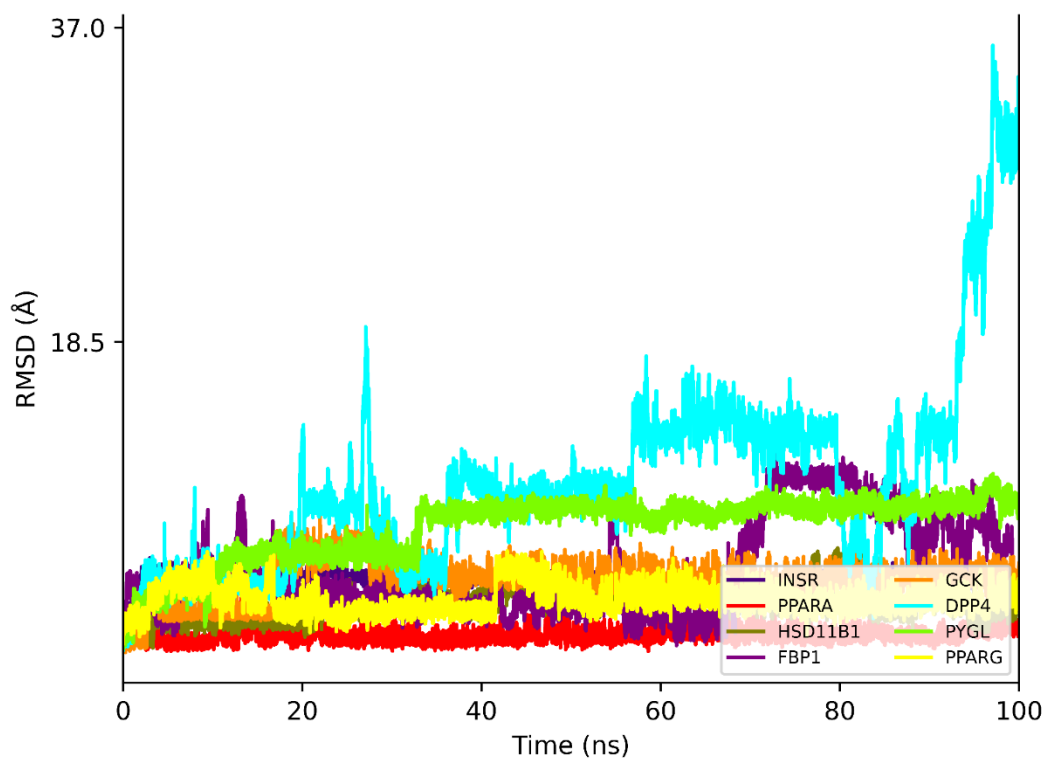
La Tabla V.8. muestra los resultados de *Docking* con las proteínas con mayor potencial para ser diana del carvacrol. Dentro de todas las proteínas obtenidas en el cálculo de *Docking*, se seleccionaron ocho para hacer una simulación de dinámica molecular; estas proteínas fueron DPP4, PYGL, HSD11B1, PPARA, PPARG, INSR, GCK y FBP1. Las simulaciones se hicieron de 100 ns y mostraron como el carvacrol interaccionaba fuertemente con PPARG, PPARA e INSR, mientras que con DPP4, FBP1 y PYGL no se producía ninguna interacción estable. En la Figura V.18. se puede observar como todas las proteínas adquieren una conformación estable exceptuando DPP4.

Tabla V.7. Resultados obtenidos por la herramienta DIA-DB del carvacrol frente a distintas proteínas implicadas en diabetes.

Proteína	Docking Score (kcal/mol)	Proteína	Docking Score (kcal/mol)
FFAR1	-7,9	PDK2	-6,2
ARK1B1	-7,2	HSD11B1	-5,9
RBP4	-7,2	PPARA	-5,8
DPP4	-6,6	NR5A2	-5,6
GCK	-6,6	INSR	-5,4
PPARD	-6,6	PYGL	-5,3
PPARG	-6,4	PTPN9	-5
AMY2A	-6,3	FBP1	-3,5
MGAM	-6,2		

El grupo colaborador liderado por el Dr. Kilic realizó ensayos de toxicidad y midió los niveles de proteína presentes en las células 3T3-L1. Los resultados mostraron que el carvacrol aumenta de forma significativa los niveles de DPP4, PYGL, HSD11B1, PPARA, PPARG, INSR, GCK y FBP1. Además, esta sobreexpresión aumentaba al combinar el uso de carvacrol con la metformina.

Figura V.18. Diagrama de RMSD de cada proteína de interés durante la simulación con el carvacrol. La leyenda muestra el color asignado a cada complejo: morado (unido a INSR), rojo (unido a PPARA), verde oscuro (unido a HSD11B1), violeta (unido a FBP1), naranja (unido a GCK), cian (unido a DPP4), verde claro (unido a PYGL) y amarillo (unido a PPARG).



5.7. RESISTENCIA A LOS ANTIBIÓTICOS

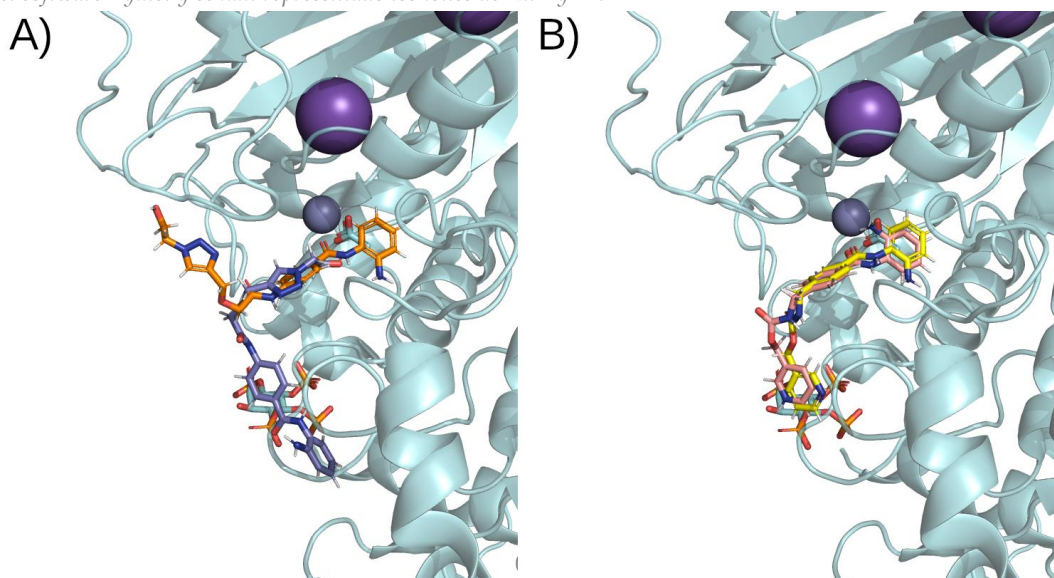
En este apartado se abordará el objetivo 2.6, “Aplicación de métodos de computación de alto rendimiento en resistencia a antibióticos”. En la actualidad, la OMS considera la resistencia a los antibióticos como un peligro público grave, debido al incremento del número de microorganismos que han adquirido resistencia a los antibióticos actuales (205). Por todo ello, es necesario seguir descubriendo nuevos compuestos con actividad antibacteriana (206). Una de las primeras barreras contra los microbios es la piel, las células epiteliales producen péptidos antimicrobianos (AMP), en los seres humanos se distinguen dos tipos principales de AMPs: las defensinas y las catelicidinas. Dentro de las catelicidinas se encuentra el péptido antimicrobiano LL37 (207), este péptido tiene una gran relevancia debido a su capacidad para destruir las bacteria y modular el sistema

inmunitario (208). También existen estudios que muestran como existen compuestos; entre los que se encuentran el butirato (209,210) y el fenilbutirato (211,212), que incrementan la producción de este péptido antimicrobiano. Sin embargo, estos compuestos tienen una vida media demasiado corta (213), por lo que es necesario descubrir nuevos compuestos que activen la producción de LL37 de una manera más estable. Uno de los compuestos identificados como inductores del gen de la catelicidina es el Entinostat (214), el cual es una fenilendiamina que inhibe la histona deacetilasa 3, aumentando de esta manera la producción de LL37.

En el presente trabajo, se ha colaborado con el grupo del Dr Gudmundur Hranf Gudmundsson (<https://english.hi.is/staff/ghrafn>) de la Universidad de Islandia. Este trabajo se ha centrado en una fenilendiamina aromatizada; concretamente la fenilendiamina HO53 (215). Esta fenilendiamina ha mostrado una inhibición de la acetilación producida por la histona deacetilasa 3, lo que provoca un aumento de la producción del péptido antimicrobiano LL37. En este trabajo, se ha conseguido explicar como la fenilendiamina HO53 interacciona con la histona deacetilasa 3.

Para comprobar cómo interaccionaba esta fenilendiamina, se realizó un estudio de BD consenso utilizando los software AutoDock Vina y Lead Finder. La proteína fue extraída de la base de datos PDB con el código 4A69 y procesada para ser utilizada en el cálculo de *Docking*. Los cálculos se realizaron usando la herramienta MetaScreener y se observó que tanto el ligando de estudio (la fenilendiamina HO53) como un compuesto ya conocido (Entinostat) interaccionaban en el sitio activo de la enzima, la zona de unión a Zn^{2+} . Los resultados de BD mostraron como ambos compuestos interaccionan con los residuos clave como la histidina 22, el ácido aspártico 93, las fenilalaninas 144 y 200 y la tirosina 298; la pose predicha por ambos software se muestra en la Figura V.19. Además, de las interacciones los valores obtenidos de *Docking score* fueron similares en ambos compuestos: el fármaco Entinostat obtuvo un valor de *score* promedio de de -9,35 kcal/mol y la fenilendiamina HO53 un valor de -8,22 kcal/mol. Estos resultados junto con los obtenidos experimentalmente por el grupo del Dr Gudmundsson, sugieren que la fenilendiamina HO53 activa la producción de los péptidos antimicrobianos debido a la inhibición de la histona deacetilasa 3.

Figura V.19. Docking de los ligandos de interés frente a la histona deacetilasa 3. A) Unión de la fenilendiamina HO53 con la histona deacetilasa 3 (cian) predicho por el Docking mediante el software AutoDock Vina (azul) y Lead Finder (naranja). B) Unión del Entinostat con la histona deacetilasa 3 (cian) predicho por el Docking mediante el software AutoDock Vina (rosa) y Lead Finder (amarillo). La visualización se ha obtenido mediante el software Pymol y se han representado los iones de Zn^{2+} y K^+ .



Los resultados del grupo liderado por el Dr. Gudmundsson mostraron que la fenilendiamina HO53 producía un cambio significativo en los niveles de expresión de los genes SLC2A4, PDK4, GCKR, PDK2, ALDOB, PFKFB3, PGAM2 y OGDHL. Este cambio era significativo en todos los genes transcurridas 24 horas después de la administración del compuesto. Además, se observó cómo había un aumento en la expresión del péptido antimicrobiano CAMP tras la administración de la fenilendiamina.

5.8. OTROS

Además de los objetivos principales, también se han realizado colaboraciones para conseguir los objetivos secundarios.

