



UNIVERSIDAD MICHOACANA DE SAN NICOLÁS DE HIDALGO
PROGRAMA INSTITUCIONAL DE MAESTRÍA EN CIENCIAS BIOLÓGICAS
FACULTAD DE MEDICINA VETERINARIA Y ZOOTECNIA

**“CARACTERIZACIÓN DEL ESPACIO-QUÍMICO Y DISEÑO DE
NUEVOS INHIBIDORES DE LA PROTEASA PRINCIPAL MPRO Y
LA PROTEASA TIPO PAPAÍNA PL-PRO DE SARS-COV-2
MEDIANTE ANÁLISIS QUIMIOINFORMÁTICO Y DE
MODELADO MOLECULAR”**

Tesis que presenta:

LB Alberto Zavala Chávez

como requisito parcial para obtener el título profesional de

Maestro en Ciencias Biológicas

en el área de Microbiología Celular y Genética Molecular

Director de Tesis:

Dr. Víctor Manuel Baizabal Aguirre

Co-Director de Tesis:

Dr. José Luis Medina Franco

Morelia, Michoacán, México. Julio del 2023

DEDICATORIA

A mis padres, abuelos y hermana.

AGRADECIMIENTOS

D.C. Víctor Manuel Baizabal Aguirre

Gracias por la oportunidad de entrar en el laboratorio, encaminarme hacia nuevos desafíos en mi formación y apoyarme en mi crecimiento profesional a través de la impartición de cursos especializados, así como la resolución de dudas.

D.C. José Luis Medina Franco

Gracias por la atención y asesoría brindada en las metodologías de Quimioinformática, agradezco también los consejos profesionales que llevaré siempre presentes en mi desarrollo como profesionalista.

D. C. José Manuel Pérez Aguilar

Gracias por su invaluable ayuda con el aprendizaje de metodologías de modelado y dinámica molecular. También agradezco su siempre tan amable disposición para apoyarme en las dudas que surgieron en el camino.

D. C. Juan José Valdez Alarcón y D. C. José Miguel Cervantes Alfaro

Gracias por la atención brindada en cada una de mis evaluaciones, por sus observaciones y consejos para mejorar mi proyecto, así como el tiempo dedicado a la revisión del mismo.

A mis compañeros de laboratorio

Marco e Isaías, les agradezco nunca dejarme solo en las adversidades de mis estudios. Siempre siendo una guía y un apoyo dentro y fuera de lo académico.

A mis padres y hermana

Por siempre creer en mí.

TABLA DE CONTENIDO

1. RESUMEN	1
2. ABSTRACT	3
3. INTRODUCCIÓN	5
4.- MARCO TEÓRICO	6
4.1 ORGANIZACIÓN GENÓMICA DE SARS-COV-2	6
4.2 ESTRUCTURA DE SARS-COV-2.....	8
4.3 CICLO DE REPLICACIÓN DE SARS-COV-2	9
4.4 ESTRUCTURA Y FUNCIÓN DE LA PROTEASA PRINCIPAL (M _{PRO}) DE SARS-COV-2.....	11
4.5 ESTRUCTURA Y FUNCIÓN DE LA PROTEASA DE TIPO PAPAÍNA (P _{LPRO}) DE SARS-COV-2	14
4.6 INFORMÁTICA Y SU IMPORTANCIA EN LA INVESTIGACIÓN ACTUAL	15
4.7 LA QUIMIOINFORMÁTICA COMO DISCIPLINA EN EL DESCUBRIMIENTO Y DISEÑO DE FÁRMACOS.....	16
4.8 EL CONCEPTO DE ESPACIO QUÍMICO	18
4.9 PROPIEDADES FÍSICOQUÍMICAS DE INTERÉS FARMACEÚTICO Y SU PAPEL COMO DESCRIPTORES MOLECULARES PARA LA SELECCIÓN DE NUEVOS COMPUESTOS	20
4.10 LA IMPORTANCIA DEL CURADO DE BASES DE DATOS DE COMPUESTOS	23
4.11 EL CONCEPTO DE “TIPO-FÁRMACO” Y SU IMPORTANCIA EN EL DESARROLLO DE REGLAS PARA EL FILTRADO DE BASES DE DATOS.....	24
4.12 MODELADO MOLECULAR EN EL DESCUBRIMIENTO Y DESARROLLO DE FÁRMACOS.....	26
4.13 SIMULACIONES DE ACOPLAMIENTO MOLECULAR Y SU PODER PREDICTIVO	28
4.14 MODELO DE FARMACÓFORO	29
4.15 ENSAYOS DE SIMULACIÓN ATOMÍSTICA DE DINÁMICA MOLECULAR	31
5. JUSTIFICACIÓN	34
6. HIPÓTESIS	35
7. OBJETIVOS	35
7.1 OBJETIVO GENERAL.....	35
7.2 OBJETIVOS ESPECÍFICOS.....	36
8.- ESTRATEGIA EXPERIMENTAL GENERAL	37

9.- MATERIALES Y MÉTODOS	37
9.1 IDENTIFICACIÓN DE COMPUESTOS ORGÁNICOS INHIBIDORES DE M _{PRO} Y PL _{PRO} REPORTADOS EN LA LITERATURA	37
9.2 CURADO DE LA BASE DE DATOS (ALVAMOLECULE).....	38
9.3 CÁLCULO DE PROPIEDADES FISCOQUÍMICAS DE INTERÉS FARMACEÚTICO.....	39
9.4 DIVERSIDAD MOLECULAR POR DISTANCIA EUCLIDIANA E ÍNDICE DE TANIMOTO.....	40
9.5 CARACTERIZACIÓN DEL ESPACIO QUÍMICO Y ANÁLISIS QUIMIOINFORMÁTICOS CON PLATFORM FOR UNIFIED MOLECULAR ANALYSIS (PUMA)	42
9.6 ANÁLISIS DE SCAFFOLDS POR CURVAS DE RECUPERACIÓN DE SISTEMAS CÍCLICOS (CSR) Y ENTROPÍA DE SHANNON	43
9.7 ACOPLAMIENTO MOLECULAR CON <i>GLIDE</i>	45
9.8 GENERACIÓN DE MODELO DE FARMACÓFORO	45
9.9 SIMULACIONES ATOMÍSTICAS DE DINÁMICA MOLECULAR DE LA INTERACCIÓN ENTRE LA PROTEASA PRINCIPAL M _{PRO} DE SARS-COV-2 Y COMPUESTOS <i>LEADS</i> CON <i>NAMD</i>	46
10.- RESULTADOS	47
10.1 IDENTIFICACIÓN DE COMPUESTOS ORGÁNICOS INHIBIDORES DE M _{PRO} Y PL _{PRO} REPORTADOS EN LA LITERATURA Y CREACIÓN DE QUIMIOTECA	47
10.2 CÁLCULO DE PROPIEDADES FISCOQUÍMICAS DE INTERÉS FARMACEÚTICO DE LOS COMPUESTOS REPORTADOS	48
10.3 VISUALIZACIÓN DEL ESPACIO QUÍMICO BASADO EN LAS SEIS PROPIEDADES FISCOQUÍMICAS MW, HB _{Acc} , HB _{Don} , nR _{otB} , ALogP Y TPSA	52
10.4 ANÁLISIS DE DIVERSIDAD MOLECULAR BASADO EN COMPARACIONES SISTEMÁTICAS POR PARES.....	56
10.5 ANÁLISIS DE LA DIVERSIDAD DE SCAFFOLDS MEDIANTE EL USO DE CURVAS DE RECUPERACIÓN DE SISTEMAS CÍCLICOS (CSR) Y ENTROPÍA DE SHANNON.....	58
10.6 DIVERSIDAD ESTRUCTURAL DE MOLÉCULAS COMPLETAS UTILIZANDO HUELLAS DACTILARES MOLECULARES (STRUCTURAL MOLECULAR FINGERPRINTS) E ÍNDICE DE SIMILITUD DE TANIMOTO.....	64
10.7 VALIDACIÓN DEL PROTOCOLO PARA ENSAYOS DE ACOPLAMIENTO MOLECULAR CON <i>GLIDE</i> DE SCHRÖDINGER	66
10.8 VALIDACIÓN PARA LA PROTEASA PRINCIPAL M _{PRO} EN COMPLEJO CON MASITINIB	67

10.9 VALIDACIÓN PARA LA PROTEASA DE TIPO PAPAÍNA PLPRO EN COMPLEJO CON EL INHIBIDOR GRL-0617.....	69
10.10 ANÁLISIS DE PROPIEDADES FISICOQUÍMICAS DE INTERÉS FARMACÉUTICO DE LOS COMPUESTOS REPORTADOS CON AFINIDAD POR LA PROTEASA PRINCIPAL MPRO	71
10.11 ANÁLISIS DE DIVERSIDAD MOLECULAR BASADO EN COMPARACIONES SISTEMÁTICAS POR PARES.....	75
10.12 ANÁLISIS DE LA DIVERSIDAD DE SCAFFOLDS MEDIANTE EL USO DE CURVAS DE RECUPERACIÓN DE SISTEMAS CÍCLICOS (CSR) Y ENTROPÍA DE SHANNON PARA COMPUESTOS CON AFINIDAD POR MPRO.....	76
10.13 ANÁLISIS DE PROPIEDADES FISICOQUÍMICAS DE INTERÉS FARMACEÚTICO PARA LOS COMPUESTOS REPORTADOS CON AFINIDAD POR LA PROTEASA DE TIPO PAPAÍNA PLPRO	80
10.14 ANÁLISIS DE DIVERSIDAD MOLECULAR BASADO EN COMPARACIONES SISTEMÁTICAS POR PARES.....	83
10.15 ANÁLISIS DE LA DIVERSIDAD DE ESCAFOLDES MEDIANTE EL USO DE CURVAS DE RECUPERACIÓN DE SISTEMAS CÍCLICOS (CSR) Y ENTROPÍA DE SHANNON PARA COMPUESTOS CON AFINIDAD POR PLPRO	83
10.16 VISUALIZACIÓN DEL ESPACIO QUÍMICO BASADO EN LAS SEIS PROPIEDADES FISICOQUÍMICAS MW, HBACC, HBDON, NROTB, ALOGP Y TPSA DE COMPUESTOS CON AFINIDAD POR LA PROTEASA PRINCIPAL MPRO	88
10.17 MODELO DE FARMACÓFORO BASADO EN CARACTERÍSTICAS MOLÉCULARES DE LOS MEJORES CON MEJOR ACOPLAMIENTO CON MPRO	98
10.18 SIMULACIÓN ATOMÍSTICA POR DINÁMICA MOLECULAR DE COMPUESTO COMP83 CON EL SITIO ACTIVO DE MPRO.....	102
10.19 VISUALIZACIÓN DEL ESPACIO QUÍMICO BASADO EN LAS SEIS PROPIEDADES FISICOQUÍMICAS MW, HBACC, HBDON, NROTB, ALOGP Y TPSA DE COMPUESTOS CON AFINIDAD POR LA PROTEASA DE TIPO PAPAÍNA PLPRO	107
11.- DISCUSIÓN.....	115
12.- CONCLUSIÓN.....	124
13.- BIBLIOGRAFÍA.....	125
14.- MATERIAL SUPLEMENTARIO.....	139

INDICE DE FIGURAS

1. Organización del genoma de SARS-CoV-2	7
2. Estructura general de SARS-CoV-2	9
3. Ciclo de replicación de SARS-CoV-2	11
4. Estructura de la proteasa principal (Mpro) de SARS-CoV-2	13
5. Estructura de la proteasa tipo papaína (PLpro) de SARS-CoV-2	15
6. Esquema representativo de la interacción entre los sistemas computacionales y su papel en el diseño de moléculas activas	28
7. Dos modelos de acoplamiento molecular	29
8. Características farmacofóricas principales	31
9. Diagrama de flujo de trabajo	37
10. Fórmulas para el cálculo de prueba de sumas de Rangos de Wilcoxon	40
11. Histogramas de propiedades fisicoquímicas de compuestos con afinidad por Mpro y PLpro	49
12. Espacio químico de compuestos con afinidad por Mpro y PLpro	54
13. Estructura de compuestos que representan datos atípicos en el espacio químico de moléculas con afinidad por Mpro	55
14. Estructura de compuestos que representan datos atípicos en el espacio químico de moléculas con afinidad por PLpro	56
15. Distancia euclidiana e índice de Tanimoto de compuestos con afinidad por Mpro y PLpro	57
16. Distribución de los 10 quimiotipos más frecuentes de compuestos con afinidad por Mpro y PLpro	62
17. Distribución de los 20 quimiotipos más frecuentes de compuestos con afinidad por Mpro y PLpro	62
18. Quimiotipos más frecuentes de moléculas con afinidad por Mpro	63
19. Quimiotipos más frecuentes de moléculas con afinidad por PLpro	63
20. Curvas CDF para la similitud por pares utilizando huellas dactilares moleculares	65
21. Mejor acoplamiento entre Masitinib-Mpro para la validación de protocolo para simulaciones de acoplamiento molecular	68
22. Interacciones intermoleculares del mejor acoplamiento Masitinib-Mpro	68

23. Mejor acoplamiento entre GRL-0617-PLpro para la validación de protocolo para simulaciones de acoplamiento molecular	70
24. Interacciones intermoleculares del mejor acoplamiento GRL-0617-PLpro	70
25. Histogramas de propiedades fisicoquímicas de compuestos con afinidad por Mpro	73
26. Gráficos de cajas de propiedades fisicoquímicas de compuestos con afinidad por Mpro	74
27. Matrices de (dis)similitud por distancia Euclidiana e índice de Tanimoto de compuestos con afinidad por Mpro	75
28. Distribución y estructura de quimiotipos más frecuentes de compuestos con actividad biológica contra Mpro	78
29. Distribución y estructura de quimiotipos más frecuentes de compuestos con afinidad predicha computacionalmente contra Mpro	79
30. Distribución y estructura de quimiotipos más frecuentes de fármacos reposicionados contra Mpro	79
31. Histograma de propiedades fisicoquímicas de compuestos con afinidad por PLpro	81
32. Gráficos de cajas de propiedades fisicoquímicas de compuestos con afinidad por PLpro	82
33. Matrices de (dis)similitud por distancia Euclidiana e índice de Tanimoto de compuestos con afinidad por PLpro	83
34. Distribución y estructura de quimiotipos más frecuentes de compuestos con actividad biológica contra PLpro	86
35. Distribución y estructura de quimiotipos más frecuentes de compuestos con afinidad predicha computacionalmente contra PLpro	87
36. Distribución y estructura de quimiotipos más frecuentes de fármacos reposicionados contra PLpro	87
37. Espacio químico en PUMA de compuestos con afinidad por Mpro	89
38. Estructura química y propiedades fisicoquímicas de Masitinib y Paxlovid	90
39. Espacio químico 2D y 3D en DataWarrior de compuestos con afinidad por Mpro	91
40. Espacio químico en 3D en DataWarrior de compuestos con afinidad predicha por métodos computacionales contra Mpro	92