5.8.1. Valoración del uso de herramientas de ML en el ámbito de la trombosis

Para valorar el uso del *Machine Learning* en el ámbito de la trombosis se utilizaron las herramientas mencionadas en el punto IV.3 "Predicción estructural de proteínas" y IV.6.2 "Dinámica molecular". Se quiso valorar la utilidad de estas herramientas (AlphaFold, Rosetta y Dinámica Molecular con Maestro-Desmond) para predecir efectos de mutaciones en el ámbito de la deficiencia de antitrombina. En este caso el grupo realizó una colaboración con el doctor Javier Corral de la Calle perteneciente al instituto murciano de investigación biosanitaria (IMIB, <https://www.imib.es/web/personal.jsf?id=3850>). Este trabajo consistió en el uso tanto de AlphaFold 2 como de simulaciones de Dinámica Molecular para predecir efectos de mutaciones en la antitrombina (código en PDB: 1AZX). El grupo colaborador obtuvo las predicciones de las mutaciones de AlphaFold 2 con la versión reducida y obtuvo el cristal de la mutación con la mayor inserción (código en PDB: 4EB1). Mientras que nuestro grupo realizó tanto las predicciones de AlphaFold 2 realizadas con la base de datos completa, como las simulaciones de Dinámica Molecular.

El grupo colaborador liderado por el Dr. Corral de la Calle es experto en la antitrombina siendo su grupo el que cristalizó la mutación de mayor inserción (M5). Debido a relevancia de cada mutación se seleccionaron las siguientes mutaciones:

- I. p.Arg79Cys (M1), produce una deficiencia de antitrombina tipo II HBS (*Heparin Binding Site*); se seleccionó debido que es un control positivo dado que, aunque causa la patología, dicha patología se produce por la imposibilidad de unirse a la heparina sin que haya un cambio estructural en la antitrombina.

- II. p.Pro112Ser (M2), produce una reducción de los niveles de antitrombina en plasma, produciendo además dímeros anómalos debido a la formación de un puente disulfuro, dichos dímeros son incapaces de salir al medio extracelular produciendo una disminución de la proteína en el plasma sanguíneo.
- III. p.Met283Val (M3), produce deficiencia antitrombina tipo II PE (*Pleiotropic Effect*), produciendo un cambio conformacional en la estructura de la antitrombina, disminuyendo tanto su actividad catalítica como su afinidad por la heparina .
- IV. p.Pro352insValPheLeuPro (M4), se produce por un *splicing* alternativo del intrón 5, al igual que ocurre en la mutación II (p.Pro112Ser), se produce una dimerización de los monómero de antitrombina por puentes disulfuro.
- V. p.Glu241_Leu242delinsValLeuValLeuValAsnThrArgThrSer (M5), esta mutación produce una proteína hiperestable de forma constitutiva, por lo que tiene baja afinidad por la heparina.

Los resultados obtenidos en las predicciones de AlphaFold 2 mostraron cómo la herramienta era incapaz de predecir correctamente ninguna de las mutaciones analizadas: ni las sustituciones de aminoácidos ni las de inserción/delección. Además, no había diferencias entre los cálculos realizados con la base de datos completa, respecto a los obtenidos con la base de datos reducida (realizados por el grupo colaborador). También se calcularon los ángulos diedros (obteniendo así su estructura secundaria) de la mutación con un cambio estructural mayor (Cristal:4EB1 y predicción de AF) y de la *Wild Type*, tal como se refleja en la Tabla V.9. Estos resultados mostraron cómo la estructura secundaria de la zona donde se localiza la mutación era predicha de forma incorrecta por AlphaFold 2. La Tabla

V.10. muestra como las predicciones forzaban las mutaciones a su forma WT, la predicción de M5 tenía mayor similitud estructural con la predicción WT que con el cristal de la PDB. Por último, se realizaron las predicciones de las mutaciones puntuales (M1-M3) usando Rosetta y, tampoco se observaron cambios en las estructuras obtenidas, tal y como se puede ver en la Tabla V.11. Además del RMSD, se analizó el IDDT (*local Distance Difference Test*), esta medida calcula la distancia a la cual se encuentran los pares de átomos de diferentes residuos y compara los valores obtenidos a nivel local entre la pose de referencia y la pose de interés (216). De esta manera, se puede establecer cómo de conservadas están ambas estructuras, la métrica IDDT tiene valores que van entre el 0 y el 1, indicando el valor 1 que las dos poses de las proteínas son exactamente iguales.

Tabla V.8. Tabla comparativa de la estructura secundaria de las estructuras de antitrombina. Se muestra la tabla desglosada por residuo, de las estructuras cristalizadas disponibles en la PDB (WT y M5) y de la predicción de AF2 de M5.

Residuo	Cristal WT: 1AZX	Cristal M5: 4EB1	Predicción AF: M5
SER204	Lámina Beta	Lámina Beta	Lámina Beta
GLU205	Giro	Lámina Beta	Giro
ALA206	Alfa Hélice	Lámina Beta	Alfa Hélice
ILE207	Lámina Beta	Lámina Beta	Alfa Hélice
ASN208	Giro	Lámina Beta	Alfa Hélice
THR211	Lámina Beta	Lámina Beta	Lámina Beta
VAL212	Alfa Hélice	Alfa Hélice	Lámina Beta
LEU213	Lámina Beta	Lámina Beta	Lámina Beta
VAL214	Lámina Beta	Lámina Beta	Lámina Beta

Tabla V.9. Métrica RMSD de las estructuras predichas por AF en comparación con los cristales obtenidos de la PDB.

Estructura Base	Estructura a Comparar	RMSD total (Å)
Predicción AF: WT	Cristal WT: 1AZX	2,431
	Predicción AF: M1	2,628
	Predicción AF: M2	2,999
	Predicción AF: M3	2,812
	Predicción AF: M4	7,355
	Predicción AF: M5	1,675
Cristal M5: 4EB1	Predicción AF: M5	4,003
	Predicción AF: WT	3,890

Tabla V.10. Tabla con las diferentes métricas de las mutaciones M1-M3 predichas por AF2. Se muestran las métricas de energía (REU: Rosseta Energy Units), RMSD y IDDT (local Distance Difference Test) de las estructuras predichas por AF2 en comparación con los cristales obtenidos de la PDB.

Predicción AF	Energía (REU)	RMSD (Å)	IDDT
M1	2,27	0,987	0,837
M2	5,88	0,974	0,850
M3	1,9	1,133	0,827

Respecto a los cálculos realizados mediante simulaciones de Dinámica Molecular: los resultados obtenidos mostraron cómo esta técnica era incapaz de cambiar la conformación de la proteína a un tiempo de 1000 ns, ni en condiciones de calor extremo (100 ° C). La Figura V.20. muestra las variaciones de RMSD de cada variante, como se puede observar en dicha figura no se encuentran variaciones en los valores de RMSD. Además, se realizó un análisis de *Free Energy Landscape* (FEL) para comprobar si había cambios en la dinámica en la conformación de menor energía (Tabla V.12. y Tabla V.13.), estos resultados junto con el análisis de RMSD mostraron que no existían diferencias entre la proteína mutante y las predicciones de AlphaFold 2 al realizar las dinámicas.

Tabla V.11. Métrica RMSD obtenidas tras el análisis FEL. Se muestran las métricas de cada una de las poses de mínimo local para las seis estructuras (WT, M1-M5) obtenidos tras las dinámicas de 100 ns en comparación con la pose inicial (a 0 ns).

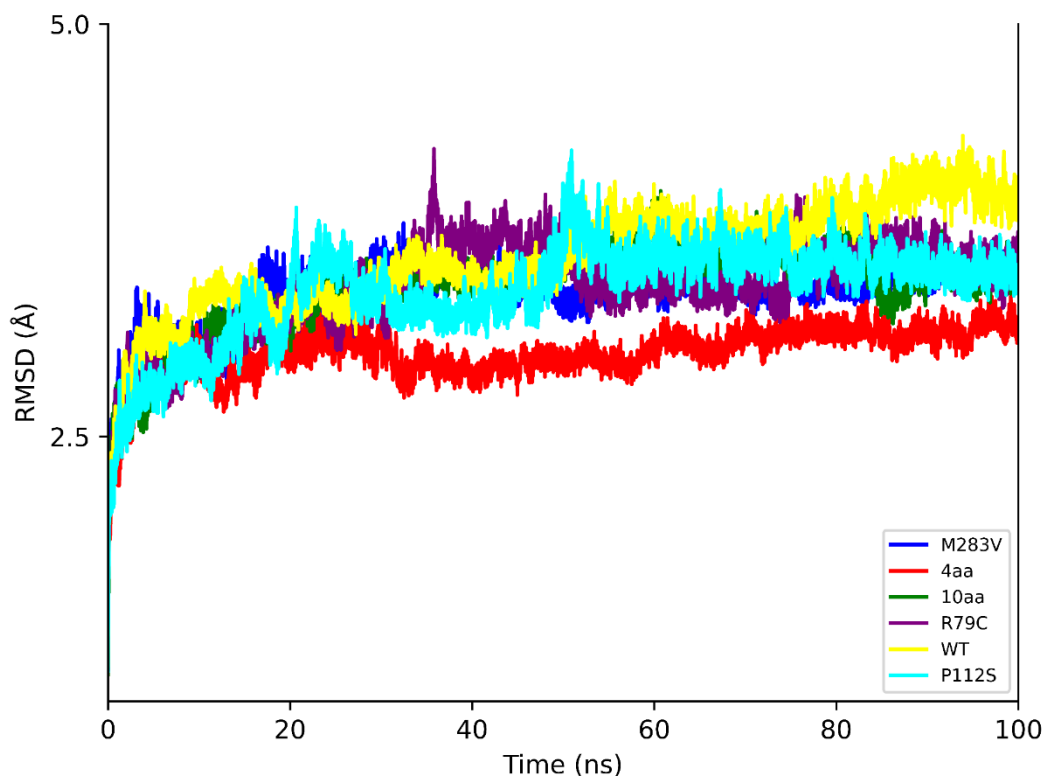
Estructura	N.º de mínimos locales	RMSD (Å) mínimo local 1	RMSD (Å) mínimo local 2	RMSD (Å) mínimo local 3
WT	2	1,877	1,646	-
M1	3	1,596	1,615	1,541
M2	3	1,820	1,832	1,718
M3	2	1,754	1,735	-
M4	2	1,703	1,614	-
M5	1	1,732	-	-

Tabla V.12. Métrica IDDT obtenidas tras el análisis FEL. Se muestran las métricas IDDT de cada una de las poses de mínimo local para las seis estructuras (WT, M1-M5) obtenidos tras las dinámicas de 100 ns en comparación con la pose inicial (a 0 ns).

Estructura	N.º de mínimos locales	IDDT mínimo local 1	IDDT mínimo local 2	IDDT mínimo local 3
WT	2	0,723	0,736	-
M1	3	0,724	0,731	0,727
M2	3	0,723	0,721	0,727
M3	2	0,717	0,734	-
M4	2	0,723	0,746	-
M5	1	0,717	-	-

Con todos estos resultados, se puede concluir que AlphaFold 2 fuerza las predicciones a su conformación nativa en todas las mutaciones estudiadas. Además, ni AlphaFold 2 ni las simulaciones de Dinámica Molecular fueron capaces de predecir el cambio conformacional realizado. Por lo que es importante seguir investigando nuevas técnicas que permitan analizar correctamente las mutaciones ocurridas en la antitrombina.