41. Espacio químico en 3D en DataWarrior de compuestos con actividad biológica reportada contra Mpro	92
42. Espacio químico en 3D en DataWarrior de fármacos reposicionados contra Mpro	93
43. Acoplamiento del compuesto MP201 en el sitio activo de Mpro	94
44. Acoplamiento del compuesto RMP26 en el sitio activo de Mpro	95
45. Familia de compuestos inhibidores de Mpro de SARS-CoV provenientes del trabajo de Turlington <i>et al.</i> (2013)	96
46. Acoplamiento del compuesto MP462 en el sitio activo de Mpro	97
47. Acoplamiento del compuesto MP463 en el sitio activo de Mpro	97
48. Acoplamiento del compuesto MP470 en el sitio activo de Mpro	98
49. Modelo de farmacóforo para buscar nuevas moléculas con afinidad por Mpro	99
50. Acoplamiento del compuesto COMP90 (ChEMBL92481) en el sitio activo de Mpro	100
51. Acoplamiento del compuesto COMP83 (ChEMBL199740) en el sitio activo de Mpro	100
52. Acoplamiento del compuesto COMP70 (ChEMBL507400) en el sitio activo de Mpro	101
53. Sistema COMP83/Mpro en su forma dimérica	103
54. Interacciones entre COMP83 y los residuos de aminoácidos del monómero A en el sitio catalítico de Mpro	104
55. Representación 2D de COMP83	105
56. Estructura química y propiedades fisicoquímicas de GRL-0617	108
57. Espacio químico en PUMA de compuestos con afinidad por PLpro	109
58. Acercamiento en el espacio químico de compuestos con actividad biológica contra PLpro	110
59. Espacio químico 2D y 3D en DataWarrior de compuestos con afinidad por PLpro	111
60. Espacio químico en 3D en DataWarrior de compuestos con afinidad predicha por métodos computacionales y fármacos reposicionados contra Mpro	111
61. Espacio químico en 3D en DataWarrior de compuestos con actividad biológica reportada contra PLpro	112

62. Acoplamiento del compuesto COMP11 (ChEMBL559234) en el sitio activo de PLpro	113
63. Acoplamiento del compuesto COMP2 (ChEMBL561061) en el sitio activo de PLpro	114

ÍNDICE DE TABLAS

1. Blanco biológico y número de compuestos obtenidos de la literatura científica	48
2. Distribución estadística de propiedades fisicoquímicas para compuestos afines a Mpro y PLpro	49
3. Contribución de cada PC en el espacio químico de compuestos con afinidad por las proteasas de SARS-CoV-2	54
4. Recuento de quimiotipos en la comparativa de compuestos con afinidad por las proteasas de SARS-CoV-2	59
5. Valores de entropía de Shannon de los 10-60 quimiotipos únicos más poblados	61
6. Distribución estadística de similitud por pares utilizando huellas dactilares moleculares ...	65
7. Valores de RMSD de mejores acoplamientos entre Masitinib-Mpro	68
8. Valores de RMSD de mejores acoplamientos entre GRL-0617-PLpro	70
9. Propiedades fisicoquímicas de Masitinib y Paxlovid	73
10. Distribución estadística de propiedades fisicoquímicas de compuestos con afinidad por Mpro	74
11. Recuento de quimiotipos para compuestos con afinidad por Mpro	77
12. Valores de SSE de los 10-60 quimiotipos únicos más poblados de compuestos con afinidad por Mpro	78
13. Distribución estadística de propiedades fisicoquímicas de compuestos con afinidad por PLpro	81
14. Propiedades fisicoquímicas de GRL-0617	82
15. Recuento de quimiotipos para compuestos con afinidad por PLpro	85
16. Valores de SSE de los 10-60 quimiotipos únicos más poblados de compuestos con afinidad por PLpro	86
17. Contribución de las seis propiedades fisicoquímicas a cada PC en el espacio químico de Mpro	89

18. Contribución de las seis propiedades fisicoquímicas a cada PC en el espacio químico de PLpro	109
--	-----

ÍNDICE DE GRÁFICAS

1. Peso molecular (MW) de compuestos con afinidad por Mpro	51
2. Peso molecular (MW) de compuestos con afinidad por PLpro	51
3. Curvas CSR para compuestos con afinidad por las proteasas de SARS-CoV-2	59
4. Curvas CSR de compuestos con afinidad por Mpro	77
5. Curvas CSR de compuestos con afinidad por PLpro	84
6. Cálculo de RMSD general de los átomos pesados de COMP83 durante 150 ns de simulación	104
7. Cálculo de distancias entre átomos de COMP83 y residuos del sitio activo de Mpro	106

ÍNDICE DE ABREVIATURAS

S: Proteína Spike o de pico de SARS-CoV-2

M: Proteína de membrana de SARS-CoV-2

E: Proteína de envoltura de SARS-CoV-2

N: Proteína de nucleocápside de SARS-CoV-2

HE: Dímeros de hemaglutinina-esterasa

CRT: Complejo de replicación-transcripción

MW: Peso molecular

LogP: Coeficiente de partición octanol/agua

HBDon: Átomos donadores de hidrógeno

HBAcc: Átomos aceptores de hidrógeno

TPSA: Área superficial polar topológica

nRotB: Número de enlaces rotables

PCA: Análisis de componentes principales

DM: Dinámica molecular

CSR: Curvas de recuperación de sistemas cíclicos

DIFAC: Diseño de fármacos asistido por computadora

1. RESUMEN

La magnitud de la pandemia por COVID-19, el aislamiento masivo y la aparente facilidad de uso de métodos computacionales ha incentivado a un gran número de grupos científicos a utilizar dichos métodos para la búsqueda *in silico* de fármacos contra COVID-19. La aplicación de diferentes herramientas *software* ha permitido probar una enorme cantidad de compuestos químicos mediante diferentes técnicas computacionales, como son el modelado y acoplamiento molecular y las simulaciones atomísticas de dinámica molecular para encontrar un fármaco capaz de inactivar al virus SARS-CoV-2 y así poner fin a la pandemia por COVID-19. Una pandemia que aún en 2023, y con ayuda de la vacunación, sigue cobrando la vida de muchas personas en el mundo. El SARS-CoV-2 contiene un genoma de ARN de 30 Kb, que codifica para 4 proteínas estructurales, 9 accesorias y 16 no estructurales que se liberan por autoproteólisis de dos poliproteínas con la ayuda de las proteasas virales Mpro y PLpro (Gordon *et al.*, 2020). Estas enzimas han sido estudiadas con profundidad y sus estructuras obtenidas por cristalografía de rayos X han permitido que, mediante técnicas de acoplamiento molecular, se conozcan las estructuras químicas de una cantidad considerable de compuestos orgánicos de bajo peso molecular con afinidad por sus sitios activos. Tal información abre una ventana de oportunidad para el descubrimiento de patrones estructurales asociados con la actividad biológica, mediante la aplicación de técnicas de quimiinformática. En este contexto, el espacio químico ha proporcionado un marco conceptual sólido para guiar el análisis de la diversidad, la clasificación de estructuras, el diseño de quimiotecas, la selección y la evaluación de las relaciones estructura-actividad dentro de uno o más grupos de compuestos químicos (Medina-Franco *et al.*, 2008). El espacio químico, que puede conceptualizarse como análogo al universo cosmológico en su inmensidad, permite organizar la diversidad molecular, postulando que moléculas diferentes ocupan regiones diferentes de un espacio matemático, donde la posición de cada molécula está definida por sus propiedades (Arús-Pous *et al.*, 2019). La visualización y análisis del espacio químico definido por compuestos reportados con afinidad por las proteasas Mpro y PLpro, nos ha

permitido, en el presente trabajo, dilucidar patrones de estructura en compuestos que se unen con alta afinidad a los sitios catalíticos de estas enzimas. Estos patrones se utilizaron como base para la búsqueda de nuevas moléculas mediante técnicas computacionales de quimioinformática y modelado molecular. Resultando en la obtención de dos moléculas orgánicas de bajo peso molecular candidatas a probarse por métodos experimentales para inhibir a la proteasa principal Mpro y otras dos para inhibir a la proteasa de tipo papaína PLpro impidiendo como consecuencia la replicación de SARS-CoV-2.

Palabras Clave: Mpro, PLpro, Quimioinformática, Espacio Químico, Acoplamiento Molecular, Proteasas virales, Simulaciones de Dinámica Molecular.

2. ABSTRACT

The magnitude of the COVID-19 pandemic, the massive isolation as well as the increasing development not only in computational infrastructure but also in specialized user-friendly software, have encouraged a large number of scientific groups to use computational methods for the *in silico* search for drugs against SARS-CoV-2. The application of different software has made it possible to test an enormous number of chemical compounds using different computational techniques, such as molecular modeling and docking and atomistic molecular dynamics simulations to find a drug capable of inactivating the SARS-CoV-2 virus and thus putting an end to the COVID-19 pandemic. A pandemic that even in 2023, and with the help of vaccination, is still claiming the lives of many people around the world. SARS-CoV-2 contains a 30 Kb RNA genome, encoding 4 structural, 9 accessory, and 16 non-structural proteins that are released by autoproteolysis of two polyproteins with the help of the viral proteases Mpro and PLpro (Gordon *et al.*, 2020). These enzymes have been studied in depth and their tertiary structures obtained by X-ray crystallography have allowed, by means of molecular docking techniques, to know the chemical structures of a considerable amount of low molecular weight organic compounds with affinity for their active sites. Such information opens a window of opportunity for the discovery of structural patterns associated with biological activity through the application of Chemoinformatic techniques. In this context, chemical space has provided a solid conceptual framework to guide diversity analysis, structure classification, compound library design, screening and evaluation of structure-activity relationships within one or more groups of chemical compounds (Medina-Franco *et al.*, 2008). Chemical space, which can be conceptualized as analogous to the cosmological universe in its immensity, makes it possible to organize molecular diversity, postulating that different molecules occupy different regions of a mathematical space, where the position of each molecule is defined by its properties (Arús-Pous *et al.*, 2019). In the present work, the visualization and analysis of the chemical space defined by compounds reported with affinity for Mpro and PLpro proteases, has allowed us to

elucidate structural patterns of compounds that bind with high affinity to the catalytic sites of these enzymes. These patterns were used as the basis for the search of new molecules via the use of the Chemoinformatics and molecular modeling computational techniques. As a result, two low molecular weight organic molecules were obtained as candidates to be tested by experimental methods to inhibit the main protease Mpro and two others to inhibit the papain-like protease PLpro, thus preventing the replication of SARS-CoV-2.

Keywords: Mpro, PLpro, Chemoinformatics, Chemical Space, Molecular Docking, Viral proteases, Molecular Dynamics Simulations.

3. INTRODUCCIÓN

La pandemia ocasionada por la liberación repentina de un nuevo coronavirus surgido en el año 2019 ha traído consigo una cantidad importante de dificultades para el mundo entero en diferentes sectores laborales y sociales, y a la comunidad científica por igual. Una de las medidas de acción tomadas por los diferentes gobiernos para frenar el impacto letal del coronavirus denominado SARS-CoV-2 por el Comité Internacional de Taxonomía de Virus (ICTV), causante de la enfermedad COVID-19 fue el aislamiento masivo. Debido a esto, millones de personas en todo el mundo fueron incapaces de continuar sus labores si éstas no podían llevarse a cabo de manera virtual y segura desde sus hogares. Por otra parte, a pesar del esfuerzo de las instancias gubernamentales y la industria farmacéutica por desarrollar nuevas terapias y vacunas para detener el contagio de las personas, el número de contagios y muertes sigue aumentando. Otro enfoque alternativo ha sido el descubrimiento o reposicionamiento y desarrollo de un fármaco capaz de inactivar al virus. Los métodos dirigidos a la búsqueda y diseño de nuevos fármacos asistido por computadora (*Computer-Aided Drug Design*, CADD, por sus siglas en inglés) han demostrado ser invaluable en la búsqueda de moléculas pequeñas con el potencial de curar diversas enfermedades. La magnitud de la pandemia por COVID-19, junto con las condiciones de aislamiento y aparente facilidad de uso de los métodos computacionales ha incentivado a que un gran número de científicos se acerquen a estas metodologías para probar una enorme cantidad de compuestos químicos con la finalidad de encontrar un fármaco capaz de inactivar al virus y con esto poner fin a la pandemia. Esto es aparentemente una buena noticia, no obstante, en la actualidad no se cuenta más que con únicamente un fármaco aprobado (Paxlovid, *Pfizer*). Lo que ha generado el uso extenso de las técnicas computacionales es una gran cantidad de información reportada y dispersa en la literatura sobre resultados de acoplamiento molecular de moléculas con potencial para inhibir a las proteasas principal (Mpro o 3CL-pro) y la proteasa de tipo papaína (PLpro) de SARS-CoV-2 (Muratov *et al.*, 2021). Por lo tanto, nuestro interés se centra en recopilar de la literatura científica la mayor cantidad de moléculas

reportadas como líderes para inhibir cualquiera de las proteasas y estudiarlas, mediante diversos análisis quimioinformáticos con la finalidad de dilucidar diferentes patrones de estructura que estén relacionados con propiedades fisicoquímicas deseables y con alta actividad biológica. De esta manera, con la ayuda de dichos patrones estructurales se podrá hacer una búsqueda de nuevas moléculas que presenten alta afinidad por Mpro y PLpro, con el potencial de probarse por métodos computacionales de dinámica molecular y enzimáticos en el laboratorio, resultando en nuevas moléculas líderes, susceptibles de ser optimizadas, desarrolladas y llevadas a fase preclínica como potenciales fármacos contra COVID-19.

4. MARCO TEÓRICO

4.1 ORGANIZACIÓN GENÓMICA DE SARS-COV-2

Los coronavirus son un grupo de virus de ARN monocatenario en sentido positivo con envoltura. Pertenecen al orden de los Nidovirales y a la familia *Coronaviridae*. Se conocen dos docenas de especies diferentes y se han dividido en cuatro géneros: alfa, beta, gamma y delta, de los cuales solo los géneros alfa y beta incluyen cepas patógenas para humanos y otros mamíferos (Kim *et al.*, 2016; Guo *et al.*, 2020).

En los últimos años, algunos coronavirus pertenecientes al género beta, a saber, el coronavirus del síndrome respiratorio agudo severo (SARS-CoV), el coronavirus del síndrome respiratorio de Oriente Medio (MERS-CoV) y el más reciente coronavirus del síndrome respiratorio agudo severo 2 (SARS-CoV-2) han causado enfermedades graves en humanos, con síntomas como daño pulmonar agudo (DPA) y síndrome de dificultad respiratoria aguda (SDRA) que pueden conducir a insuficiencia pulmonar y provocar la muerte. A partir de diciembre de 2019 el mundo quedó inmerso en un estado de pandemia debido a la enfermedad denominada COVID-19 (abreviatura de *Coronavirus Disease*, 2019), ocasionada por el virus SARS-CoV-2, un coronavirus de ARN monocatenario de sentido positivo de aproximadamente 30 kilobases, que comparte una identidad de secuencia de alrededor del 80% con la del SARS-CoV. El ARN genómico de SARS-CoV-2

comprende 14 marcos de lectura abiertos (ORFs) de los cuales dos, ORF1a y ORF1b, que se superponen por un desplazamiento de marco ribosómico (-1), abarcan dos tercios del genoma y se traducen en las poliproteínas pp1a y pp1ab, respectivamente (Alanagreh *et al.*, 2020; Helmy *et al.*, 2020). Estas poliproteínas son procesadas por dos proteasas virales, la proteasa principal Mpro y la proteasa de tipo papaína PLpro para producir 16 proteínas no estructurales (Nsps), algunas de las cuales, junto con factores del hospedero, forman el complejo de replicación-transcripción (CRT), que constituye el eje central para la replicación y transcripción del genoma viral (Chen *et al.*, 2020; Romano *et al.*, 2020). El tercio restante del genoma tiene sus ORFs superpuestos y codifican para cuatro proteínas estructurales principales: la proteína Spike o de pico (S), la proteína de membrana (M), la proteína de envoltura (E) y la proteína de la nucleocápside (N), además de otras proteínas accesorias codificadas en los ORF3a, ORF6, ORF7a, ORF7b, ORF8, ORF9 y ORF10, que pueden interferir con la respuesta inmune innata del huésped (Fig. 1) (Arya *et al.*, 2021).

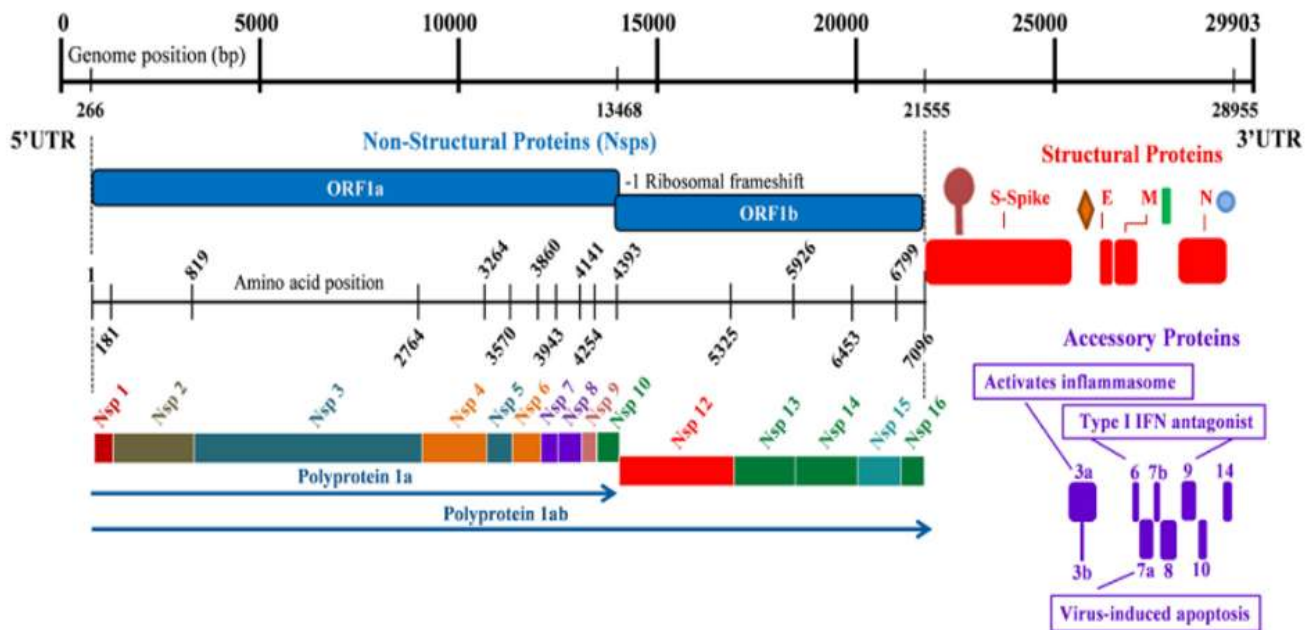


Figura 1: Organización del genoma de SARS-CoV-2. El genoma viral codifica para 16 proteínas no estructurales (Nsps) requeridas para la replicación y transcripción del genoma viral. Las Nsps están codificadas en los primeros 2/3 del genoma, mientras que el último tercio codifica para proteínas estructurales involucradas en el ensamblaje de nuevas

partículas virales, así como de otras proteínas accesorias interferentes con la respuesta inmune del huésped (Arya *et al.*, 2021).

4.2 ESTRUCTURA DE SARS-COV-2

SARS-CoV-2 es un beta-coronavirus esférico o pleomórfico de un tamaño entre 60 y 140 nm que posee cinco estructuras principales, las cuales incluyen a las proteínas Spike o de pico (S), de membrana (M), de envoltura (E), de nucleocápside (N) y dímeros de hemaglutinina-esterasa (HE), además del ARN como material genético (Fehr & Perlman, 2015; Walls *et al.*, 2020; Wu *et al.*, 2020).

La proteína S es una glicoproteína que se proyecta en la superficie periférica del virus, presentando prominencias en forma de picos que forman un aro, asemejándose a una corona bajo un microscopio electrónico. Representan las estructuras más grandes del virus y son esenciales para la entrada del mismo en las células hospederas (Robson, 2020). Para poder unirse a la membrana de la célula huésped, la proteína S es escindida por una proteasa celular de tipo furina (TMPRSS2), separando a la proteína S en dos polipéptidos distintos, S1 y S2. S1 es responsable de la unión con el receptor celular, la enzima convertidora de Angiotensina II (ACE2), mientras que S2 conduce a la fusión de las membranas viral y celular favoreciendo su entrada a la célula (Poduri *et al.*, 2020).

La proteína M es la más abundante del coronavirus y es responsable del desarrollo de una respuesta humoral específica, siendo además capaz de neutralizar los anticuerpos desarrollados en pacientes con SARS (Liu *et al.*, 2010).

Las proteínas E son glicoproteínas presentes en pequeñas cantidades que funcionan como canales iónicos, los cuales permiten la liberación del material genético viral hacia la célula hospedera (Kong *et al.*, 2020). Por tanto, aunque estas proteínas no son necesarias para la replicación viral, son fundamentales para la patogénesis mediante el ensamblaje y liberación del ARN del virus.

Las proteínas N son proteínas fosforiladas que permiten la unión del ARN viral en una conformación de tipo “perla en cuerda” o “beads-on-a-string”. Se sabe que estas proteínas ayudan en la encapsulación del material genético en las partículas del virus al unir la red del genoma viral unida a la proteína con el complejo de replicación-transcripción (CRT) (Mu *et al.*, 2020; Walls *et al.*, 2020).

La proteína HE se une a los ácidos siálicos en las glicoproteínas de la superficie celular y contiene actividad de acetil-esterasa. Estas actividades mejoran la entrada del coronavirus en las células y su propagación a través de la mucosa (Fig.2) (Zeng *et al.*, 2008; Lang *et al.*, 2020).

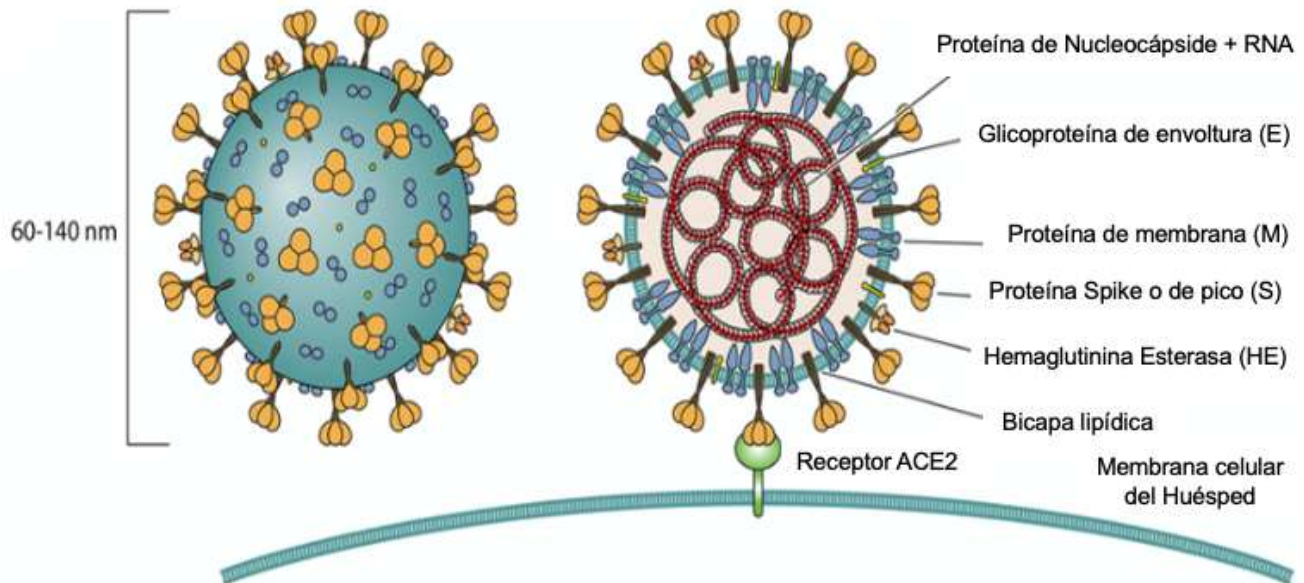


Figura 2: Estructura general de SARS-CoV-2. (Tomado y modificado de Safiabadi Tali *et al.*, 2021)

4.3 CICLO DE REPLICACIÓN DE SARS-COV-2

El ciclo de replicación de SARS-CoV-2 comienza cuando el virus inyecta su genoma en la célula hospedera. La entrada del genoma viral puede lograrse ya sea a través de endosomas o por fusión directa de la envoltura viral con la membrana de la célula hospedera. En particular, la proteína S de SARS-CoV-2 se une a la enzima convertidora de angiotensina 2 (ACE2), que es un receptor de superficie celular altamente específico para este virus y que se expresa en los pulmones, las arterias, el corazón, los riñones y el intestino, lo que permite la entrada del virus a las células del humano. Después de la interacción con ACE2, la proteína S se divide en dos subunidades, S1 y S2, por acción de una proteasa extracelular de tipo furina (TMPRSS2) asociada a la superficie de la célula hospedera. Mientras que S1 se une a ACE2, la subunidad S2 es escindida y activada por TMPRSS2, lo que resulta en una fusión de la membrana viral y la liberación del ARN genómico (ARNg) en el

citoplasma de la célula hospedera (Guo *et al.*, 2020; Hoffmann *et al.*, 2020; Xu *et al.*, 2020).

Una vez dentro del citoplasma, el ARNg es traducido por los ribosomas del huésped. Esto último da como resultado dos poliproteínas, pp1a y pp1ab que se escinden proteolíticamente en 16 proteínas no estructurales (Nsp 1-16) por acción de las proteasas virales, Mpro y PLpro.

Varias Nsps, junto con otros factores, se ensamblan para formar el complejo CRT dentro de la célula hospedera infectada. Mientras que las Nsp 2-11 desempeñan un papel de apoyo, las Nsp 12-16 proporcionan la función enzimática requerida para la replicación y transcripción del genoma viral dentro del CRT. En este proceso, la cadena de ARN (+) se replica en ARN (-) y posteriormente la cadena negativa se usa ya sea como molde para la síntesis de ARN (+) que formará parte de las nuevas partículas virales o para la transcripción de ARNm subgenómicos que serán traducidos por la maquinaria del hospedero en las proteínas estructurales: S, M, E, N y las proteínas accesorias. Las proteínas M, S y E se insertan en el retículo endoplásmico (RE) y luego se trasladan al sitio de ensamblaje viral, que es el compartimento intermedio RE-Golgi. En este sitio, la proteína N se une a la cadena de ARNg (+) para producir un complejo de nucleoproteínas, las cuales junto con las proteínas estructurales tienen un proceso de ensamblado y maduración para después brotar del Golgi en forma de pequeñas vesículas (Zhou *et al.*, 2020; Fung *et al.*, 2014). Estas vesículas viajan a la membrana de la célula hospedera donde finalmente los viriones maduros de SARS-CoV-2 se liberan a la región extracelular por exocitosis. Los viriones recién liberados pueden infectar a un nuevo conjunto de células sanas conduciendo a la progresión de la enfermedad o diseminarse al ambiente a través de la apertura de la boca, la respiración o situaciones similares promoviendo el contagio a personas sanas (Fig.3) (Ibrahim., 2020; Shereen *et al.*, 2020; Arya *et al.*, 2021).