Figura V.20. Diagrama de RMSD de la antitrombina. La leyenda muestra el color asignado a cada variante: azul (M283V), rojo (Inserción de 4 aminoácidos), verde (Inserción de 10 aminoácidos), morado (R79C), amarillo (WT), y cian (P112S).



5.8.2. Desarrollo de MetaScreener

También durante esta tesis, también se ha realizado una colaboración en el desarrollo de la herramienta MetaScreener, en la cual, se ha realizado una mejora dado que nuestro grupo es el desarrollador de dicha herramienta y, parte de la tesis actual ha consistido en testar y mejorar las herramientas desarrolladas por otros miembros del grupo de investigación. Concretamente se ha trabajado en la mejora del cálculo del *Docking* (tanto en *Virtual Screening* como en *Blind Docking*) usando el software de Lead Finder para mantener las cargas originales proporcionadas en el archivo de input de los ligandos. Respecto al software AutoDock Vina, se ha mejorado el cálculo de las predicciones proteína-ligando mediante el software PLIP (<https://plip-tool.biotec.tu-dresden.de/plip-web/plip/index>) gracias a la implementación del uso de Pymol para evitar problemas con las moléculas en formato pdbqt. En el script consenso de *Blind Docking* y *Virtual Screening* se ha

trabajado en la mejora y optimización de los dos scripts originales (el primer script que une los trabajos y el segundo que calcula las interacciones y el *Docking score* final).

VI – CONCLUSIONES

VI -CONCLUSIONES

- A protocol (TOLEDO) has been created to massively and efficiently launch multiple Molecular Dynamics simulations in high-performance computing environments in a simple and fast way.
- The software TOLEDO can be used in a wide variety of contexts where the relevance of the disease is less important than the thorough characterization of the protein of interest.
 - Several *in silico* techniques have been exploited to find novel compounds against multiple pathologies.
- The MetaScreener tool has been used in many contexts:
 - It facilitated the discovery of several inhibitors for the treatment of various types of cancer.
 - It has been used to research Alzheimer's disease and its possible therapies.
 - The improvement of antibiotic resistance is becoming more prevalent.
- The MetaScreener tool has been combined with Molecular Dynamics and used in multiple fields:
 - Understanding the molecular mechanism of quercetin to treat retinitis pigmentosa.
 - Using phytoprostanes to treat diseases related to wound healing.

- Collaborations have been carried out in numerous projects including the development of new software and the application of existing software in different fields.
 - Use of AlphaFold and Molecular Dynamics simulation in antithrombin deficiency.

VII - LIMITACIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN

VII - LIMITACIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN

- Actualmente TOLEDO solamente se puede utilizar para analizar automáticamente sistemas proteína-ligando.
- TOLEDO sigue necesitando el uso de la interfaz gráfica de usuario para realizar cálculos de proteínas con membrana.
- Es necesario encontrar nuevos datasets con información actualizada y que tengan datos suficientes para poder crear modelos eficaces de ML.
- Las técnicas *in silico* siguen necesitando de validación mediante técnicas *in vitro*/ *in vivo* y estas técnicas requieren mucho más tiempo para realizarse.

Futuras líneas de investigación

- Futuras versiones de TOLEDO van a implementar análisis de sistemas proteína-proteína y mejorar los análisis actuales proteína-ligando.
- La herramienta SIBILA va a seguir siendo testada con más modelos de ML combinados con Dinámica Molecular.
- Seguir implementando las técnicas *in silico* para descubrir nuevos tratamientos para enfermedades.

VIII - REFERENCIAS BIBLIOGRÁFICAS

VIII - REFERENCIAS BIBLIOGRÁFICAS

1. Megari K. Quality of life in chronic disease patients. *Health Psychol Res.* 2013 Sep 23;1(3):27.
2. Jung M. Challenges of Multimorbidities in the Era of an Aging Population. *Health Care Manag.* 2016 Apr;35(2):134–43.
3. Schluger NW, Koppaka R. Lung Disease in a Global Context. A Call for Public Health Action. *Ann Am Thorac Soc.* 2014 Mar;11(3):407–16.
4. Capingana DP, Magalhães P, Silva AB, Gonçalves MA, Baldo MP, Rodrigues SL, et al. Prevalence of cardiovascular risk factors and socioeconomic level among public-sector workers in Angola. *BMC Public Health.* 2013 Dec;13(1):732.
5. Zhang J, Kern-Allely S, Yu T, Price RK. HIV and tuberculosis co-infection in East Asia and the Pacific from 1990 to 2017: results from the Global Burden of Disease Study 2017. *J Thorac Dis.* 2019 Sep;11(9):3822–35.
6. Hussey L, Turner S, Thorley K, McNamee R, Agius R. Work-related sickness absence as reported by UK general practitioners. *Occup Med.* 2012 Mar 1;62(2):105–11.
7. Di Gennaro F, Pizzol D, Marotta C, Antunes M, Racalbutto V, Veronese N, et al. Coronavirus Diseases (COVID-19) Current Status and Future Perspectives: A Narrative Review. *Int J Environ Res Public Health.* 2020 Apr 14;17(8):2690.
8. Chowell G, Abdirizak F, Lee S, Lee J, Jung E, Nishiura H, et al. Transmission characteristics of MERS and SARS in the healthcare setting: a comparative study. *BMC Med.* 2015 Dec;13(1):210.
9. Zeng Z, Zhan J, Chen L, Chen H, Cheng S. Global, regional, and national dengue burden from 1990 to 2017: A systematic analysis based on the global burden of disease study 2017. *EClinicalMedicine.* 2021 Feb;32:100712.

10. WHO Ebola Response Team. Ebola Virus Disease in West Africa — The First 9 Months of the Epidemic and Forward Projections. *N Engl J Med*. 2014 Oct 16;371(16):1481–95.
11. Nichols E, Szeoke CEI, Vollset SE, Abbasi N, Abd-Allah F, Abdela J, et al. Global, regional, and national burden of Alzheimer’s disease and other dementias, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol*. 2019 Jan;18(1):88–106.
12. Thanarajasingam G, Hubbard JM, Sloan JA, Grothey A. The Imperative for a New Approach to Toxicity Analysis in Oncology Clinical Trials. *J Natl Cancer Inst*. 2015 Oct;107(10):djv216.
13. Blüher M. Obesity: global epidemiology and pathogenesis. *Nat Rev Endocrinol*. 2019 May;15(5):288–98.
14. Tashiro H, Brenner MK. Immunotherapy against cancer-related viruses. *Cell Res*. 2017 Jan;27(1):59–73.
15. Fernández-Menéndez S, Fernández-Morán M, Fernández-Vega I, Pérez-Álvarez A, Villafani-Echazú J. Epstein–Barr virus and multiple sclerosis. From evidence to therapeutic strategies. *J Neurol Sci*. 2016 Feb;361:213–9.
16. Campisi J, Yaswen P. Aging and cancer cell biology, 2009. *Aging Cell*. 2009 Jun;8(3):221–5.
17. Hou Y, Dan X, Babbar M, Wei Y, Hasselbalch SG, Croteau DL, et al. Ageing as a risk factor for neurodegenerative disease. *Nat Rev Neurol*. 2019 Oct;15(10):565–81.
18. Bertoglia MP, Gormaz JG, Libuy M, Sanhueza D, Gajardo A, Srur A, et al. The population impact of obesity, sedentary lifestyle, and tobacco and alcohol consumption on the prevalence of type 2 diabetes: Analysis of a health population survey in Chile, 2010. Guerrero-Romero F, editor. *PLOS ONE*. 2017 May 25;12(5):e0178092.
19. Olvera Alvarez HA, Kubzansky LD, Campen MJ, Slavich GM. Early life stress, air pollution, inflammation, and disease: An integrative review and immunologic model of social-environmental adversity and lifespan health. *Neurosci Biobehav Rev*. 2018 Sep;92:226–42.

20. Thomson EM. Air Pollution, Stress, and Allostatic Load: Linking Systemic and Central Nervous System Impacts. *J Alzheimers Dis.* 2019 Jun 4;69(3):597–614.
21. Anichini G, Gandolfo C, Fabrizi S, Miceli GB, Terrosi C, Gori Savellini G, et al. Seroprevalence to Measles Virus after Vaccination or Natural Infection in an Adult Population, in Italy. *Vaccines.* 2020 Feb 3;8(1):66.
22. Mera RM, Suaya JA, Amrine-Madsen H, Hoge CS, Miller LA, Lu EP, et al. Increasing Role of *Staphylococcus aureus* and Community-Acquired Methicillin-Resistant *Staphylococcus aureus* Infections in the United States: A 10-Year Trend of Replacement and Expansion. *Microb Drug Resist.* 2011 Jun;17(2):321–8.
23. Wei R, Ge F, Huang S, Chen M, Wang R. Occurrence of veterinary antibiotics in animal wastewater and surface water around farms in Jiangsu Province, China. *Chemosphere.* 2011 Mar;82(10):1408–14.
24. Van Boeckel TP, Brower C, Gilbert M, Grenfell BT, Levin SA, Robinson TP, et al. Global trends in antimicrobial use in food animals. *Proc Natl Acad Sci.* 2015 May 5;112(18):5649–54.
25. Caskurlu H, Culpan M, Erol B, Turan T, Vahaboglu H, Caskurlu T. Changes in Antimicrobial Resistance of Urinary Tract Infections in Adult Patients over a 5-Year Period. *Urol Int.* 2020;104(3–4):287–92.
26. Kolch W. Coordinating ERK/MAPK signalling through scaffolds and inhibitors. *Nat Rev Mol Cell Biol.* 2005 Nov;6(11):827–37.
27. Xin P, Xu X, Deng C, Liu S, Wang Y, Zhou X, et al. The role of JAK/STAT signaling pathway and its inhibitors in diseases. *Int Immunopharmacol.* 2020 Mar;80:106210.
28. Traut T. Allosteric regulatory enzymes. New York: Springer; 2008. 250 p.
29. Gu J, Gong E, Zhang B, Zheng J, Gao Z, Zhong Y, et al. Multiple organ infection and the pathogenesis of SARS. *J Exp Med.* 2005 Aug 1;202(3):415–24.
30. Kulenkampff K, Wolf Perez AM, Sormanni P, Habchi J, Vendruscolo M. Quantifying misfolded protein oligomers as drug targets and biomarkers in Alzheimer and Parkinson diseases. *Nat Rev Chem.* 2021 Feb 15;5(4):277–94.

31. Sturdee AP, Bodley-Tickell AT, Archer A, Chalmers RM. Long-term study of *Cryptosporidium* prevalence on a lowland farm in the United Kingdom. *Vet Parasitol.* 2003 Oct;116(2):97–113.
32. Larsen J, Petersen A, Sørnum M, Stegger M, Van Alphen L, Valentiner-Branth P, et al. Meticillin-resistant *Staphylococcus aureus* CC398 is an increasing cause of disease in people with no livestock contact in Denmark, 1999 to 2011. *Eurosurveillance* [Internet]. 2015 Sep 17 [cited 2024 Sep 3];20(37). Available from: <https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2015.20.37.30021>
33. Vialle-Valentin CE, Serumaga B, Wagner AK, Ross-Degnan D. Evidence on access to medicines for chronic diseases from household surveys in five low- and middle-income countries. *Health Policy Plan.* 2015 Oct;30(8):1044–52.
34. Makoge V, Vaandrager L, Maat H, Koelen M. Poverty and health among CDC plantation labourers in Cameroon: Perceptions, challenges and coping strategies. Barry AE, editor. *PLoS Negl Trop Dis.* 2017 Nov 20;11(11):e0006100.
35. Challa JM, Getachew T, Debella A, Merid M, Atnafe G, Eyeberu A, et al. Inadequate Hand Washing, Lack of Clean Drinking Water and Latrines as Major Determinants of Cholera Outbreak in Somali Region, Ethiopia in 2019. *Front Public Health.* 2022 May 6;10:845057.
36. Rassi A, Rassi A, Marin-Neto JA. Chagas disease. *The Lancet.* 2010 Apr;375(9723):1388–402.
37. Shiffman J. Donor funding priorities for communicable disease control in the developing world. *Health Policy Plan.* 2006 Aug 30;21(6):411–20.
38. Alemayehu C, Mitchell G, Nikles J. Barriers for conducting clinical trials in developing countries- a systematic review. *Int J Equity Health.* 2018 Dec;17(1):37.
39. Sardana D, Zhu C, Zhang M, Gudivada RC, Yang L, Jegga AG. Drug repositioning for orphan diseases. *Brief Bioinform.* 2011 Jul 1;12(4):346–56.
40. Barrett PM, Alagely A, Topol EJ. Cystic fibrosis in an era of genomically guided therapy. *Hum Mol Genet.* 2012 Oct 15;21(R1):R66–71.

41. McKone EF, Borowitz D, Drevinek P, Griese M, Konstan MW, Wainwright C, et al. Long-term safety and efficacy of ivacaftor in patients with cystic fibrosis who have the Gly551Asp-CFTR mutation: a phase 3, open-label extension study (PERSIST). *Lancet Respir Med*. 2014 Nov;2(11):902–10.
42. Zhou Q, Jin J fen, Zhu L ling, Chen M, Xu H min, Wang H fen, et al. The optimal choice of medication administration route regarding intravenous, intramuscular, and subcutaneous injection. *Patient Prefer Adherence*. 2015 Jul;9:23.
43. Dagan R, Hoberman A, Johnson C, Leibovitz EL, Arguedas A, Rose FV, et al. Bacteriologic and clinical efficacy of high dose amoxicillin/clavulanate in children with acute otitis media: *Pediatr Infect Dis J*. 2001 Sep;20(9):829–37.
44. Tarver CP, Noorily AD, Sakai CS. First place – resident clinical science award 1993: A comparison of cocaine vs. lidocaine with oxymetazoline for use in nasal procedures. *Otolaryngol Neck Surg*. 1993 Oct;109(4):653–9.
45. Gordh T, Gordh TE, Lindqvist K, Warner DS. Lidocaine: The Origin of a Modern Local Anesthetic. *Anesthesiology*. 2010 Dec 1;113(6):1433–7.
46. Krohe M, Eek D, Mazar I, Horsfield A, Pompilus F, Friebe R, et al. Patient-reported preferences for oral versus intravenous administration for the treatment of cancer: a review of the literature. *Patient Prefer Adherence*. 2016 Aug;Volume 10:1609–21.
47. Sertkaya A, Wong HH, Jessup A, Beleche T. Key cost drivers of pharmaceutical clinical trials in the United States. *Clin Trials*. 2016 Apr;13(2):117–26.
48. DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: New estimates of R&D costs. *J Health Econ*. 2016 May;47:20–33.
49. Chorniy A, Bailey J, Civan A, Maloney M. Regulatory review time and pharmaceutical research and development. *Health Econ*. 2021 Jan;30(1):113–28.
50. Braunholtz DA, Edwards SJL, Lilford RJ. Are randomized clinical trials good for us (in the short term)? Evidence for a “trial effect”. *J Clin Epidemiol*. 2001 Mar;54(3):217–24.

51. Locock L, Smith L. Personal benefit, or benefiting others? Deciding whether to take part in clinical trials. *Clin Trials*. 2011 Feb;8(1):85–93.
52. Alomar M, Tawfiq AM, Hassan N, Palaian S. Post marketing surveillance of suspected adverse drug reactions through spontaneous reporting: current status, challenges and the future. *Ther Adv Drug Saf*. 2020 Jan;11:204209862093859.
53. Van Norman GA. Limitations of Animal Studies for Predicting Toxicity in Clinical Trials. *JACC Basic Transl Sci*. 2019 Nov;4(7):845–54.
54. Doke SK, Dhawale SC. Alternatives to animal testing: A review. *Saudi Pharm J*. 2015 Jul;23(3):223–9.
55. Banerjee P, Siramshetty VB, Drwal MN, Preissner R. Computational methods for prediction of in vitro effects of new chemical structures. *J Cheminformatics*. 2016 Dec;8(1):51.
56. Goodman J, Chandna A, Roe K. Trends in animal use at US research facilities. *J Med Ethics*. 2015 Jul;41(7):567–9.
57. Nowocin AK, Brown K, Edwards LA, Meader L, Hill JI, Wong W. An Extraperitoneal Technique for Murine Heterotopic Cardiac Transplantation. *Am J Transplant*. 2015 Sep;15(9):2491–4.
58. Ohta H, Komatsu T, Muta K, Koizumi M, Iguchi Y, Okano HJ. Marmoset angiography just by percutaneous puncture of the caudal ventral artery. *Ai J*, editor. *PLOS ONE*. 2021 Apr 28;16(4):e0250576.
59. Bhutani P, Joshi G, Raja N, Bachhav N, Rajanna PK, Bhutani H, et al. U.S. FDA Approved Drugs from 2015–June 2020: A Perspective. *J Med Chem*. 2021 Mar 11;64(5):2339–81.
60. Falcone R, Lombardi P, Filetti M, Duranti S, Pietragalla A, Fabi A, et al. Oncologic Drugs Approval in Europe for Solid Tumors: Overview of the Last 6 Years. *Cancers*. 2022 Feb 11;14(4):889.
61. Chiu PKF, Shen X, Wang G, Ho CL, Leung CH, Ng CF, et al. Enhancement of prostate cancer diagnosis by machine learning techniques: an algorithm development and validation study. *Prostate Cancer Prostatic Dis*. 2022 Dec;25(4):672–6.

62. Chen S, Jian T, Chi C, Liang Y, Liang X, Yu Y, et al. Machine Learning-Based Models Enhance the Prediction of Prostate Cancer. *Front Oncol.* 2022 Jul 6;12:941349.
63. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJWL. Machine Learning methods for Quantitative Radiomic Biomarkers. *Sci Rep.* 2015 Aug 17;5(1):13087.
64. Gandelman JS, Byrne MT, Mistry AM, Polikowsky HG, Diggins KE, Chen H, et al. Machine learning reveals chronic graft- *versus* -host disease phenotypes and stratifies survival after stem cell transplant for hematologic malignancies. *Haematologica.* 2019 Jan;104(1):189–96.
65. Manzari MT, Shamay Y, Kiguchi H, Rosen N, Scaltriti M, Heller DA. Targeted drug delivery strategies for precision medicines. *Nat Rev Mater.* 2021 Feb 2;6(4):351–70.
66. Sicklick JK, Kato S, Okamura R, Schwaederle M, Hahn ME, Williams CB, et al. Molecular profiling of cancer patients enables personalized combination therapy: the I-PREDICT study. *Nat Med.* 2019 May;25(5):744–50.
67. Gunasekaran K, Nussinov R. How Different are Structurally Flexible and Rigid Binding Sites? Sequence and Structural Features Discriminating Proteins that Do and Do not Undergo Conformational Change upon Ligand Binding. *J Mol Biol.* 2007 Jan;365(1):257–73.
68. Koshland DE. The Key–Lock Theory and the Induced Fit Theory. *Angew Chem Int Ed Engl.* 1995 Jan 3;33(23–24):2375–8.
69. Whitesides GM, Krishnamurthy VM. Designing ligands to bind proteins. *Q Rev Biophys.* 2005 Nov;38(4):385–95.
70. Besnard J, Ruda GF, Setola V, Abecassis K, Rodriguiz RM, Huang XP, et al. Automated design of ligands to polypharmacological profiles. *Nature.* 2012 Dec;492(7428):215–20.
71. Schaefer HF. A history of ab initio computational quantum chemistry: 1950–1960. *Tetrahedron Comput Methodol.* 1988;1(2):97–102.
72. Hurley AC. Thermochemistry in the Hartree-Fock Approximation. In: *Advances in Quantum Chemistry* [Internet]. Elsevier; 1973 [cited 2024 Sep

- 4]. p. 315–34. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0065327608605675>
73. Perdew JP, Levy M. Physical Content of the Exact Kohn-Sham Orbital Energies: Band Gaps and Derivative Discontinuities. *Phys Rev Lett*. 1983 Nov 14;51(20):1884–7.
74. Becke AD. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys Rev A*. 1988 Sep 1;38(6):3098–100.
75. Ugi I, Bauer J, Brandt J, Friedrich J, Gasteiger J, Jochum C, et al. New Applications of Computers in Chemistry. *Angew Chem Int Ed Engl*. 1979 Feb;18(2):111–23.
76. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, et al. Gaussian~16 Revision C.01. 2016.
77. James JP. MOPAC2016. Stewart Stewart Comput Chem Colo Springs CO USAHTTPOpenMOPACnet 2016.
78. Barca GMJ, Bertoni C, Carrington L, Datta D, De Silva N, Deustua JE, et al. Recent developments in the general atomic and molecular electronic structure system. *J Chem Phys*. 2020 Apr 21;152(15):154102.
79. Case DA, Belfon K, Ben-Shalom IY, Brozell SR, Cerutti DS, Cheatham TE, et al. AMBER 2020. University of California, San Francisco; 2020.
80. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, et al. The Amber biomolecular simulation programs. *J Comput Chem*. 2005 Dec;26(16):1668–88.
81. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. *J Comput Chem*. 2005 Dec;26(16):1781–802.
82. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers JR, et al. The protein data bank: A computer-based archival file for macromolecular structures. *J Mol Biol*. 1977 May;112(3):535–42.
83. Chermette H. Chemical reactivity indexes in density functional theory. *J Comput Chem*. 1999 Jan 15;20(1):129–54.
84. Karplus M, Petsko GA. Molecular dynamics simulations in biology. *Nature*. 1990 Oct;347(6294):631–9.