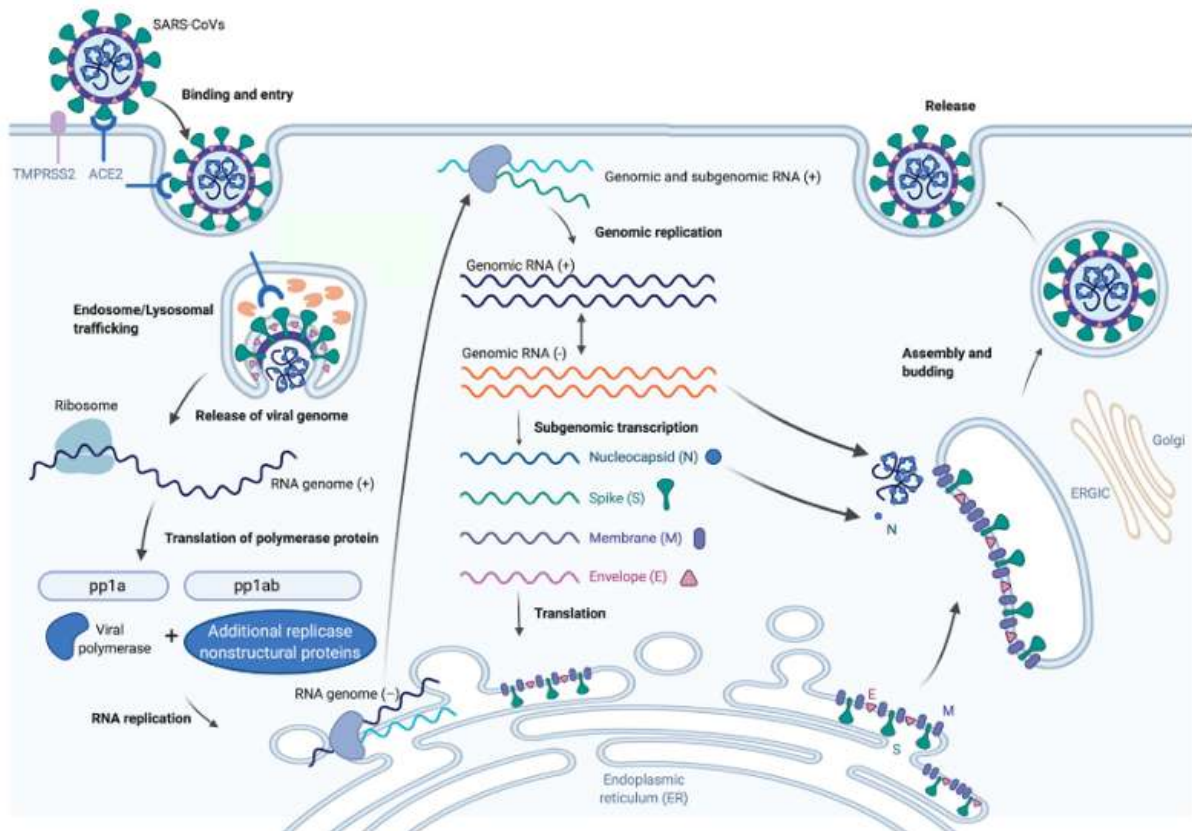


Figura 3: Ciclo de replicación de SARS-CoV-2. Tomado y modificado de Harrison *et al.*, 2020.

4.4 ESTRUCTURA Y FUNCIÓN DE LA PROTEASA PRINCIPAL (Mpro) DE SARS-COV-2

La proteasa principal (*Main protease*, Mpro, por sus siglas en inglés) de SARS-CoV-2 es una cisteína proteasa involucrada en la mayoría de los eventos de escisión en la maduración de las poliproteínas precursoras pp1a y pp1ab en al menos 11 sitios conservados, generando 12 proteínas efectoras no estructurales con un papel fundamental en la replicación y transcripción viral, por lo que la inhibición de la actividad de esta enzima bloquearía la replicación viral, convirtiéndola en una diana biológica atractiva para el diseño de fármacos antivirales (Dai *et al.*, 2020). En marzo 2020, la estructura de Mpro de SARS-CoV-2 fue purificada, cristalizada y resuelta a 1.75 Å por difracción de rayos X (Zhang *et al.*, 2020). Esta estructura, sin duda, ha facilitado la identificación de nuevos y mejores compuestos inhibidores de la actividad proteasa.

Mpro es una proteína conservada entre los coronavirus y varias características comunes se comparten entre sus sustratos. Las Mpro de SARS-CoV y SARS-CoV-2 tienen una identidad de secuencia del 96%, por lo que su estructura tridimensional es muy similar, con una desviación cuadrática media (RMSD) entre las dos estructuras de 0.53 Å para todas las posiciones de C α (PDB: 2BX4). Además, sus valores de eficiencia catalítica (k_{cat}/K_m) y la constante de disociación (K_d) de dimerización son también similares (Zhang *et al.*, 2020).

En su forma activa, Mpro es un homodímero compuesto por dos protómeros, donde cada protómero tiene tres dominios (Fig. 4). El dominio I (residuos 8-101) y el dominio II (residuos 102-184) tienen una estructura de barril β -antiparalela y albergan el sitio de unión al sustrato en una hendidura entre ellos. El dominio III (residuos 198-303) contiene cinco hélices α dispuestas en un cúmulo globular en gran parte antiparalelo, está conectado al dominio II por una región de bucle largo (residuos 185-200) y participa en la regulación de la dimerización de Mpro, a través de la interacción iónica entre Glu290 de un protómero y Arg4 del otro (Zhang *et al.*, 2020). El dímero formado por Mpro de SARS-CoV-2 tiene una interfaz de contacto de aproximadamente 1394 Å, predominantemente entre el dominio II del protómero A y los residuos de la región N-terminal del protómero B, con las dos protómeros orientados perpendicularmente entre sí. La dimerización de la enzima es necesaria para la actividad catalítica, ya que la región N-terminal de cada uno de los dos protómeros interacciona con Glu166 del otro protómero, ayudando a dar forma al bolsillo S1 del sitio de unión al sustrato (Fig. 4) (Zhang *et al.*, 2020).

A partir de la estructura de Mpro se han identificado los residuos importantes del sitio activo (His41 y Cys145), así como los 11 sitios de corte en la poliproteína (replicasa viral pp1ab de ~790 kDa). Más aún, ahora también es posible conocer la secuencia de reconocimiento específica (en este documento se indicará con una diagonal el sitio de corte en el contexto de una secuencia de aminoácidos). De todas las proteínas del SARS-CoV-2, la proteasa Mpro es la más investigada en los coronavirus y dado que está altamente conservada entre SARS-CoV-2, SARS-CoV, MERS-CoV y otros coronavirus, la investigación previa sobre esta enzima puede

servir como base para el diseño de fármacos capaces de inhibirla (Zhou et al., 2020). Actualmente, diferentes grupos trabajan en el descubrimiento y optimización de compuestos orgánicos dirigidos al sitio activo de Mpro, lo que demuestra la enorme importancia de la combinación de métodos computacionales y experimentales para acelerar el descubrimiento de nuevas drogas. Es muy importante enfatizar que no existen proteasas de humano que reconozcan la secuencia de corte Leu-Gln/Ser-Ala-Gly, lo que resulta en una probabilidad muy baja de que los compuestos que se identifiquen y optimicen causen citotoxicidad en células de humano. Por lo tanto, el compuesto identificado representaría una opción viable, efectiva y segura para inhibir la replicación de SARS-CoV-2, eliminar el virus y restaurar la homeostasis.

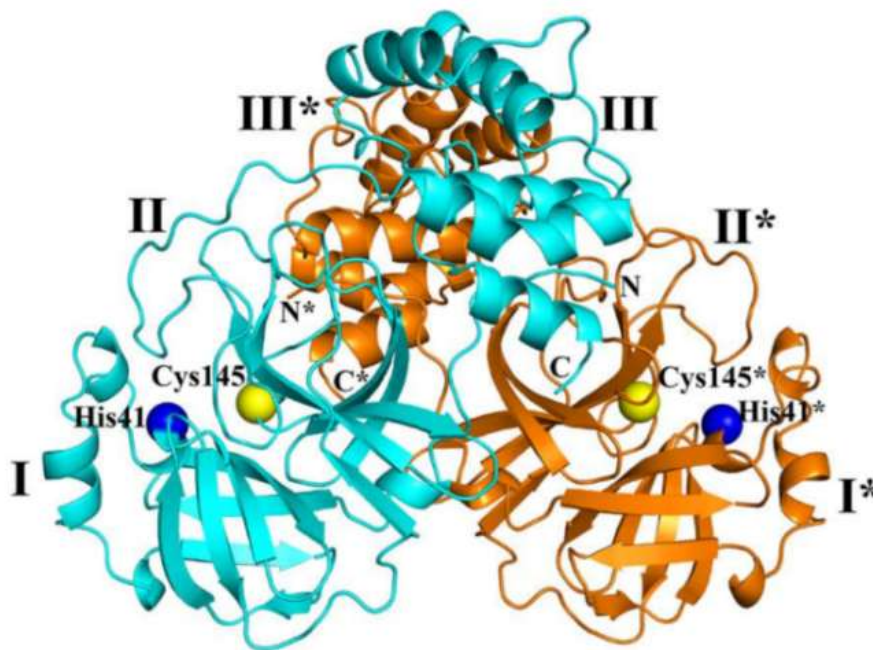


Figura 4. Estructura de la proteasa principal Mpro de SARS-CoV-2. Los residuos His41 y Cys145 son importantes en la catálisis. Los números romanos indican los dominios de cada monómero (azul y naranja). Las letras N y C indican el amino y carboxilo terminal, respectivamente (Tomado de Zhang *et al.*, 2020).

4.5 ESTRUCTURA Y FUNCIÓN DE LA PROTEASA DE TIPO PAPAÍNA (PLpro) DE SARS-COV-2

La proteasa de tipo papaína (*Papain Like-Protease*, PLpro, por sus siglas en inglés) es al igual que Mpro, una cisteína proteasa que escinde proteolíticamente a las poliproteínas virales pp1a y pp1ab. Se han identificado 3 sitios de corte, lo que genera las proteínas no estructurales Nsp1, Nsp2 y Nsp3 (Harcourt *et al.*, 2004). La secuencia de corte para la proteólisis mediada por PLpro es Leu-X-Gly-Gly/X (la diagonal indica el sitio de corte y la X representa a cualquier aminoácido). Debido a que tanto la ubiquitina (Ub) como el producto génico estimulado por interferón 15 (ISG15) tienen la secuencia Leu-X-Gly-Gly en su extremo C-terminal, PLpro tiene actividad en ambas moléculas (Barretto *et al.*, 2005; Linder *et al.*, 2007). Se sabe que modificaciones post-traduccionales de las moléculas de señalización, en forma de ubiquitinación y de activación por ISG15 activan las respuestas inmunitarias innatas y la actividad enzimática de PLpro puede interferir negativamente en dichas respuestas (Welchman *et al.*, 2005). Las PLpro de SARS-CoV y SARS-CoV-2 tienen una identidad de secuencia del 83%. La enzima consiste de un dominio similar a la ubiquitina en el extremo amino que está separado del dominio catalítico. Este dominio catalítico adopta una arquitectura de mano derecha abierta, con los subdominios pulgar, palma y dedos. El subdominio pulgar está formado por cuatro hélices α , el subdominio palma es una lámina β de seis hebras y el subdominio de dedos está formado por una lámina β de cuatro hebras, retorcida y antiparalela (Fig. 5). Un ion de zinc está coordinado por cuatro residuos de cisteína en conformación tetraédrica en la región de las yemas de los dedos; este arreglo es esencial para la integridad estructural y la actividad de la enzima (Ratia *et al.*, 2006). El sitio activo de PLpro está ubicado en la hendidura entre los subdominios pulgar y palma. Este sitio contiene una tríada catalítica de residuos de cisteína, histidina y ácido aspártico. Un bucle β flexible (Bucle BL2) del dominio de la palma, presente cerca de la entrada del sitio activo, actúa como una puerta (*gate*). Esta puerta se observa en una conformación abierta en las estructuras PLpro no ligadas. Tras la unión de un inhibidor, esta puerta sobre la entrada de la hendidura catalítica se cierra a través de un mecanismo de ajuste inducido, formando interacciones intermoleculares con

el ligando (Fig. 5). El bucle puede asumir múltiples conformaciones, dependiendo del tamaño y tipo de la molécula inhibidora (Ratia *et al.*, 2008; Báez-Santos *et al.*, 2014; Rut *et al.*, 2020). La consideración de la estructura y conformación de dicha puerta será crucial para los futuros esfuerzos de diseño de fármacos basados en la estructura contra PLpro (Arya *et al.*, 2021).