85. Tidor B. Molecular dynamics simulations. *Curr Biol*. 1997 Sep;7(9):R525–7.
86. Harvey MJ, De Fabritiis G. High-throughput molecular dynamics: the powerful new tool for drug discovery. *Drug Discov Today*. 2012 Oct;17(19–20):1059–62.
87. Ge H, Wang Y, Li C, Chen N, Xie Y, Xu M, et al. Molecular Dynamics-Based Virtual Screening: Accelerating the Drug Discovery Process by High-Performance Computing. *J Chem Inf Model*. 2013 Oct 28;53(10):2757–64.
88. Sierra-Sosa D, Garcia-Zapirain B, Castillo C, Oleagordia I, Nuno-Solinis R, Urtaran-Laresgoiti M, et al. Scalable Healthcare Assessment for Diabetic Patients Using Deep Learning on Multiple GPUs. *IEEE Trans Ind Inform*. 2019 Oct;15(10):5682–9.
89. Akintoye SB, Han L, Zhang X, Chen H, Zhang D. A Hybrid Parallelization Approach for Distributed and Scalable Deep Learning. *IEEE Access*. 2022;10:77950–61.
90. Batool M, Ahmad B, Choi S. A structure-based drug discovery paradigm. *Int J Mol Sci*. 2019;20(11).
91. Misura KMS, Chivian D, Rohl CA, Kim DE, Baker D. Physically realistic homology models built with ROSETTA can be more accurate than their templates. *Proc Natl Acad Sci*. 2006 Apr 4;103(14):5361–6.
92. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*. 2018 Jul 2;46(W1):W296–303.
93. Webb B, Sali A. Comparative Protein Structure Modeling Using MODELLER. *Curr Protoc Bioinforma* [Internet]. 2016 Jun [cited 2024 Sep 5];54(1). Available from: <https://currentprotocols.onlinelibrary.wiley.com/doi/10.1002/cpbi.3>
94. Bender BJ, Cisneros A, Duran AM, Finn JA, Fu D, Lokits AD, et al. Protocols for Molecular Modeling with Rosetta3 and RosettaScripts. *Biochemistry*. 2016 Aug 30;55(34):4748–63.

95. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021 Aug 26;596(7873):583–9.
96. Torres PHM, Sodero ACR, Jofily P, Silva-Jr FP. Key Topics in Molecular Docking for Drug Design. *Int J Mol Sci*. 2019 Sep 15;20(18):4574.
97. Castro-Alvarez A, Costa A, Vilarrasa J. The Performance of Several Docking Programs at Reproducing Protein–Macrolide-Like Crystal Structures. *Molecules*. 2017 Jan 17;22(1):136.
98. Trott O, Olson AJ. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*. 2010;31(2):455–61.
99. Stroganov OV, Novikov FN, Stroylov VS, Kulkov V, Chilov GG. Lead Finder: An Approach To Improve Accuracy of Protein–Ligand Docking, Binding Energy Estimation, and Virtual Screening. *J Chem Inf Model*. 2008 Dec 22;48(12):2371–85.
100. Imbernón B, Serrano A, Bueno-Crespo A, Abellán JL, Pérez-Sánchez H, Cecilia JM. METADOCK 2: a high-throughput parallel metaheuristic scheme for molecular docking. *Bioinformatics*. 2021 Jul 12;37(11):1515–20.
101. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J Med Chem*. 2004 Mar 1;47(7):1739–49.
102. McNutt AT, Francoeur P, Aggarwal R, Masuda T, Meli R, Ragoza M, et al. GNINA 1.0: molecular docking with deep learning. *J Cheminformatics*. 2021 Jun 9;13(1):43.
103. Tapia-Abellán A, Angosto-Bazarra D, Martínez-Banaclocha H, de Torre-Minguela C, Cerón-Carrasco JP, Pérez-Sánchez H, et al. MCC950 closes the active conformation of NLRP3 to an inactive state. *Nat Chem Biol*. 2019 Jun;15(6):560–4.
104. Zhao H, Caflisch A. Molecular dynamics in drug design. *Eur J Med Chem*. 2015 Feb;91:4–14.
105. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level

- parallelism from laptops to supercomputers. *SoftwareX*. 2015 Sep;1–2:19–25.
106. Bowers KJ, Sacerdoti FD, Salmon JK, Shan Y, Shaw DE, Chow E, et al. Molecular dynamics---Scalable algorithms for molecular dynamics simulations on commodity clusters. In: *Proceedings of the 2006 ACM/IEEE conference on Supercomputing - SC '06* [Internet]. Tampa, Florida: ACM Press; 2006 [cited 2024 Feb 8]. p. 84. Available from: <http://portal.acm.org/citation.cfm?doid=1188455.1188544>
 107. Case DA, Aktulga HM, Belfon K, Cerutti DS, Cisneros GA, Cruzeiro VWD, et al. AmberTools. *J Chem Inf Model*. 2023 Oct 23;63(20):6183–91.
 108. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM : A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem*. 1983 Jun;4(2):187–217.
 109. Lill MA. Multi-dimensional QSAR in drug discovery. *Drug Discov Today*. 2007 Dec;12(23–24):1013–7.
 110. Kensert A, Alvarsson J, Norinder U, Spjuth O. Evaluating parameters for ligand-based modeling with random forest on sparse data sets. *J Cheminformatics*. 2018 Dec;10(1):49.
 111. Wolber G, Langer T. LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters. *J Chem Inf Model*. 2005 Jan 1;45(1):160–9.
 112. Borges NM, Sartori GR, Ribeiro JFR, Rocha JR, Martins JBL, Montanari CA, et al. Similarity search combined with docking and molecular dynamics for novel hAChE inhibitor scaffolds. *J Mol Model*. 2018 Jan;24(1):41.
 113. Ebalunode JO, Zheng W. Unconventional 2D Shape Similarity Method Affords Comparable Enrichment as a 3D Shape Method in Virtual Screening Experiments. *J Chem Inf Model*. 2009 Jun 22;49(6):1313–20.
 114. Helguera A, Combes R, Gonzalez M, Cordeiro MN. Applications of 2D Descriptors in Drug Design: A DRAGON Tale. *Curr Top Med Chem*. 2008 Dec 1;8(18):1628–55.
 115. Jaeger S, Fulle S, Turk S. Mol2vec: Unsupervised Machine Learning Approach with Chemical Intuition. *J Chem Inf Model*. 2018 Jan 22;58(1):27–35.

116. Datta R, Das D, Das S. Efficient lipophilicity prediction of molecules employing deep-learning models. *Chemom Intell Lab Syst.* 2021 Jun;213:104309.
117. Boyles F, Deane CM, Morris GM. Learning from the ligand: using ligand-based features to improve binding affinity prediction. Elofsson A, editor. *Bioinformatics.* 2020 Feb 1;36(3):758–64.
118. Menden MP, Iorio F, Garnett M, McDermott U, Benes CH, Ballester PJ, et al. Machine Learning Prediction of Cancer Cell Sensitivity to Drugs Based on Genomic and Chemical Properties. Raghava GPS, editor. *PLoS ONE.* 2013 Apr 30;8(4):e61318.
119. Tao L, Zhang P, Qin C, Chen SY, Zhang C, Chen Z, et al. Recent progresses in the exploration of machine learning methods as in-silico ADME prediction tools. *Adv Drug Deliv Rev.* 2015 Jun;86:83–100.
120. Pereira J, Simpkin AJ, Hartmann MD, Rigden DJ, Keegan RM, Lupas AN. High-accuracy protein structure prediction in CASP14. *Proteins Struct Funct Bioinforma.* 2021 Dec;89(12):1687–99.
121. Yoo AB, Jette MA, Grondona M. SLURM: Simple Linux Utility for Resource Management. In: Feitelson D, Rudolph L, Schwiegelshohn U, editors. *Job Scheduling Strategies for Parallel Processing [Internet].* Berlin, Heidelberg: Springer Berlin Heidelberg; 2003 [cited 2022 Feb 25]. p. 44–60. (Goos G, Hartmanis J, van Leeuwen J, editors. *Lecture Notes in Computer Science*; vol. 2862). Available from: http://link.springer.com/10.1007/10968987_3
122. Gentsch W. Sun Grid Engine: towards creating a compute power grid. In: *Proceedings First IEEE/ACM International Symposium on Cluster Computing and the Grid [Internet].* Brisbane, Qld., Australia: IEEE Comput. Soc; 2001 [cited 2024 Sep 6]. p. 35–6. Available from: <http://ieeexplore.ieee.org/document/923173/>
123. Banks JL, Beard HS, Cao Y, Cho AE, Damm W, Farid R, et al. Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J Comput Chem.* 2005 Dec;26(16):1752–80.

124. Roos K, Wu C, Damm W, Reboul M, Stevenson JM, Lu C, et al. OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. *J Chem Theory Comput.* 2019 Mar 12;15(3):1863–74.
125. Lu C, Wu C, Ghoreishi D, Chen W, Wang L, Damm W, et al. OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space. *J Chem Theory Comput.* 2021 Jul 13;17(7):4291–300.
126. Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chem Inf Comput Sci.* 1988 Feb 1;28(1):31–6.
127. Gasteiger J, Marsili M. Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron.* 1980 Jan;36(22):3219–28.
128. Halgren TA. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J Comput Chem.* 1996 Apr;17(5–6):490–519.
129. Wang E, Sun H, Wang J, Wang Z, Liu H, Zhang JZH, et al. End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. *Chem Rev.* 2019 Aug 28;119(16):9478–508.
130. Jacobson MP, Friesner RA, Xiang Z, Honig B. On the Role of the Crystal Environment in Determining Protein Side-chain Conformations. *J Mol Biol.* 2002 Jul;320(3):597–608.
131. Jacobson MP, Pincus DL, Rapp CS, Day TJJ, Honig B, Shaw DE, et al. A hierarchical approach to all-atom protein loop prediction. *Proteins Struct Funct Bioinforma.* 2004 May;55(2):351–67.
132. Felling A, Seeber M, Fanelli F. PSNtools for standalone and web-based structure network analyses of conformational ensembles. *Comput Struct Biotechnol J.* 2022;20:640–9.
133. Ribeiro MT, Singh S, Guestrin C. arXiv.org. 2016 [cited 2024 Sep 10]. ‘Why Should I Trust You?’: Explaining the Predictions of Any Classifier. Available from: <https://arxiv.org/abs/1602.04938v3>

134. Lundberg S, Lee SI. arXiv.org. 2017 [cited 2024 Sep 10]. A Unified Approach to Interpreting Model Predictions. Available from: <https://arxiv.org/abs/1705.07874v2>
135. Abel R, Paredes Ramos M, Chen Q, Pérez-Sánchez H, Coluzzi F, Rocco M, et al. Computational Prediction of Potential Inhibitors of the Main Protease of SARS-CoV-2. *Front Chem*. 2020 Dec 23;8:590263.
136. Pérez-Sánchez H, den-Haan H, Peña-García J, Lozano-Sánchez J, Martínez Moreno ME, Sánchez-Pérez A, et al. DIA-DB: A Database and Web Server for the Prediction of Diabetes Drugs. *J Chem Inf Model*. 2020 Sep 28;60(9):4124–30.
137. Sarmah S, Pahari S, Belwal VK, Jana M, Singha Roy A. Elucidation of molecular interaction of bioactive flavonoid luteolin with human serum albumin and its glycated analogue using multi-spectroscopic and computational studies. *J Mol Liq*. 2020 Nov;318:114147.
138. Zhang C, Wang R, Zhang G, Gong D. Mechanistic insights into the inhibition of quercetin on xanthine oxidase. *Int J Biol Macromol*. 2018 Jun 1;112:405–12.
139. Ernst ME, Fravel MA. Febuxostat: A selective xanthine-oxidase/xanthine-dehydrogenase inhibitor for the management of hyperuricemia in adults with gout. *Clin Ther*. 2009 Nov 1;31(11):2503–18.
140. Szliszka E, Skaba D, Czuba ZP, Krol W. Inhibition of Inflammatory Mediators by Neobavaisoflavone in Activated RAW264.7 Macrophages. *Molecules*. 2011 May 3;16(5):3701–12.
141. Lee S, Yun B, Kim M, Park C, Lee W, Oh HM, et al. Phenolic Compounds Isolated from *Psoralea corylifolia* Inhibit IL-6-induced STAT3 Activation. *Planta Med*. 2012 Jun;78(09):903–6.
142. Jiménez A, Selga A, Torres JL, Julià L. Reducing Activity of Polyphenols with Stable Radicals of the TTM Series. Electron Transfer versus H-Abstraction Reactions in Flavan-3-ols. *Org Lett*. 2004 Nov 1;6(24):4583–6.
143. Cao H, Pauff JM, Hille R. X-ray Crystal Structure of a Xanthine Oxidase Complex with the Flavonoid Inhibitor Quercetin. *J Nat Prod*. 2014 Jul 25;77(7):1693–9.

144. Lee SG, Lee H, Nam TG, Eom SH, Heo HJ, Lee CY, et al. Neuroprotective Effect of Caffeoylquinic Acids from *Artemisia princeps* Pampanini against Oxidative Stress-Induced Toxicity in PC-12 Cells. *J Food Sci.* 2011;76(2):C250–6.
145. Mikulic-Petkovsek M, Samoticha J, Eler K, Stampar F, Veberic R. Traditional Elderflower Beverages: A Rich Source of Phenolic Compounds with High Antioxidant Activity. *J Agric Food Chem.* 2015 Feb 11;63(5):1477–87.
146. Abdel Motaal A, Ezzat SM, Tadros MG, El-Askary HI. In vivo anti-inflammatory activity of caffeoylquinic acid derivatives from *Solidago virgaurea* in rats. *Pharm Biol.* 2016 Dec 1;54(12):2864–70.
147. Wan Y, Qian J, Li Y, Shen Y, Chen Y, Fu G, et al. Inhibitory mechanism of xanthine oxidase activity by caffeoylquinic acids *in vitro*. *Int J Biol Macromol.* 2021 Aug 1;184:843–56.
148. Mehmood A, Li J, Rehman AU, Kobun R, Llah IU, Khan I, et al. Xanthine oxidase inhibitory study of eight structurally diverse phenolic compounds. *Front Nutr [Internet].* 2022 Sep 20 [cited 2024 Sep 10];9. Available from: <https://www.frontiersin.org/journals/nutrition/articles/10.3389/fnut.2022.966557/full>
149. Zhai L, Ladomersky E, Lenzen A, Nguyen B, Patel R, Lauing KL, et al. IDO1 in cancer: a Gemini of immune checkpoints. *Cell Mol Immunol.* 2018 May;15(5):447–57.
150. Munn DH, Shafizadeh E, Attwood JT, Bondarev I, Pashine A, Mellor AL. Inhibition of T Cell Proliferation by Macrophage Tryptophan Catabolism. *J Exp Med.* 1999 May 3;189(9):1363–72.
151. Yue EW, Sparks R, Polam P, Modi D, Douty B, Wayland B, et al. INCB24360 (Epacadostat), a Highly Potent and Selective Indoleamine-2,3-dioxygenase 1 (IDO1) Inhibitor for Immuno-oncology. *ACS Med Chem Lett.* 2017 May 11;8(5):486–91.
152. Balog A, Lin T an, Maley D, Gullo-Brown J, Kandoussi EH, Zeng J, et al. Preclinical Characterization of Linrodostat Mesylate, a Novel, Potent, and

- Selective Oral Indoleamine 2,3-Dioxygenase 1 Inhibitor. *Mol Cancer Ther.* 2021 Mar 1;20(3):467–76.
153. Kumar S, Jaipuri FA, Waldo JP, Potturi H, Marcinowicz A, Adams J, et al. Discovery of indoximod prodrugs and characterization of clinical candidate NLG802. *Eur J Med Chem.* 2020 Jul 15;198:112373.
154. Nelp MT, Kates PA, Hunt JT, Newitt JA, Balog A, Maley D, et al. Immunomodulating enzyme indoleamine 2,3-dioxygenase is effectively inhibited by targeting its apo-form. *Proc Natl Acad Sci.* 2018 Mar 27;115(13):3249–54.
155. McKie SJ, Neuman KC, Maxwell A. DNA topoisomerases: Advances in understanding of cellular roles and multi-protein complexes via structure-function analysis. *BioEssays.* 2021;43(4):2000286.
156. Vann KR, Oviatt AA, Osheroff N. Topoisomerase II Poisons: Converting Essential Enzymes into Molecular Scissors. *Biochemistry.* 2021 Jun 1;60(21):1630–41.
157. Wendorff TJ, Schmidt BH, Heslop P, Austin CA, Berger JM. The Structure of DNA-Bound Human Topoisomerase II Alpha: Conformational Mechanisms for Coordinating Inter-Subunit Interactions with DNA Cleavage. *J Mol Biol.* 2012 Dec 7;424(3):109–24.
158. Shigematsu H, Ozaki S, Yasui D, Yamamoto H, Zaitso J, Taniyama D, et al. Overexpression of topoisomerase II alpha protein is a factor for poor prognosis in patients with luminal B breast cancer. *Oncotarget.* 2018 Jun 1;9(42):26701–10.
159. Matias-Barrios VM, Radaeva M, Ho CH, Lee J, Adomat H, Lallous N, et al. Optimization of New Catalytic Topoisomerase II Inhibitors as an Anti-Cancer Therapy. *Cancers.* 2021 Jan;13(15):3675.
160. Rosendahl AH, Perks CM, Zeng L, Markkula A, Simonsson M, Rose C, et al. Caffeine and Caffeic Acid Inhibit Growth and Modify Estrogen Receptor and Insulin-like Growth Factor I Receptor Levels in Human Breast Cancer. *Clin Cancer Res.* 2015 Apr 14;21(8):1877–87.
161. Magalhães CS de, Takarada JE, Carvalho NC, Carvalho D do C, de Andrade FL, Ferreira EB, et al. The Coffee Protective Effect on Catalase

- System in the Preneoplastic Induced Rat Liver. *J Chem.* 2016;2016(1):8570321.
162. Nazari PMS, Riedl J, Pabinger I, Ay C. The role of podoplanin in cancer-associated thrombosis. *Thromb Res.* 2018 Apr 1;164:S34–9.
163. Suzuki-Inoue K, Kato Y, Inoue O, Kaneko MK, Mishima K, Yatomi Y, et al. Involvement of the Snake Toxin Receptor CLEC-2, in Podoplanin-mediated Platelet Activation, by Cancer Cells *♦. *J Biol Chem.* 2007 Sep 7;282(36):25993–6001.
164. Sasaki T, Shirai T, Tsukiji N, Tamura S, Ichikawa J, Osada M, et al. Functional characterization of recombinant snake venom rhodocytin: rhodocytin mutant blocks CLEC-2/podoplanin-dependent platelet aggregation and lung metastasis. *J Thromb Haemost.* 2018 May 1;16(5):960–72.
165. Pan Y, Yago T, Fu J, Herzog B, McDaniel JM, Mehta-D’Souza P, et al. Podoplanin requires sialylated O-glycans for stable expression on lymphatic endothelial cells and for interaction with platelets. *Blood.* 2014 Dec 4;124(24):3656–65.
166. Simon T, Gagliano T, Giamas G. Direct Effects of Anti-Angiogenic Therapies on Tumor Cells: VEGF Signaling. *Trends Mol Med.* 2017 Mar 1;23(3):282–92.
167. Sharifi-Rad J, Rayess YE, Rizk AA, Sadaka C, Zgheib R, Zam W, et al. Turmeric and Its Major Compound Curcumin on Health: Bioactive Effects and Safety Profiles for Food, Pharmaceutical, Biotechnological and Medicinal Applications. *Front Pharmacol [Internet].* 2020 Sep 15 [cited 2024 Sep 10];11. Available from: <https://www.frontiersin.org/journals/pharmacology/articles/10.3389/fphar.2020.01021/full>
168. Kim JH, Shim JS, Lee SK, Kim KW, Rha SY, Chung HC, et al. Microarray-based Analysis of Anti-angiogenic Activity of Demethoxycurcumin on Human Umbilical Vein Endothelial Cells: Crucial Involvement of the Down-regulation of Matrix Metalloproteinase. *Jpn J Cancer Res.* 2002;93(12):1378–85.

169. Thuphairo K, Sornchan P, Suttisansanee U. Bioactive Compounds, Antioxidant Activity and Inhibition of Key Enzymes Relevant to Alzheimer's Disease from Sweet Pepper (*Capsicum annuum*) Extracts. *Prev Nutr Food Sci*. 2019 Sep;24(3):327–37.
170. Budryn G, Grzelczyk J, Jaśkiewicz A, Żyżelewicz D, Pérez-Sánchez H, Cerón-Carrasco JP. Evaluation of butyrylcholinesterase inhibitory activity by chlorogenic acids and coffee extracts assed in ITC and docking simulation models. *Food Res Int*. 2018 Jul 1;109:268–77.
171. Arya A, Chahal R, Rao R, Rahman MH, Kaushik D, Akhtar MF, et al. Acetylcholinesterase Inhibitory Potential of Various Sesquiterpene Analogues for Alzheimer's Disease Therapy. *Biomolecules*. 2021 Mar;11(3):350.
172. Jaipea S, Saehlim N, Sutcharitruk W, Athipornchai A, Ingkaninan K, Saeeng R. Synthesis of piperine analogues as AChE and BChE inhibitors for the treatment of Alzheimer's disease. *Phytochem Lett*. 2023 Feb 1;53:216–21.
173. Liu P, Cheng M, Guo J, Cao D, Luo J, Wan Y, et al. Dual functional antioxidant and butyrylcholinesterase inhibitors for the treatment of Alzheimer's disease: Design, synthesis and evaluation of novel melatonin-alkylbenzylamine hybrids. *Bioorg Med Chem*. 2023 Jan 15;78:117146.
174. Fukuyama K, Kakio S, Nakazawa Y, Kobata K, Funakoshi-Tago M, Suzuki T, et al. Roasted Coffee Reduces β -Amyloid Production by Increasing Proteasomal β -Secretase Degradation in Human Neuroblastoma SH-SY5Y Cells. *Mol Nutr Food Res*. 2018;62(21):1800238.
175. Mancini RS, Wang Y, Weaver DF. Phenylindanes in Brewed Coffee Inhibit Amyloid-Beta and Tau Aggregation. *Front Neurosci* [Internet]. 2018 Oct 12 [cited 2024 Sep 10];12. Available from: <https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2018.00735/full>
176. Smith JR, Jamie JF, Guillemin GJ. Kynurenine-3-monooxygenase: a review of structure, mechanism, and inhibitors. *Drug Discov Today*. 2016 Feb 1;21(2):315–24.