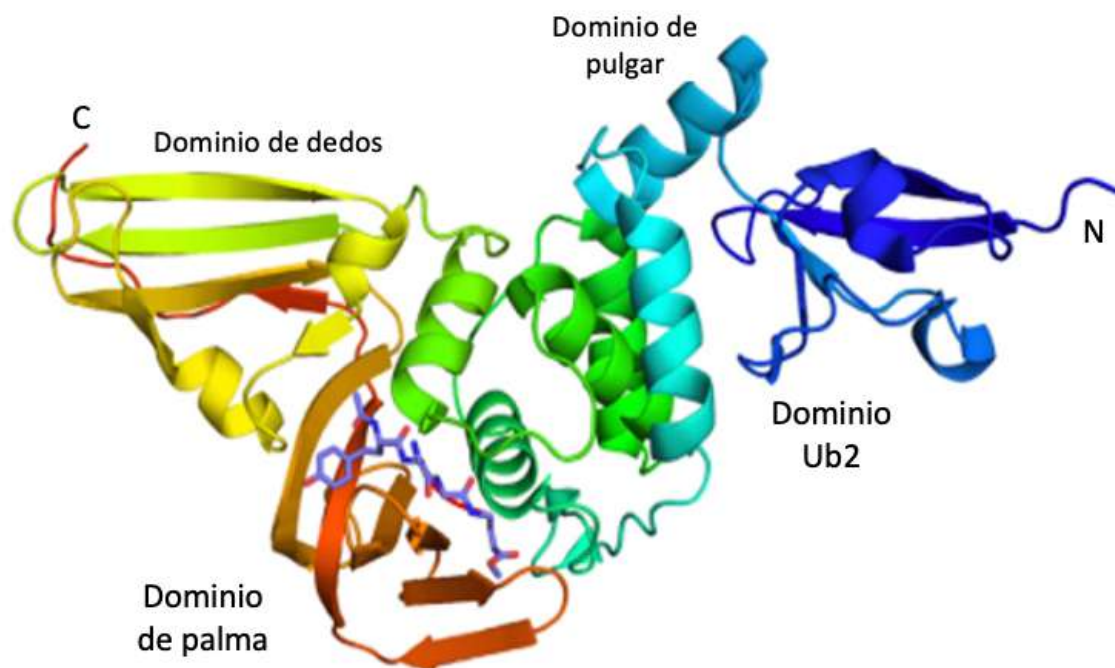


Figura 5. Estructura de la proteasa de tipo papáina PLpro de SARS-CoV-2. (PDB ID: 6WX4) mostrando diferentes subdominios (pulgares, mano, dedos y de unión a ubiquitina). El bucle β flexible del dominio palma está en conformación de puerta cerrada con un inhibidor ligado al sitio catalítico de la enzima (color morado). Las letras N y C indican el amino y carboxilo terminal, respectivamente (tomado y modificado de Arya *et al.*, 2021).

4.6 INFORMÁTICA Y SU IMPORTANCIA EN LA INVESTIGACIÓN ACTUAL

La informática, definida por el diccionario de la universidad de Cambridge como “*el estudio de la estructura, el comportamiento y las interacciones de los sistemas naturales y artificiales que almacenan, procesan y comunican información*” (Cambridge Dictionary, 2020), ha generado un cambio importante en el desarrollo de la investigación en ciencia básica y aplicada en la actualidad. Las computadoras son importantes en este rubro, siendo de importancia crítica para transformar rápidamente datos en información e información en conocimiento, validado por expertos en el área de interés (Gasteiger, 2020). Disciplinas como la biología y la

química dependen de la generación, manejo y recuperación de una gran cantidad de información obtenida de diferentes fuentes como observaciones experimentales, mediciones, simulaciones, etc., lo que vuelve casi imposible desarrollar investigación relevante en estas áreas sin el uso de computadoras y herramientas de software especializadas. La informática no solo es esencial para almacenar, procesar, analizar y difundir datos heterogéneos en biología, química y otras áreas, sino que, además, el uso de algoritmos y otras aplicaciones informáticas son esenciales para derivar modelos predictivos que a menudo complementan y guían las prácticas experimentales. Las predicciones de propiedades químicas y biológicas y las simulaciones de sistemas moleculares se han convertido en componentes integrales de la investigación básica y aplicada (López-López *et al.*, 2020).

4.7 LA QUIMIOINFORMÁTICA COMO DISCIPLINA EN EL DESCUBRIMIENTO Y DISEÑO DE FÁRMACOS

La Quimioinformática es una disciplina independiente que tiene una amplia gama de aplicaciones en química (Varnek & Baskin, 2011). Existen en la literatura en inglés varios términos publicados que hacen referencia a la misma disciplina ("Chemical informatics", "Cheminformatics", "Chemoinformatics", "Chem informatics") y cuyo uso depende mayormente de la región geográfica (Miranda-salas *et al.*, 2023). No obstante, para el presente texto se empleará el término Quimioinformática para referirse al enfoque computacional utilizado para solucionar problemas relacionados con la elucidación y representación de la estructura, el diseño y síntesis de compuestos químicos, así como la recopilación, análisis y gestión de datos químicos con un enfoque significativo en el descubrimiento de fármacos (López-López *et al.*, 2020). Brown y Bristol introdujeron en 1998 el término Quimioinformática, definido como el uso de la tecnología y la gestión de la información para transformar los datos en información y la información en conocimiento con el fin último de tomar mejores decisiones rápidamente en el área de la identificación y organización de compuestos líderes en el descubrimiento de fármacos (Brow & Bristol, 1998). La Quimioinformática se relaciona con disciplinas

como la bioinformática y la química teórica y computacional, donde comparten similitudes como el uso de una gran cantidad de datos y el estudio de ciertos fenómenos como el acoplamiento molecular de ligando-proteína, la química cuántica y el estudio de campos de fuerza respectivamente. No obstante, tiene diferencias importantes respecto a estas áreas, como la forma de sus modelos moleculares, conceptos básicos, mecanismos de inferencia y dominios de aplicación (López-López *et al.*, 2020; Varnek y Basil, 2011). Mientras que los modelos moleculares usados en química cuántica (conjuntos de núcleos y electrones), en modelado molecular basado en campos de fuerza (conjuntos de átomos y enlaces “clásicos”) y en bioinformática (estructura y función de biomoléculas, como proteínas y ácidos nucleicos), la Quimioinformática trata las moléculas como grafos moleculares o vectores relacionados con descriptores asociados a características particulares como propiedades fisicoquímicas, actividad biológica, geometría tridimensional, etc (Varnek y Basil, 2011). La Quimioinformática tiene una amplia gama de aplicaciones en diferentes áreas de investigación, incluido el diseño de fármacos, la Quimiogenómica y la farmacología de sistemas, así como la química médica, analítica, orgánica, cosmética, alimentaria e inorgánica y la investigación de productos naturales, entre otros (Kirchmair, 2020; Medina-Franco & Saldívar-González, 2020).

El objetivo común de todas las disciplinas relacionadas con la informática es convertir la información en conocimiento. Muchos problemas del mundo real requieren enfoques multidisciplinarios y la colaboración de dos o más campos de investigación. Los proyectos de descubrimiento y desarrollo de fármacos son un ejemplo destacado de esta colaboración entre principalmente la bioinformática y la quimioinformática (Santibáñez-Morán *et al.*, 2020). El modelado de interacciones proteína-ligando requiere predecir la disposición tridimensional del objetivo molecular, por lo general una proteína (una tarea de interés primario para la bioinformática) y el almacenamiento y procesamiento de moléculas pequeñas con potencial farmacéutico (una tarea de interés primordial para la quimioinformática). En este rubro la quimioinformática se enfrenta al desafío de cantidades excesivamente grandes de moléculas pequeñas con potencial actividad

farmacéutica, estudiando las relaciones de estructura-actividad asociadas (Schuurman y Leszczynski, 2008). La elaboración de quimiotecas, el modelado por homología de dianas biológicas, el cribado virtual basado en la estructura, los ensayos de dinámica molecular, entre otros métodos, son ejemplos comunes de técnicas de diseño de fármacos asistido por computadora, que a menudo requieren la combinación de enfoques bioinformáticos y quimioinformáticos. Se deduce entonces que existe un potencial grande para explorar más y explotar la superposición entre bioinformática y quimioinformática (López-López *et al.*, 2020).

4.8 EL CONCEPTO DE ESPACIO QUÍMICO

“El espacio químico puede verse como análogo al universo cosmológico en su inmensidad, con compuestos químicos que pueblan el espacio en lugar de estrellas” (Zhao *et al.*, 2008). Esta analogía permite concientizar al investigador acerca de la compleja tarea que implica el descubrimiento o desarrollo de un nuevo fármaco. En este rubro el espacio químico ha proporcionado un marco conceptual sólido para guiar el análisis de la diversidad, la clasificación de estructuras, el diseño de bibliotecas, la selección de compuestos y la evaluación de las relaciones estructura-propiedad y estructura-actividad (Medina-Franco *et al.*, 2008). Arús-Pous y colaboradores definen espacio químico como *“un concepto para organizar la diversidad molecular, postulando que diferentes moléculas ocupan diferentes regiones de un espacio matemático donde la posición de cada molécula está definida por sus propiedades”* (Arús-Pous *et al.*, 2019). El espacio químico es entonces en sí una noción subjetiva y general que depende principalmente de la elección del número y tipo de descriptores que definen el espacio M-dimensional (Medina-Franco *et al.*, 2021). Estos descriptores pueden ser cualquier conjunto de números que definan el espacio de una forma ordenada, lógica y racional, además de que pueden ser adecuados para definir el espacio deseado y aplicar el concepto para una variedad de aplicaciones, dependiendo de los objetivos del proyecto. El tipo de descriptores es distintivo de las diferentes disciplinas informáticas, de tal manera que contribuyen a dar forma a disciplinas como la bioinformática, la quimioinformática, la informática biomédica, entre otras (López-López *et al.*, 2021)

Los tipos comunes de descriptores que se han utilizado para definir el espacio químico de moléculas orgánicas pequeñas incluyen propiedades moleculares que están asociadas a que una molécula sea “*de tipo fármaco*” o “*drug-like*” en inglés, como lo son propiedades ADME (absorción, distribución, metabolismo y excreción), toxicidad y otras características farmacéuticamente relevantes (Medina-Franco *et al.*, 2021).

Hablar del espacio químico en su totalidad, es referirse a un número de compuestos que excede el número de partículas elementales en el universo cosmológico (Dobson, 2004). Generar computacionalmente las estructuras de estos compuestos es una tarea actualmente imposible y pensar en sintetizarlas puede ser incluso inimaginable. De acuerdo con Varnek y Baskin (2011), el objetivo de la quimioinformática es encontrar una forma racional de representar este espacio químico literalmente infinito y obtener información que pueda convertirse en conocimiento. La capacidad de navegar por este espacio químico es extremadamente importante ya que podría resolver preguntas sobre el origen, evolución, funcionamiento e incluso destino del universo.

Si se piensa ahora en la fracción del espacio químico en donde se llevan a cabo los procesos biológicos, es decir, el espacio químico biológicamente relevante, se cae en la cuenta de que el tamaño de esta fracción es muy pequeño. Esto puede explicarse debido a que los compuestos biológicamente activos no se distribuyen uniformemente por todo el espacio químico, sino que forman regiones muy compactas en él, así como las galaxias del universo cosmológico (Zhao *et al.*, 2008). El organismo vivo más simple posee un genoma que codifica para al menos 1,000 proteínas diferentes, mientras que el genoma del humano codifica para al menos 100,000, es decir, 100 veces más. Aun así, estos números son tan solo una pequeña fracción del número teórico de proteínas existentes posibles (Lander, 2001), y estas macromoléculas se forman únicamente por la combinación de 20 aminoácidos y el tamaño promedio de una proteína natural es de aproximadamente 300 residuos, por lo que si tan solo se produjera una sola molécula de cada proteína

posible, la masa molecular combinada de estos compuestos sería mayor a la del universo (Dobson, 2004). Es entonces preciso decir que el grupo de proteínas biológicamente activas es un grupo extremadamente selecto al ser moléculas con un alto nivel de complejidad. Si se piensa ahora en el enorme número de moléculas pequeñas con potencial farmacéutico, las expectativas son fascinantes (Dobson, 2004). Aquí radica la importancia de generar estrategias eficientes para la navegación y exploración del espacio químico, ya que el desarrollo de nuevos compuestos biológicamente activos y el diseño de nuevos fármacos para la medicina depende en gran medida de ello (Zhao *et al.*, 2008).

4.9 PROPIEDADES FISICOQUÍMICAS DE INTERÉS FARMACEÚTICO Y SU PAPEL COMO DESCRIPTORES MOLECULARES PARA LA SELECCIÓN DE NUEVOS COMPUESTOS

Con respecto al descubrimiento de fármacos, se ha reportado que los fármacos y/o candidatos a fármacos tienden a tener distribuciones similares de propiedades fisicoquímicas (Tian *et al.*, 2015). Por lo tanto, el entendimiento de estas propiedades se ha convertido en una parte integral del descubrimiento de *hits* computacionales y ha permitido la selección precisa de compuestos *hits* que funcionan como puntos de partida adecuados para la identificación de nuevos candidatos clínicos que posean una mayor probabilidad de éxito farmacocinético y seguridad general (Agoni *et al.*, 2020). La evaluación de las propiedades fisicoquímicas de los compuestos para determinar su similitud con fármacos aprobados ocurre en las primeras etapas del proceso de descubrimiento. Para este fin, se han utilizado varios descriptores moleculares que evalúan una combinación de estas propiedades fisicoquímicas complejas y permiten la optimización e identificación de moléculas, teniendo en cuenta su seguridad, selectividad y eficacia (Li *et al.*, 2009). Algunas de estas propiedades fisicoquímicas son el peso molecular (MW), el coeficiente de partición octanol/agua (LogP), átomos donadores de hidrógeno (HBDon), átomos aceptores de hidrógeno (HBAcc), área superficial polar topológica (TPSA) y el número de enlaces rotables (nRotB), todos abreviados por sus siglas en inglés. El MW se define como la suma de las masas atómicas de todos

los átomos que componen a una molécula de un compuesto específico. Se sabe que el tamaño y la forma de un fármaco pueden determinar su orientación y la velocidad de movimiento a través de las membranas biológicas (Agatonovic-Krustrin *et al.*, 2002). En 1997 Lipinski y colaboradores encontraron que apenas el 11% de compuestos que pasaban a una fase clínica II tenían un MW > 500 y el porcentaje es de 8 % para compuestos con un MW > 600. Esto refleja el hecho de que los compuestos con un MW más alto tienen menos probabilidades de ser activos por vía oral que los de MW más bajo (Lipinski *et al.*, 1997). Además, después de la administración oral, el fármaco debe ser capaz de moverse a través de las membranas celulares (absorción transcelular) o entre espacios estrechos entre las células de la mucosa gastrointestinal (absorción paracelular). Tanto la difusión transcelular como la paracelular dependen de algunas propiedades fisicoquímicas dentro de las cuales se encuentra el peso molecular, la polaridad y la lipofiliidad. Aunque para la mayoría de los fármacos el mecanismo predominante de absorción es por vía transcelular, la vía paracelular es sumamente importante, de manera que compuestos con un valor de peso molecular alto no tendrán una buena absorción en el epitelio intestinal ya que su tamaño limita el paso a través de los espacios paracelulares. La ruta paracelular, por lo tanto, restringe el paso a moléculas predominantemente pequeñas, lo que corresponde a la regla de cinco que indica una buena absorción para compuestos con un MW < 500 (Van den Waterbeemd, 2000; Lipinski *et al.*, 1997). Una segunda propiedad fisicoquímica importante y la mejor conocida para predecir si un compuesto será capaz de cruzar las membranas biológicas es la determinación de la lipofiliidad, usando el logaritmo del coeficiente de partición entre octanol y agua ($\log P$) (Van den Waterbeemd *et al.*, 2001). El coeficiente de partición se define como la relación entre la solubilidad de un fármaco ionizado en un solvente orgánico y la solubilidad del mismo fármaco ionizado en un medio acuoso. Se han utilizado una variedad de disolventes orgánicos para medir los coeficientes de partición; sin embargo, se considera que el n-octanol es el estándar. El coeficiente de partición, designado como P, es un término adimensional, es decir, sin unidades y, a menudo, se expresa como un valor logarítmico, $\log P$ (Harrold y Zavod, 2018). El $\log P$ de cualquier molécula es el