177. Savitz J. The kynurenine pathway: a finger in every pie. *Mol Psychiatry*. 2020 Jan;25(1):131–47.
178. Zhang S, Sakuma M, Deora GS, Levy CW, Klausing A, Breda C, et al. A brain-permeable inhibitor of the neurodegenerative disease target kynurenine 3-monooxygenase prevents accumulation of neurotoxic metabolites. *Commun Biol*. 2019 Jul 24;2(1):1–10.
179. Zwilling D, Huang SY, Sathyaikumar KV, Notarangelo FM, Guidetti P, Wu HQ, et al. Kynurenine 3-Monooxygenase Inhibition in Blood Ameliorates Neurodegeneration. *Cell*. 2011 Jun 10;145(6):863–74.
180. Cherrak SA, Merzouk H, Mokhtari-Soulimane N. Potential bioactive glycosylated flavonoids as SARS-CoV-2 main protease inhibitors: A molecular docking and simulation studies. *PLOS ONE*. 2020 Oct 15;15(10):e0240653.
181. Dias MC, Pinto DCGA, Silva AMS. Plant Flavonoids: Chemical Characteristics and Biological Activity. *Molecules*. 2021 Jan;26(17):5377.
182. Kennan A, Aherne A, Humphries P. Light in retinitis pigmentosa. *Trends Genet TIG*. 2005 Feb;21(2):103–10.
183. Mendes HF, van der Spuy J, Chapple JP, Cheetham ME. Mechanisms of cell death in rhodopsin retinitis pigmentosa: implications for therapy. *Trends Mol Med*. 2005 Apr;11(4):177–85.
184. Fanelli F, Seeber M. Structural insights into retinitis pigmentosa from unfolding simulations of rhodopsin mutants. *FASEB J Off Publ Fed Am Soc Exp Biol*. 2010 Sep;24(9):3196–209.
185. Ortega JT, Parmar T, Jastrzebska B. Flavonoids enhance rod opsin stability, folding, and self-association by directly binding to ligand-free opsin and modulating its conformation. *J Biol Chem*. 2019 May 17;294(20):8101–22.
186. Mostafa DG, Khaleel EF, Badi RM, Abdel-Aleem GA, Abdeen HM. Rutin hydrate inhibits apoptosis in the brains of cadmium chloride-treated rats via preserving the mitochondrial integrity and inhibiting endoplasmic reticulum stress. *Neurol Res*. 2019 Jul;41(7):594–608.
187. Perdices L, Fuentes-Broto L, Segura F, Cuenca N, Orduna-Hospital E, Pinilla I. Epigallocatechin Gallate Slows Retinal Degeneration, Reduces

- Oxidative Damage, and Modifies Circadian Rhythms in P23H Rats. *Antioxidants*. 2020 Aug 8;9(8):718.
188. Ortega JT, Parmar T, Golczak M, Jastrzebska B. Protective Effects of Flavonoids in Acute Models of Light-Induced Retinal Degeneration. *Mol Pharmacol*. 2021 Jan;99(1):60–77.
189. Pires DEV, Ascher DB, Blundell TL. DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. *Nucleic Acids Res*. 2014 Jul 1;42(W1):W314–9.
190. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr*. 1993;26(2):283–91.
191. Sánchez SM, Domínguez-Perles R, Montoro-García S, Gabaldón JA, Guy A, Durand T, et al. Bioavailable phytoprostanes and phytofurans from *Gracilaria longissima* have anti-inflammatory effects in endothelial cells. *Food Funct*. 2020 Jun 24;11(6):5166–78.
192. Lara-Guzmán OJ, Gil-Izquierdo Á, Medina S, Osorio E, Álvarez-Quintero R, Zuluaga N, et al. Oxidized LDL triggers changes in oxidative stress and inflammatory biomarkers in human macrophages. *Redox Biol*. 2018 May 1;15:1–11.
193. Markovič T, Jakopin Ž, Dolenc MS, Mlinarič-Raščan I. Structural features of subtype-selective EP receptor modulators. *Drug Discov Today*. 2017 Jan 1;22(1):57–71.
194. Iyú D, Jüttner M, Glenn JR, White AE, Johnson AJ, Fox SC, et al. PGE1 and PGE2 modify platelet function through different prostanoid receptors. *Prostaglandins Other Lipid Mediat*. 2011 Feb 1;94(1):9–16.
195. Medina S, Gil-Izquierdo Á, Durand T, Ferreres F, Domínguez-Perles R. Structural/Functional Matches and Divergences of Phytoprostanes and Phytofurans with Bioactive Human Oxylipins. *Antioxidants*. 2018 Nov;7(11):165.
196. Ong KL, Stafford LK, McLaughlin SA, Boyko EJ, Vollset SE, Smith AE, et al. Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021. *The Lancet*. 2023 Jul 15;402(10397):203–34.

197. Gourgari E, Wilhelm EE, Hassanzadeh H, Aroda VR, Shoulson I. A comprehensive review of the FDA-approved labels of diabetes drugs: Indications, safety, and emerging cardiovascular safety data. *J Diabetes Complications*. 2017 Dec 1;31(12):1719–27.
198. Foretz M, Guigas B, Viollet B. Metformin: update on mechanisms of action and repurposing potential. *Nat Rev Endocrinol*. 2023 Aug;19(8):460–76.
199. Dutta S, Shah RB, Singhal S, Dutta SB, Bansal S, Sinha S, et al. Metformin: A Review of Potential Mechanism and Therapeutic Utility Beyond Diabetes. *Drug Des Devel Ther*. 2023 Jun 26;17:1907–32.
200. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, et al. PubChem Substance and Compound databases. *Nucleic Acids Res*. 2016 Jan 4;44(D1):D1202–13.
201. Suntres ZE, Coccimiglio J, Alipour M. The Bioactivity and Toxicological Actions of Carvacrol. *Crit Rev Food Sci Nutr*. 2015 Feb 23;55(3):304–18.
202. Mir M, Permana AD, Ahmed N, Khan GM, Rehman A ur, Donnelly RF. Enhancement in site-specific delivery of carvacrol for potential treatment of infected wounds using infection responsive nanoparticles loaded into dissolving microneedles: A proof of concept study. *Eur J Pharm Biopharm*. 2020 Feb 1;147:57–68.
203. Singh AK, Yadav D, Sharma N, Jin JO. Dipeptidyl Peptidase (DPP)-IV Inhibitors with Antioxidant Potential Isolated from Natural Sources: A Novel Approach for the Management of Diabetes. *Pharmaceuticals*. 2021 Jun;14(6):586.
204. Bower AM, Real Hernandez LM, Berhow MA, de Mejia EG. Bioactive Compounds from Culinary Herbs Inhibit a Molecular Target for Type 2 Diabetes Management, Dipeptidyl Peptidase IV. *J Agric Food Chem*. 2014 Jul 2;62(26):6147–58.
205. Murray CJL, Ikuta KS, Sharara F, Swetschinski L, Robles Aguilar G, Gray A, et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet*. 2022 Feb;399(10325):629–55.
206. Monserrat-Martinez A, Gambin Y, Sierecki E. Thinking Outside the Bug: Molecular Targets and Strategies to Overcome Antibiotic Resistance. *Int J Mol Sci*. 2019 Jan;20(6):1255.

207. Gudmundsson GH, Agerberth B, Odeberg J, Bergman T, Olsson B, Salcedo R. The Human Gene FALL39 and Processing of the Cathelin Precursor to the Antibacterial Peptide LL-37 in Granulocytes. *Eur J Biochem.* 1996;238(2):325–32.
208. Mookherjee N, Anderson MA, Haagsman HP, Davidson DJ. Antimicrobial host defence peptides: functions and clinical potential. *Nat Rev Drug Discov.* 2020 May;19(5):311–32.
209. Raqib R, Sarker P, Bergman P, Ara G, Lindh M, Sack DA, et al. Improved outcome in shigellosis associated with butyrate induction of an endogenous peptide antibiotic. *Proc Natl Acad Sci.* 2006 Jun 13;103(24):9178–83.
210. Raqib R, Sarker P, Mily A, Alam NH, Arifuzzaman ASM, Rekha RS, et al. Efficacy of sodium butyrate adjunct therapy in shigellosis: a randomized, double-blind, placebo-controlled clinical trial. *BMC Infect Dis.* 2012 May 10;12:111.
211. Bekele A, Gebreselassie N, Ashenafi S, Kassa E, Aseffa G, Amogne W, et al. Daily adjunctive therapy with vitamin D3 and phenylbutyrate supports clinical recovery from pulmonary tuberculosis: a randomized controlled trial in Ethiopia. *J Intern Med.* 2018;284(3):292–306.
212. Mily A, Rekha RS, Kamal SMM, Arifuzzaman ASM, Rahim Z, Khan L, et al. Significant Effects of Oral Phenylbutyrate and Vitamin D3 Adjunctive Therapy in Pulmonary Tuberculosis: A Randomized Controlled Trial. *PLOS ONE.* 2015 Sep 22;10(9):e0138340.
213. Iannitti T, Palmieri B. Clinical and Experimental Applications of Sodium Phenylbutyrate. *Drugs RD.* 2011 Sep;11(3):227–49.
214. Miraglia E, Nylén F, Johansson K, Arnér E, Cebula M, Farmand S, et al. Entinostat up-regulates the CAMP gene encoding LL-37 via activation of STAT3 and HIF-1 α transcription factors. *Sci Rep.* 2016 Sep 16;6(1):33274.
215. Myszor IT, Parveen Z, Ottosson H, Bergman P, Agerberth B, Strömberg R, et al. Novel aroylated phenylenediamine compounds enhance antimicrobial defense and maintain airway epithelial barrier integrity. *Sci Rep.* 2019 May 8;9(1):7114.

216. Mariani V, Biasini M, Barbato A, Schwede T. IDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics*. 2013 Nov;29(21):2722–8.