resultado de las contribuciones aditivas y las interrelaciones de todos sus grupos funcionales, por lo que cada grupo funcional ayuda a determinar la naturaleza hidrofóbica o hidrofílica global de la molécula (Harrold y Zavod, 2018). Con respecto a la polaridad de la molécula, se encuentra la propiedad de área superficial polar (PSA), un descriptor de superficie molecular ampliamente utilizado en el estudio del transporte de fármacos, como la absorción intestinal (Li *et al.*, 2005) y la penetración de la barrera hematoencefálica (BBB) (Strazielle y Ghersi-Egea, 2005). El PSA se define como la suma de las contribuciones al área de superficie molecular de los átomos polares como el oxígeno, el nitrógeno y sus respectivos hidrógenos (Prasanna y Doerksen, 2008). Como tal, el PSA está relacionado con la capacidad de un fármaco para formar enlaces de hidrógeno y, por lo tanto, es un descriptor de enlaces de hidrógeno. La capacidad de un átomo para donar (HBDon) o aceptar (HBAcc) átomos de hidrógeno está relacionada con la polaridad de la molécula y con esto su capacidad para formar puentes de hidrógeno con otras moléculas, funcionando como un puente entre dos átomos electronegativos. De esta manera, los grupos funcionales que forman puentes de hidrógeno son capaces de formar este tipo de enlaces con moléculas de agua y aumentar la solubilidad en agua (Harrold y Zavod, 2018). Las moléculas con muchos átomos donadores de hidrógeno y un PSA grande tienden a tener valores de permeabilidad bajos (Clark, 1999). El cálculo del PSA puede ser complicado por la necesidad de decidir y luego calcular con precisión la geometría molecular 3D adecuada de la molécula en estudio (Clark, 1999). El área superficial polar topológica (TPSA), es una variante del PSA que hace uso de las contribuciones de grupos funcionales de una gran base de datos de estructuras. El TPSA es una medida conveniente del PSA que evita la necesidad de calcular la estructura 3D del ligando o decidir cuál es la conformación biológica relevante (Prasanna & Doerksen, 2008). Por último, el número de enlaces rotables (nRotB) es un parámetro que ayuda a evaluar la rigidez de la molécula. Un enlace rotable es cualquier enlace que no forme parte de ningún anillo, que no esté unido a un átomo terminal o que no esté unido a un átomo de hidrógeno. Los enlaces de amida C-N no se cuentan debido a su alta barrera de rotación. La rigidez de una molécula no solo ayuda al acoplamiento con su receptor

sino que también ayuda a la biodisponibilidad del compuesto. Veber y colaboradores investigaron un conjunto de datos de 1100 compuestos con datos de biodisponibilidad oral en ratas y descubrieron que aquellos con $n\text{RotB} < 10$ y un $\text{PSA} < 140 \text{ \AA}^2$ tenían una mayor probabilidad de lograr una biodisponibilidad oral superior al 20% (Veber *et al.*, 2002). El estudio de las propiedades fisicoquímicas de los conjuntos de datos de moléculas es una guía invaluable para la selección de aquellos compuestos con mayor probabilidad de éxito, por lo que su revisión y evaluación es sumamente importante.

4.10 LA IMPORTANCIA DEL CURADO DE BASES DE DATOS DE COMPUESTOS

Cuidar la calidad de los datos, tanto químicos como biológicos, provenientes de bases de datos públicas es de gran importancia para cualquier ensayo en general. No obstante, en quimioinformática, utilizar estructuras incorrectas en los análisis planteados puede traer repercusiones significativas en los resultados de las pruebas y su interpretación (Williams *et al.*, 2012; Fourches *et al.*, 2010). La calidad de los datos varía para diferentes bases de datos, dependiendo del rigor en el proceso de curado. No sería razonable esperar el mismo grado de curación en una base de datos con diez millones de compuestos y una que cuenta con apenas cientos o unos pocos miles de ellos. Además, incluso disponiendo de los recursos y tiempo para hacerlo, existen varios factores que hacen que un mismo compuesto parezca tener diferentes estructuras en diferentes bases de datos (Hersey *et al.*, 2015). En publicaciones científicas, las estructuras de los compuestos a menudo se dibujan en una forma que tiene relevancia para el contexto del artículo. Por ejemplo, en un resultado de acoplamiento molecular, una molécula de naturaleza ácida o básica podría dibujarse como una molécula cargada ya sea negativa o positivamente, ya que esta es la forma relevante para su unión con la proteína (Hersey *et al.*, 2015). Actualmente, la mayoría de las estructuras químicas en la literatura aparecen como imágenes y no en un formato de estructura legible por los equipos de cómputo universales, de manera que el proceso de extraer las estructuras y cargarlas en una base de datos implica volver a dibujar la estructura a partir de la imagen del artículo,

utilizar un software convertidor de imágenes a un formato específico o descargarla de alguna otra fuente pública o privada en un formato legible, con un posible cambio respecto a la estructura reportada. Este proceso inevitablemente introduce errores de estructura que alteran la calidad de los datos e interfieren con los resultados de los análisis (Hersey *et al.*, 2015). Revisar cada una de las estructuras en una base de datos de manera manual es una tarea agotadora y puede conducir a errores en la estandarización, aun cuando la base de datos no sea tan grande. Para facilitar el proceso, en la actualidad existen programas capaces de evaluar los compuestos en busca de errores estructurales o características no favorables en el contexto farmacológico, incluso algunos proveedores de bases de datos cuentan con su propio software de estandarización para su uso de manera libre. Las reglas para la validación y estandarización dependen de los objetivos del usuario y algunas de las más comunes incluyen la verificación de presencia de átomos estándar, valencia usual, conversión de grupos funcionales como los grupos nitro en representaciones coherentes, eliminación de estructuras múltiples, entre otros. Una vez localizadas las fallas, la estandarización consiste en corregir las estructuras en sus respectivas observaciones, convertirlas en una forma tautomérica canónica, estandarizar la representación de las estructuras aromáticas, así como la estereoquímica de los compuestos, entre otros. El curado de la base de datos es un paso crucial en los análisis informáticos y aunque el uso de estas herramientas facilitan el proceso, queda en manos del usuario hacer su propia evaluación en el contexto de sus propios intereses de investigación (Hersey *et al.*, 2015).

4.11 EL CONCEPTO DE “TIPO-FÁRMACO” Y SU IMPORTANCIA EN EL DESARROLLO DE REGLAS PARA EL FILTRADO DE BASES DE DATOS

El cálculo de propiedades fisicoquímicas de las moléculas se ha utilizado a menudo para evaluar la similitud entre fármacos aprobados y aquellos que se proponen como candidatos computacionales (Lipinski *et al.*, 2001; Veber *et al.*, 2002). Con la finalidad de filtrar aquellas moléculas con propiedades no deseables de las bases de datos y reducir el riesgo de desgaste en etapas tardías del descubrimiento de fármacos, los químicos medicinales y computacionales han propuesto el concepto de “*Drug-likeness*” que se puede traducir como “*con similitud a fármacos*” o “*de tipo*”

fármaco” (Lipinski *et al.*, 1997; Sadowski, 2000). *Drug-likeness* se puede definir como un equilibrio complejo de varias propiedades moleculares y características estructurales que determinan si una molécula en particular es un buen candidato para convertirse en un fármaco. Estas propiedades principalmente engloban la lipofilidad, distribución electrónica, características enlazantes de átomos de hidrógeno, tamaño y flexibilidad de las moléculas y algunas otras características farmacofóricas que influyen en el comportamiento de una molécula en un organismo vivo, por ejemplo en el transporte, afinidad por las proteínas, reactividad, toxicidad, actividad metabólica, estabilidad, entre otros (Turner, 2007). Actualmente se continúa estudiando una gran cantidad de moléculas, tanto similares como no similares, a fármacos con la finalidad de encontrar motivos estructurales que se asocien con características de “*Drug-likeness*” (Turner, 2007). Un enfoque común para medir las características de *Drug-likeness* de un conjunto de moléculas es el seguimiento de filtros o reglas que definen los límites aceptables de ciertas propiedades fisicoquímicas y estructurales para fármacos y candidatos computacionales. El filtro o conjunto de reglas más famoso es el llamado “Regla de los 5 (Ro5)” propuesta por Lipinski y colaboradores en 1997, que define los límites de cuatro propiedades fisicoquímicas moleculares para que un compuesto sea activo al suministrarse por vía oral (Lipinski *et al.*, 1997). De acuerdo con las reglas de Lipinski, la mala absorción y/o permeabilidad de un fármaco es más probable cuando su estructura química cumple con dos o más de los siguientes criterios: MW mayor a 500 g/mol, el valor de logP calculado es mayor a 5, existen más de 5 HBDon (-NH-, -OH) y el número de HBAcc es mayor a 10 (-N=, -O-) (Lipinski *et al.*, 1997). Según las reglas de Lipinski, es menos probable que las moléculas químicas que violan estas reglas posean muchas características deseables para convertirse en fármacos; no obstante, es importante señalar que la regla de 5 no estima la absorción oral de manera cuantitativa, ni tampoco compuestos que no rompen ninguno de los criterios de Ro5 son necesariamente biodisponibles por vía oral. Es decir, la Ro5 no categoriza de manera definitiva los compuestos con buena o mala absorción, aunque es simple, rápida y proporciona un grado razonable de clasificación (Turner, 2007).

Con el tiempo, han ido apareciendo otras reglas o filtros basados en propiedades moleculares que se han propuesto para comprender la composición estructural de fármacos y moléculas similares y evaluar las características de “tipo fármaco” de diferentes conjuntos de moléculas (Tian *et al.*, 2015). Un ejemplo son las reglas de Veber, donde el número de nRotB y la suma de donadores y aceptores de hidrógeno así como el PSA juegan un papel predominante por encima del MW en la selección de compuestos para administración oral. Las reglas de Veber sugieren que un compuesto que cumpla con ya sea poseer una $PSA \leq 140 \text{ \AA}^2$ y un número de nRotB ≤ 10 , o bien, que la suma de sus donadores y aceptores de hidrógeno sea ≤ 12 y el número de nRotB ≤ 10 , es un compuesto que cumple con criterios que lo hacen candidato a administrarse de manera oral (Veber *et al.*, 2002).

Estos y otros conjuntos de reglas para la selección y el diseño de potenciales fármacos pueden proporcionar pautas útiles que ayuden a guiar el proceso de descubrimiento y a evitar pérdidas en etapas tardías de la elaboración de un fármaco; sin embargo, las reglas no son absolutas y existen una gran cantidad de compuestos que incumplen con algunas y han llegado a ser fármacos aprobados. Algunos ejemplos de fármacos con peso molecular elevado y con una biodisponibilidad oral inesperadamente alta incluyen a la faloidina (toxina de hongos), antamanida (antitoxina), ciclosporina (inmunosupresor) y vinorelbina (agente antitumoral) (Veber *et al.*, 2002). De ello se deduce, que existe una clara necesidad de modelos cuantitativos capaces de examinar grandes bases de datos químicas, para identificar nuevas entidades químicas con buena solubilidad y absorción oral que representarían opciones interesantes en el descubrimiento y desarrollo de futuros fármacos (Turner, 2007).

4.12 MODELADO MOLECULAR EN EL DESCUBRIMIENTO Y DESARROLLO DE FÁRMACOS

Actualmente el uso de herramientas computacionales en el descubrimiento y desarrollo de fármacos tiene un gran poder predictivo que ayuda a hacer el proceso más eficiente, seguro y económico. El modelado molecular se usa para simular el

comportamiento de sistemas químicos y biológicos utilizando diversos algoritmos computacionales. Es una de las técnicas que ocupan un lugar de liderazgo en el mundo y tiene una amplia aplicabilidad en diversos temas como son el descubrimiento y diseño de drogas, el diseño y mejora de biomateriales y la espectroscopía. El modelado molecular podría definirse como un tipo de trabajo computarizado que aplica las leyes de la física, con base en resultados experimentales, para el análisis de moléculas. Este trabajo incluye el análisis del tipo y número de átomos, la naturaleza de los enlaces, la longitud y ángulos de enlace, los ángulos diedros, la energía molecular, la optimización geométrica, los cálculos de entalpía de interacción y la frecuencia vibracional de sistemas moleculares. Los métodos de modelado pueden también describir con notable precisión las características nucleofílicas o electrofílicas de las moléculas, sus potenciales electrostáticos y, muy importante, predecir propiedades estructurales y biológicas que conduzcan a un entendimiento de las relaciones entre la estructura y actividad para el diseño racional de drogas de consumo humano (Fig.6).