IX – ANEXOS

IX - ANEXOS

ANEXO 1. Producción científica y calidad de las revistas

Artículo	Revista	Año	Índice de Impacto	Categoría	Cuartil
Curcumin-Based Inhibitors of Thrombosis and Cancer Metastasis Promoting Factor CLEC 2 from Traditional Medicinal Species Curcuma longa	Evidence-Based Complementary and Alternative Medicine	2022	2.65	INTEGRATIVE & COMPLEMENTARY MEDICINE	Q3
Flavonoids improve the stability and function of P23H rhodopsin slowing down the progression of retinitis pigmentosa in mice	Journal of Neuroscience Research	2022	4.2	NEUROSCIENCES	Q2
Physiologically relevant curcuminoids inhibit angiogenesis via VEGFR2 in human aortic endothelial cells	Food and Chemical Toxicology	2022	4.3	TOXICOLOGY	Q1
				FOOD SCIENCE & TECHNOLOGY	Q2

Evaluation of Activity of Sesquiterpene Lactones and Chicory Extracts as Acetylcholinesterase Inhibitors Assayed in Calorimetric and Docking Simulation Studies	Nutrients	2022	5.9	NUTRITION & DIETETICS	Q1
A Phytoprostane from <i>Gracilaria longissima</i> Increases Platelet Adhesion to Leukocytes and Endothelial Cell Migration by Potential Binding to EP3 Prostaglandin Receptor	INTERNATIONAL JOURNAL OF MOLECULAR SCIENCES	2023	4.9	BIOCHEMISTRY & MOLECULAR BIOLOGY	Q1
				CHEMISTRY, MULTIDISCIPLINARY	Q2
Aroylated phenylenediamine HO53 modulates innate immunity, histone acetylation and metabolism	Molecular Immunology	2023	3.2	BIOCHEMISTRY & MOLECULAR BIOLOGY	Q2
				IMMUNOLOGY	Q3

Effect of Inhibiting Butyrylcholinesterase Activity Using Fractionated Coffee Extracts Digested In Vitro in Gastrointestinal Tract: Docking Simulation and Calorimetric and Studies	Nutrients	2023	4.8	NUTRITION & DIETETICS	Q1
Computational insights on the antioxidant, antinitrosant, and xanthine oxidase inhibitory capacities of neobavaisoflavone	NEW JOURNAL OF CHEMISTRY	2023	2.7	CHEMISTRY, MULTIDISCIPLINARY	Q2
Effects of In Vitro Digestion of Polyphenols from Coffee on Binding Parameters to Human Topoisomerase II α	Molecules	2023	4.2	BIOCHEMISTRY & MOLECULAR BIOLOGY CHEMISTRY, MULTIDISCIPLINARY	Q2
Evaluation of mode of indoleamine 2,3-dioxygenase 1 inhibition by 4,7-dichloroquinolines	European Journal of Medicinal Chemistry Reports	2023	4.0	CHEMISTRY, MEDICINAL	Q2

Exploring the Antioxidant Properties of Caffeoylquinic and Feruloylquinic Acids: A Computational Study on Hydroperoxyl Radical Scavenging and Xanthine Oxidase Inhibition	Antioxidants	2023	6.0	CHEMISTRY, MEDICINAL	Q1
				FOOD SCIENCE & TECHNOLOGY	
				BIOCHEMISTRY & MOLECULAR BIOLOGY	
Analysis of AlphaFold and molecular dynamics structure predictions of mutations in serpins	PLoS One	2024	2.9	MULTIDISCIPLINARY SCIENCES	Q1
Comparative assessment of different anti-CD147/Basigin 2 antibodies as a potential therapeutic anticancer target by molecular modeling and dynamic simulation	MOLECULAR DIVERSITY	2024	3.9	CHEMISTRY, APPLIED	Q2
				CHEMISTRY, MULTIDISCIPLINARY	
				CHEMISTRY, MEDICINAL	

TOLEDO: Enhancing Maestro GUI for Non-Expert Users to Perform Massive MD Simulations	Journal of Biomolecular Structure and Dynamics	2024	2.7	BIOPHYSICS	Q2
				BIOCHEMISTRY & MOLECULAR BIOLOGY	Q3
Interaction of an anticancer oxygenated propenylbenzene derivatives with human topoisomerase II α and actin: molecular modeling and isothermal titration calorimetry studies	Journal of Thermal Analysis and Calorimetry	2024	3.0	CHEMISTRY, ANALYTICAL	Q1
				CHEMISTRY, PHYSICAL	Q3
				THERMODYNAMICS	Q2
Research into how carvacrol and metformin affect several human proteins in a hyperglycemic condition: A comparative study in silico and in vitro	Archives of Biochemistry and Biophysics	2024	3.8	BIOCHEMISTRY & MOLECULAR BIOLOGY	Q2
				BIOPHYSICS	Q1

Vitamin-C-dependent downregulation of the citrate metabolism pathway potentiates pancreatic ductal adenocarcinoma growth arrest	Molecular Oncology	2024	5.0	ONCOLOGY	Q1
Identification of potent inhibitors of kynurenine-3-monooxygenase from natural products: in silico and in vitro approaches	Heliyon	2024	3.4	MULTIDISCIPLINARY SCIENCES	Q1

ANEXO 2. Participaciones en Congresos

Comunicación	Congreso	Año	Organizador	Tipo de Comunicación	Tipo de Congreso
Optimización de simulaciones de dinámica molecular de forma paralela.	VII Jornadas de Investigación y Doctorado 'ODS con Ciencia'	2021	UCAM	Comunicación Oral	Regional
Optimization of molecular dynamic simulations using parallel programming.	VIII Bioinformatics Student Symposium		RSG-ISCB	Póster	Nacional
TOLEDO: a new software that is able to run and analyze Desmond Molecular Dynamics simulations using parallel programming.	II Symposium on Chemical and Physical Sciences for Young Researchers	2022	Universidad de Murcia	Comunicación Oral	Regional

Optimization of Molecular Dynamics Simulations Using Parallel Computing	Online Bilateral Bioinformatics Seminar		Dokuz Eylul University	Comunicación Oral	Internacional (Turquía)
TOLEDO: a tool that allows to run and analyze Desmond Molecular Dynamics simulations with higher throughput than Maestro GUI.	17th German Conference on Cheminformatics and EuroSAMPL Satellite Workshop	2022	German Chemical Society	Póster	Internacional (Alemania)
TOLEDO: una nueva herramienta para lanzar y analizar dinámicas moleculares de forma paralela con alta eficiencia a través de Maestro-Desmond	VIII Jornadas de Investigación y Doctorado: Ética en la Investigación Científica	2022	UCAM	Comunicación Oral	Regional

<p>TOLEDO. A new tool that allows for the improvement of the throughput of desmond molecular dynamics using HPC systems.</p>	<p>II Congreso Internacional de Investigación Biosanitaria para Jóvenes</p>	<p>2022</p>	<p>Universidad de Murcia</p>	<p>Comunicación Oral</p>	<p>Regional</p>
<p>TOLEDO: una herramienta que permite optimizar las simulaciones de Maestro-Desmond.</p>	<p>II Congreso Estatal de Estudiantes de Biociencias (CEEBI)</p>	<p>2023</p>	<p>CEEBI</p>	<p>Póster</p>	<p>Nacional</p>
<p>Optimización de simulaciones de dinámica molecular a través de la computación de altas prestaciones.</p>	<p>II Congreso Estatal de Estudiantes de Biociencias (CEEBI)</p>	<p>2023</p>	<p>CEEBI</p>	<p>Comunicación Oral</p>	<p>Nacional</p>

Aplicaciones del software TOLEDO para dinámica molecular en el ámbito del descubrimiento de fármacos.	IX Jornadas de Investigación y Doctorado: "Impacto Social de la Ciencia"	2023	UCAM	Comunicación Oral	Regional
Avances y Limitaciones de las Tecnologías de Predicción Proteica en el contexto de la trombosis	X Jornadas de Investigación y Doctorado "Bienestar Psicológico en los Estudios de Doctorado: repercusión en la Calidad de los Programas"	2024	UCAM	Comunicación Oral	Nacional
TOLEDO: Enhancing Tools for Extended Molecular Dynamics Simulation in Biomedical Research.	XVII Congreso Anual de Biotecnología (BAC Murcia 2024)		BiotecMur		

ANEXO 3. Software y licencias

Los cálculos realizados durante esta tesis han requerido la instalación de diversos software, algunos de los cuales tienen licencia libre y otros que han requerido la instalación de una licencia.

Los cálculos de *Docking* y similaridad farmacofórica fueron realizados mediante el uso del software MetaScreener (diseñado por nuestro grupo). Este software se ha descargado en cada clúster mediante las instrucciones presentes en su GitHub (<https://github.com/bio-hpc/metascreeener>). MetaScreener se puede utilizar para lanzar cálculos usando los siguientes software:

- AutoDock Vina: Este software es de licencia libre y puede ser utilizado por cualquier usuario. La licencia es Apache 2.0.
- Lead Finder: Este software requiere de licencia para poder ser ejecutado, habiendo dos tipos de licencia: académica y comercial. Para actualizar la licencia se introduce el siguiente comando desde el directorio `metascreeener/MetaScreener`:

```
./external_sw/leadFinder/leadfinder -lic-update File.lic
```

Siendo el archivo `File.lic` el archivo con la licencia.
- Ligand Scout: Este software también requiere de licencia para poder ser ejecutado. Para actualizar la licencia se introduce el siguiente comando desde el directorio `metascreeener/MetaScreener`:

```
./external_sw/ligandScout/ligandscout_activation -s XXXXXXXXXX
```

Siendo `XXXXXXXXXX` el serial que da la licencia.

AutoDock Vina está instalado por defecto al descargar MetaScreener, los software Lead Finder y Ligand Scout deben ser instalados por el usuario en el directorio `metascreeener/Metascreener/external_sw/`.

Los cálculos de Dinámica Molecular fueron realizados utilizando el motor de Dinámica Molecular Desmond, para ejecutar este software es necesario instalar la licencia de Maestro-Desmond.

Para instalar la licencia en el cliente (*host*), se puede hacer utilizando una clave o un archivo de licencia, en ambos casos el usuario deber ser administrador o superusuario:

- Con archivo de licencia: Se puede realizar tanto usando la interfaz gráfica de usuario como la línea de comandos.
 - Interfaz gráfica de usuario:

Se abre el software de configuración de Maestro, se selecciona la pestaña “*Help*” y dentro de ella la opción: “*Configure Licensing ...*”. Se abre una ventana para configurar la licencia y en la pestaña “*Add Licenses*”, se selecciona la opción “*I have a license file*”, se pulsa el botón “*Browse*”, se añade el archivo con la licencia y se pulsa en “*Install License*”.
 - Línea de comandos:

En el directorio con la instalación de Maestro-Desmond, se ejecuta el siguiente comando:

```
$SCHRODINGER/licadmin INSTALL -c archivo_licencia.txt
```

Siendo `archivo_licencia.txt` el directorio de forma absoluta o relativa donde se encuentra el archivo de licencia.
- Con clave: Al igual que con el archivo de licencia, se puede realizar tanto usando la interfaz gráfica de usuario como la línea de comandos.
 - Interfaz gráfica de usuario:

Se abre el software de configuración de Maestro, se selecciona la pestaña “*Help*” y dentro de ella la opción: “*Configure Licensing ...*”. Se abre una ventana para configurar la licencia y en la pestaña “*Add Licenses*”, se selecciona la opción “*I have a key*”, se introduce la clave en el cuadro de texto y se pulsa en “*Install License*”.

- Línea de comandos:

En el directorio con la instalación de Maestro-Desmond, se ejecuta el siguiente comando:

```
§SCHRODINGER/licadmin INSTALL -k XXXXX
```

Siendo XXXXX la clave que contiene la licencia.

Para instalar la licencia en otros ordenadores que no sean el cliente principal, se debe indicar en qué servidor se tiene instalado la licencia principal. Para ello, se utiliza la interfaz gráfica. Se abre el software de configuración de Maestro, se selecciona la pestaña *“Help”* y dentro de ella la opción: *“Configure Licensing ...”*. Se abre una ventana para configurar la licencia y en la pestaña *“Add Licenses”*, se selecciona la opción *“I can identify my license server”*, se selecciona el servidor donde se tiene instalada la licencia y se añade el puerto 27008. Es importante que el servidor principal tenga abierto el puerto 53000 para poder hacer la conexión entre ambas máquinas.

El software de Maestro-Desmond debe actualizarse, incluyendo las licencias para usar TOLEDO.