En 2011, Sutch y colaboradores indicaron que la tasa promedio de fracaso en ensayos clínicos de nuevos medicamentos era del 81% (DiMasi *et al.*, 2010). El 30% de estos fracasos son debido a la falta de eficacia, mientras que los riesgos en la seguridad clínica y toxicológica representan otro 30% (Kola & Landis, 2004). El fracaso en las últimas etapas del desarrollo de un fármaco es muy caro, sobre todo si dicho fármaco alcanza los ensayos clínicos de fase II o III antes de que ocurran los problemas. Estos problemas enfatizan la importancia de la identificación de moléculas que tengan el potencial para convertirse en fármacos y que lleguen al mercado en la etapa inicial de su descubrimiento. En 1991, la industria farmacéutica observó una tasa de fracaso del 40%, debido a problemas de biodisponibilidad y farmacocinética, pero la incorporación de los principios de ADMET (absorción, distribución, metabolismo, excreción y toxicidad) en la etapa temprana del proceso de desarrollo de fármacos ha reducido el porcentaje al 10% (Kola & Landis, 2004; Tsaioun *et al.*, 2009). Un correcto uso de diferentes técnicas computacionales puede

eficientizar de manera importante el proceso para la obtención de un nuevo fármaco, especialmente en las etapas de descubrimiento y optimización del principio activo.

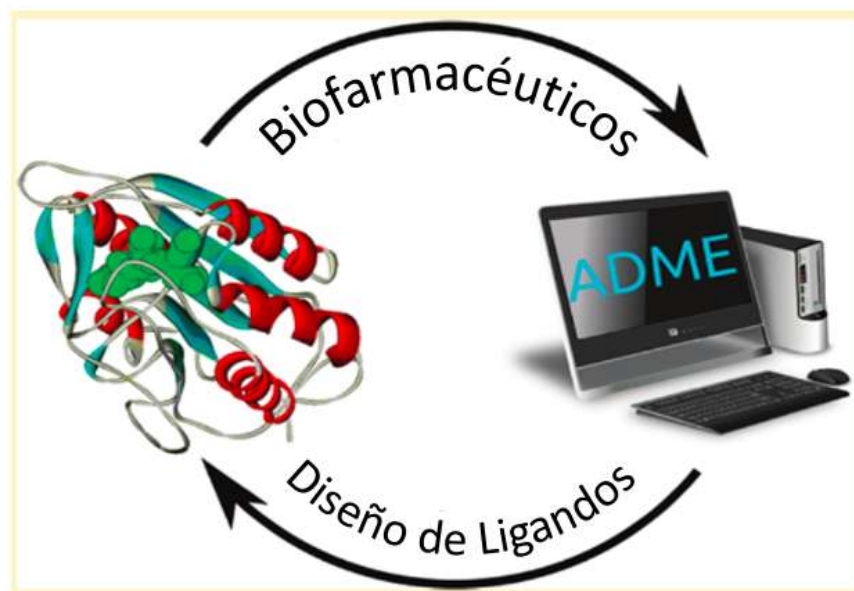


Figura 6. Esquema representativo de la interrelación entre los sistemas computacionales y su papel en el diseño de moléculas activas. desarrollados para descubrir, diseñar y optimizar compuestos orgánicos y el potencial que representa en la industria farmacéutica la conversión de estos compuestos orgánicos en fármacos de consumo humano (adaptado de Sutch *et al.*, 2012).

4.13 SIMULACIONES DE ACOPLAMIENTO MOLECULAR Y SU PODER PREDICTIVO

El acoplamiento molecular es un método de diseño de fármacos basado en la estructura que simula la interacción molecular y predice el modo de unión y la afinidad entre receptores y ligandos. Esta tecnología se ha utilizado ampliamente en el campo de la investigación del diseño de fármacos, reduciendo los costos de investigación y mejorando en gran medida la eficiencia en el descubrimiento de moléculas con potencial farmacéutico. El acoplamiento molecular consiste en simular la conformación óptima según la complementariedad y la preorganización de un compuesto en la cavidad de alguna molécula más grande. En este contexto, el cribado virtual es una técnica computacional utilizada ampliamente para predecir la unión de un gran número de ligandos extraídos de una base de datos (Quimiotecas) a un receptor biológico, con el objetivo en particular de identificar los

compuestos más prometedores (*hits*) para su posterior estudio. Dado el número cada vez mayor de estructuras de proteínas y ácidos nucleicos conocidas, el cribado virtual sigue creciendo como método eficaz para el descubrimiento de nuevos inhibidores y fármacos (Cosconati *et al.*, 2010).

Estas técnicas se enfrentan a varios desafíos, como el muestreo de las diversas conformaciones de moléculas flexibles y el cálculo de energías de unión absolutas en un entorno acuoso. No obstante, el campo ha tenido un importante éxito en la predicción con gran exactitud de nuevos ligandos junto con sus estructuras unidas al receptor, por lo que ofrece una ruta práctica para descubrir nuevos compuestos reactivos e información importante para la investigación farmacéutica (Shoichet, 2004).

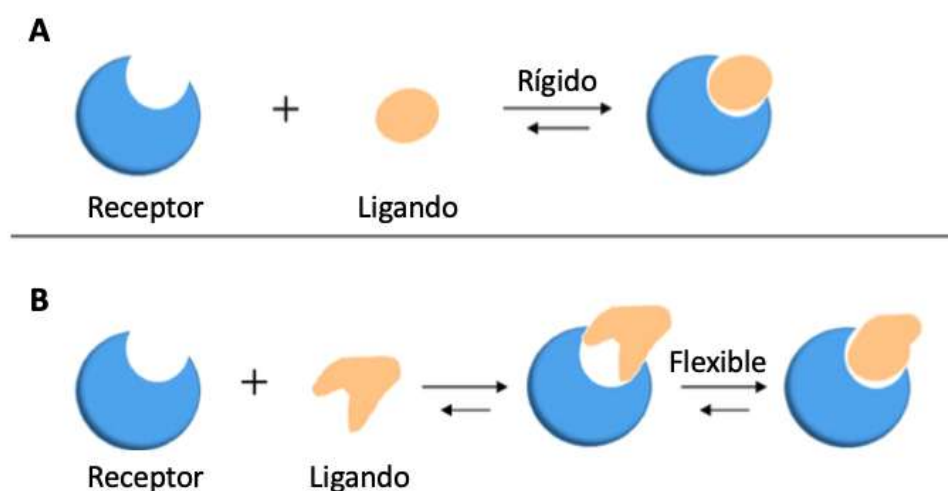


Figura 7. Dos modelos de acoplamiento molecular. A) modelo de candado llave. B) modelo de ajuste inducido (adaptado de Fan *et al.*, 2019).

4.14 MODELO DE FARMACÓFORO

Un modelo de farmacóforo se define como un arreglo tridimensional de las características mínimas necesarias estéricas y electrónicas para asegurar interacciones óptimas con un blanco farmacológico específico, lo cual desencadenará o bloqueará una respuesta biológica (Saldívar-González *et al.*, 2016). El modelo de farmacóforo es una técnica de diseño de fármacos asistido por

computadora “DIFAC”, que se emplea comúnmente para consultar en quimiotecas las moléculas de interés con las propiedades deseadas y extraerlas para su posterior utilización y estudio (Voet *et al.*, 2013).

El modelo de farmacóforo se basa en la teoría de que moléculas con funcionalidad química y disposición espacial similar, tendrán actividad biológica sobre la misma diana farmacológica. Las características en un modelo de farmacóforo se representan como entidades geométricas tales como esferas, planos y vectores, e indican las propiedades químicas dentro de una molécula que la hacen capaz de crear interacciones con su blanco biológico. Las características más importantes para la generación de un modelo de farmacóforo son: HBAcc, HBDon, áreas hidrofóbicas (H), grupos positivos y negativos ionizables (PI/NI), anillos aromáticos (AR) y áreas de coordinación de metales (Fig. 8). Pueden añadirse restricciones de tamaño adicionales en forma de volúmenes de forma o exclusión (XVOL) para representar el tamaño y la forma del bolsillo de unión (Vuorinen & Schuster, 2015).

Los modelos de farmacóforo pueden generarse utilizando dos enfoques diferentes. El primero es el modelo de farmacóforo “*basado en la estructura*” donde se utiliza la información estructural de la proteína diana, ya sea enzimas o receptores, para identificar compuestos que puedan utilizarse como fármacos. El segundo enfoque es el modelo de farmacóforo “*basado en el ligando*”, el cual consiste en el desarrollo de un modelo de farmacóforo comúnmente en 3D, utilizando las propiedades fisicoquímicas de moléculas que se conoce que tienen actividad biológica y una orientación similar con respecto a la misma diana biológica. Esto permite la extracción de características químicas en común y, por tanto, la construcción del farmacóforo (Giordano *et al.*, 2022).

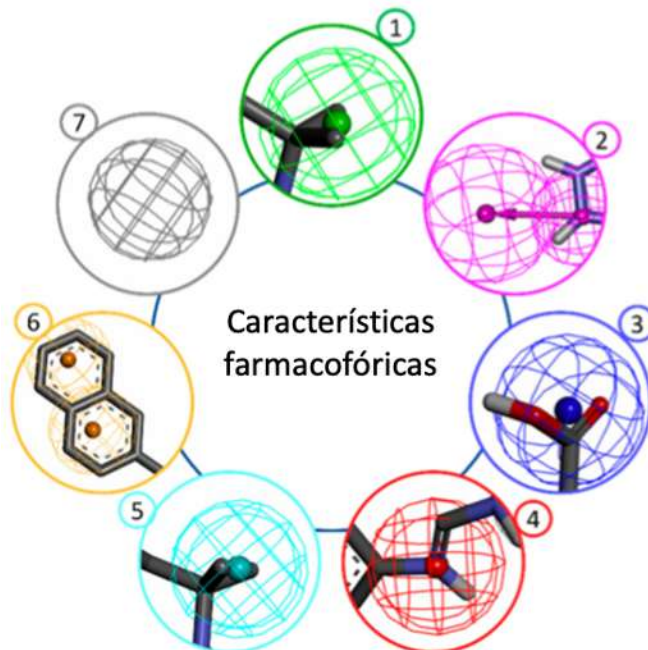


Figura 8. Características farmacofóricas principales. 1.- Aceptor de puente de Hidrógeno (HBAcc), 2.- Donador de puente de Hidrógeno (HBDon), 3.- Grupo negativo ionizable (NI), 4.- Grupo positivo ionizable (PI), 5.- Grupo hidrofóbico (H), 6.- Grupo aromático (AR), 7.- Volumen de exclusión (XVOL).

4.15 ENSAYOS DE SIMULACIÓN ATOMÍSTICA DE DINÁMICA MOLECULAR

En su artículo titulado “*Molecular Dynamics for all (2018)*” Hollingsworth y Dror plantean una analogía donde se pretende explicarle a un alienígena lo que es una bicicleta, cómo funciona y cómo repararla si esta se avería. Si se le mostrase al alienígena una imagen de la bicicleta, sería para él complicado entender el funcionamiento de la misma; no obstante, si se le mostrase una película de alguien montando una bicicleta o aún mejor, un modelo virtual donde el alienígena pudiera experimentar mover los componentes como por ejemplo un pedal y observar como las ruedas giran, así como otros componentes, sería para él de gran ayuda para comprender el concepto. En biología, entender el funcionamiento de una proteína o alguna otra biomolécula representa un desafío similar, por lo que estructuras a nivel atómico son muy útiles y generan información importante sobre el funcionamiento de la biomolécula en cuestión (Hollingsworth & Dror, 2018).

Las simulaciones atomísticas de dinámica molecular (DM) predicen cómo se moverá cada átomo de una proteína u otro sistema molecular a través del tiempo, dadas las condiciones en las que se estudia el sistema, basándose en un modelo físico que gobierna las interacciones interatómicas (Karplus y McCammon, 2002). Estas simulaciones pueden capturar una amplia variedad de procesos como el cambio conformacional, la unión de ligandos y el plegamiento de proteínas, revelando las posiciones de todos los átomos en una resolución temporal de femtosegundos (10^{-15} s). Además, las simulaciones pueden predecir cómo responderán las biomoléculas, cuando existe alguna perturbación sobre ellas, el caso de una mutación, fosforilación, protonación o la adición o eliminación de un ligando (Hollingsworth & Dror, 2018).

En una DM, se tratan las biomoléculas y el agua de solvatación como partículas que interactúan entre sí a través de una función clásica de energía potencial llamada "*campo de fuerza*" (Milardi & Pappalardo, 2015). Dadas las posiciones y velocidades de todos los átomos de un sistema, es posible calcular la fuerza ejercida sobre cada átomo por todos los demás átomos. Por tanto, se pueden utilizar las leyes del movimiento de Newton para predecir la posición espacial de cada átomo en función del tiempo. Esto sucede en repetidas ocasiones, actualizando la posición y velocidad de cada átomo del sistema, hasta que la trayectoria resultante es, en esencia, una película tridimensional que describe las transiciones conformacionales a nivel atomístico del sistema en cada punto durante el intervalo de tiempo simulado (Hollingsworth & Dror, 2018).

Para asegurar la estabilidad numérica, los pasos de tiempo en una simulación deben ser cortos, típicamente solo unos pocos femtosegundos cada uno. La mayoría de los eventos de interés bioquímico tienen lugar en escalas de tiempo de nanosegundos, microsegundos o un poco más. Por tanto, una simulación típica implica millones o miles de millones de pasos de tiempo. Este hecho, combinado con las millones de interacciones entre átomos evaluadas durante cada paso de

tiempo, hacen que las simulaciones sean muy exigentes desde el punto de vista computacional (Hollingsworth & Dror, 2018).

El poder de estas simulaciones radica en la posibilidad de capturar la posición y el movimiento de cada átomo en un sistema respecto al tiempo, algo que actualmente es muy difícil obtener empleando cualquier técnica experimental, pero además, es posible controlar cuidadosamente las condiciones de la simulación, como la conformación inicial de una proteína, los ligandos que están unidos a ella, si tiene mutaciones o modificaciones postraduccionales, qué otras moléculas están presentes en su entorno, su estado de protonación, temperatura, voltaje a través de una membrana, entre otras. De manera que aunque en la actualidad los recursos computacionales para realizar simulaciones son cada vez más accesibles, se requiere de un alto nivel de experiencia para formular preguntas clave capaces de abordarse mediante ésta técnica, diseñar las simulaciones e interpretar los resultados, generando información y conocimiento valioso sujeto a probarse por técnicas experimentales (Hollingsworth & Dror, 2018).

5. JUSTIFICACIÓN

El número de casos y de defunciones de la pandemia ocasionada por SARS-CoV-2 continúan aumentando, aún después de tres años de distanciamiento social, las repercusiones no solo en el rubro de salud, sino también en los sectores sociales y económicos siguen de manifiesto. Actualmente se registran, según datos de la Universidad Johns Hopkins, Baltimore, USA, un total de 676,609,955 casos en todo el mundo y 6,881,955 muertes. Solo en los Estados Unidos Mexicanos estas cifras alcanzan los 7,483,444 casos confirmados y se contabilizan 333,188 defunciones al 08 de Junio del 2023 (consultado en <https://coronavirus.jhu.edu/map.html>). El enorme esfuerzo de la comunidad científica, así como el importante aporte económico de diferentes industrias e instancias de gobierno en el mundo, han logrado que a la fecha se cuenten ya con once vacunas aprobadas para uso de emergencia por la organización mundial de la salud (WHO), que aportan un nivel importante de defensa para la población en contra del patógeno (COVOVAX, *Serum Institute of India*; Nuvaxovid, *Novavax*; Spikevax, *Moderna*; Comirnaty, *Pfizer/BionTech*; Convidecia, *Cansino*; Jcovden, *Johnson & Johnson*; Vaxweria, *Oxford/AztraZeneca*; Covishield, *Serum Institute of India*; Covaxin, *Bharat Biotech*; Covilo, *Sinopharm*; CoronaVac, *Sinovac*) y recientemente un fármaco aprobado (Paxlovid, *Pfizer*), el cual aún no se distribuye y comercializa de manera regular. Después de un tardado proceso de vacunación, obstaculizado en parte por la baja aceptación de algunos individuos debido al pronto desarrollo y manufactura de las vacunas, ha permitido el regreso a una situación de “nueva normalidad” y con esto el término del confinamiento. No obstante, se siguen contabilizando casos de COVID-19 en la actualidad, algunos de los cuales terminan en la desafortunada muerte de las personas contagiadas. A la fecha, se cuenta tan solo con Paxlovid para tratar la enfermedad COVID19, lo cual es insuficiente para tratar a las personas con dicha enfermedad. Aunque teóricamente hay varias maneras posibles de controlar la infección por coronavirus humanos, incluyendo vacunas, anticuerpos monoclonales, terapias basadas en oligonucleótidos y péptidos, el desarrollo efectivo de estas terapias puede tardar desde meses hasta varios años. Debido a la urgencia por encontrar un tratamiento contra SARS-CoV-2, esta propuesta se enfocará a evaluar

compuestos orgánicos de bajo peso molecular que han sido ya reportados por diferentes grupos de investigación como moléculas líderes y potenciales fármacos inhibidores de las proteasas Mpro y PLpro de SARS-CoV-2. Estas enzimas han sido propuestas por la comunidad científica como blancos idóneos para el diseño de fármacos debido a su función fundamental en la replicación del virus y a su nivel de conservación en los coronavirus. El diseño de fármacos dirigido a enzimas conservadas como Mpro y PLpro pueden ser efectivos y de amplio espectro, por lo que el estudio y dilucidación del espacio químico de los compuestos encontrados permitirá identificar patrones estructurales asociados con una alta probabilidad de actividad biológica, lo cual constituirá un punto de partida importante para la búsqueda de nuevos compuestos, su posterior optimización y su evaluación experimental.

6. HIPÓTESIS

El estudio del espacio químico relacionado con las moléculas reportadas con actividad inhibitoria contra las proteasas Mpro y PLpro de SARS-CoV-2 permite no solamente identificar patrones estructurales asociados con alta actividad biológica sino también proponer una molécula candidato con el potencial de convertirse en un fármaco contra SARS-CoV-2.

7. OBJETIVOS

7.1 OBJETIVO GENERAL

Estudiar el espacio químico relacionado con los compuestos reportados que presentan afinidad por los sitios activos de las proteasas Mpro y PLpro de SARS-CoV-2. Seleccionar a los mejores candidatos para la búsqueda y/o diseño de nuevas moléculas con potencial para inhibir la actividad enzimática de las proteasas.

7.2 OBJETIVOS ESPECÍFICOS

1. Elaboración de una quimioteca de compuestos con afinidad por los sitios activos de las proteasas Mpro y PLpro de SARS-CoV-2
2. Generar el espacio químico de los compuestos recopilados, utilizando descriptores de relevancia farmacéutica mediante técnicas de quimioinformática.
3. Selección de compuestos con propiedades fisicoquímicas idóneas mediante el análisis del espacio químico y simulaciones de acoplamiento molecular.
4. Evaluación de las energías de interacción y análisis de la conformación adoptada en el sitio activo de las proteasas Mpro y PLpro de compuestos con estructuras prometedoras por simulaciones de acoplamiento molecular.
5. Análisis de patrones estructurales de los compuestos con mejor acoplamiento, búsqueda de nuevos compuestos a partir de dichos patrones y análisis por acoplamiento molecular de los compuestos resultantes.
6. Realizar ensayos de simulación atomística por DM de los compuestos orgánicos que muestren los mejores parámetros de acoplamiento con al menos una de las dos proteasas de SARS-CoV-2, Mpro y PLpro.

8.- ESTRATEGIA EXPERIMENTAL GENERAL

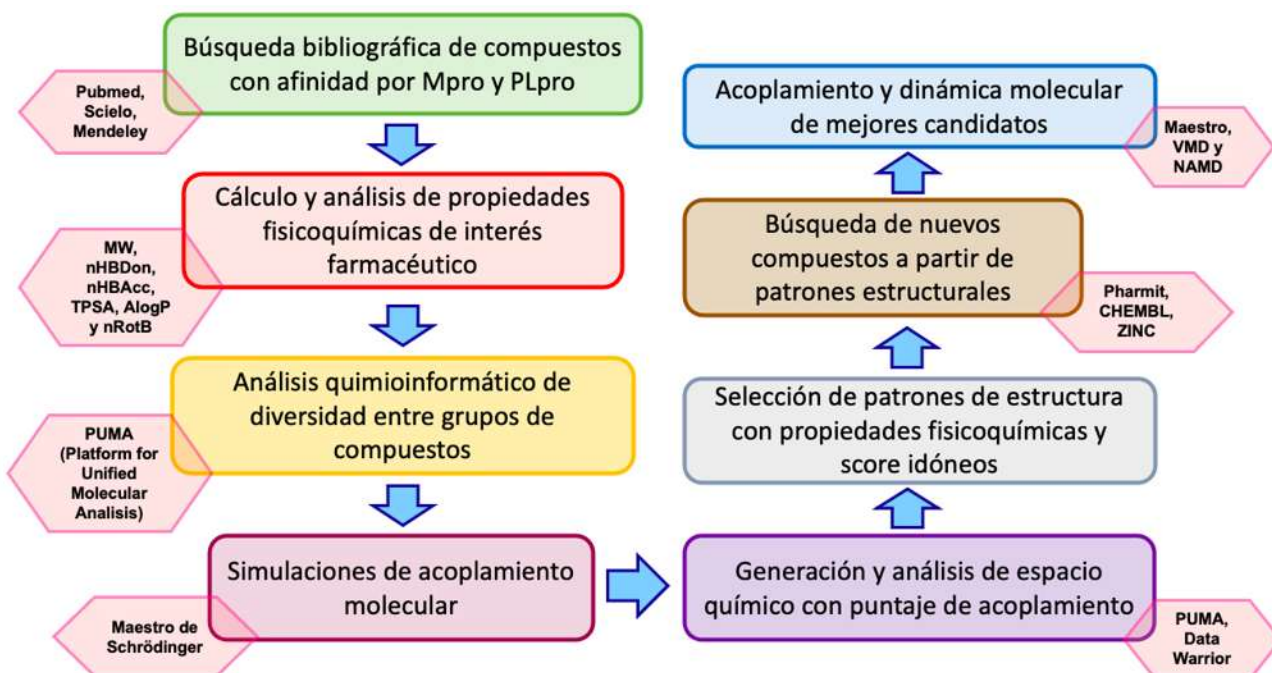


Figura 9. Diagrama de flujo de trabajo. El diagrama muestra la secuencia de etapas para obtener compuestos orgánicos *leads* con alta afinidad por el sitio activo de Mpro y PLpro de SARS-CoV-2.

9.- MATERIALES Y MÉTODOS

9.1 IDENTIFICACIÓN DE COMPUESTOS ORGÁNICOS INHIBIDORES DE Mpro Y PLpro REPORTADOS EN LA LITERATURA

La identificación de compuestos inhibidores de Mpro y PLpro se logrará con una revisión bibliográfica en diferentes gestores de información como PubMed, Scielo y Mendeley. Se creará un compilado de la información de los compuestos que incluirá el país y año de la publicación, el blanco biológico (Mpro, PLpro o ambas), el origen del compuesto, si es natural se incluye la especie, quimioteca, metodología del ensayo de acoplamiento molecular, puntaje (*score*), y actividad biológica para aquellos compuestos que la reportan. Una vez seleccionados los compuestos y con las estructuras de Mpro y PLpro preparadas se analizará la interacción de los compuestos con el sitio activo, mediante acoplamiento molecular *in silico* en dos

softwares diferentes. De estas pruebas, se seleccionarán los mejores candidatos con base en su energía de interacción (entalpía) y la geometría de acoplamiento (orientación en el sitio activo). En este paso es importante comenzar a visualizar las principales interacciones no covalentes que se establecen entre el sitio activo y los compuestos. Se incluirán también controles positivos que consistirán en ensayos de acoplamiento entre compuestos previamente reportados que se unen con alta afinidad al sitio activo de Mpro (Dai *et al.*, 2020).

9.2 CURADO DE LA BASE DE DATOS (AlvaMolecule)

AlvaMolecule es un software que permite visualizar, analizar, curar y estandarizar una base de datos de compuestos químicos. Este software es gratuito para uso académico y funciona como una hoja de trabajo en la que se puede visualizar la base de datos como una cuadrícula de moléculas o bien como una hoja de cálculo. Los datos se suministran al software en un archivo de extensión .sdf con los compuestos representados por su código SMILES y un identificador. Para la revisión de los compuestos, AlvaMolecule proporciona 10 verificadores que ayudan a identificar estructuras erróneas o a filtrar características estructurales específicas: estructuras múltiples, valencia inusual, enlace covalente/iónico, carga total, isótopos, átomos cargados, átomos que no son carbono, set de átomos no estándar (H,C,N,O,P,S,F,Cl,Br,I), aromaticidad y átomos radicales. Además, el software tiene la función de estandarizar las moléculas y puede calcular un set de 88 descriptores moleculares y propiedades fisicoquímicas como el peso molecular, refractividad molar, área polar topológica superficial (TPSA), estimaciones de volumen molecular, modelos de LogP, entre otros. Estas propiedades ayudan a filtrar los compuestos y a preparar una base de datos para análisis posteriores. La versión que se utilizó es la 1.0.4 para un software de 64 bits disponible para descarga en el sitio web de AlvaScience (<https://www.alvascience.com/alvamolecule/>).

9.3 CÁLCULO DE PROPIEDADES FÍSICOQUÍMICAS DE INTERÉS FARMACEÚTICO

El cálculo de seis propiedades fisicoquímicas de interés farmacéutico (MW, CLogP, TPSA, nHBDon, nHBAcc y nRotB) se llevó a cabo en la plataforma PUMA (Platform for Unified Molecular Analysis) (González-Medina & Medina-Franco, 2017). Estas propiedades describen la polaridad de la molécula (HBDon, HBAcc, LogP), flexibilidad (TPSA, nRotB) y tamaño (MW), y permiten comparar un conjunto de moléculas con valores de propiedades fisicoquímicas propias de fármacos aprobados y sugerir si estos compuestos podrían ser activos por vía oral (Bickerton *et al.*, 2012). Para probar si los valores de los resultados seguían una distribución normal se realizaron pruebas de normalidad de Shapiro-Wilk y para evaluar si existía diferencia significativa entre los grupos se utilizó la prueba de sumas de rangos de Wilcoxon. El test de suma de rangos de Wilcoxon es el análogo al test t-student para estadística paramétrica donde se contrasta si una población de n número de observaciones es estadísticamente igual o diferente a otra, pero en el caso de que se tenga una distribución ya sea desconocida o no normal de los datos. Esta es una prueba de estadística no paramétrica donde la medida comparativa determinante de si existe diferencia o no entre las poblaciones es su valor de mediana y no de media. Para esto se hace un ordenamiento de todas las observaciones y se les asigna un valor de rango de acuerdo a su magnitud, la suma de los rangos (R_1 y R_2) para cada conjunto de datos son valores necesarios en el cálculo de los valores U. El valor más bajo de la variable U, será utilizada en el cálculo del valor de probabilidad Z. El cual mediante su comparación con los valores de probabilidad establecidos en tablas para el estadístico dará la pauta para decidir si las poblaciones son estadísticamente iguales o diferentes.

$$\begin{array}{l} \textcircled{1} \\ U_1 = n_1 \cdot n_2 + \frac{n_1(n_1 + 1)}{2} - R_1 \\ \textcircled{2} \\ U_2 = n_1 \cdot n_2 + \frac{n_2(n_2 + 1)}{2} - R_2 \\ U = \min(U_1, U_2) \end{array} \quad \begin{array}{l} \textcircled{3} \\ Z = \frac{U - \left(\frac{n_1 \cdot n_2}{2}\right)}{\sqrt{\frac{n_1 \cdot n_2(n_1 + n_2 + 1)}{12}}} \\ \text{Si } Z \leq Z_\alpha \quad \rightarrow \text{ Se acepta la } H_0 \\ \text{Si } Z > Z_\alpha \quad \rightarrow \text{ Se rechaza la } H_0 \end{array}$$

$$H_0: Me_1 = Me_2$$

Figura. 10. Fórmulas para el cálculo de la prueba de sumas de rangos de Wilcoxon.

Los valores de la variable U son determinantes del rango significativo para considerar en la prueba, en las fórmulas 1 y 2 se calculan los valores de U para cada conjunto de datos, donde n_1 y n_2 son el número de datos para cada conjunto y R_1 y R_2 son las sumas de los rangos de los datos correspondientes a cada conjunto. En la fórmula 3 se observa la fórmula para calcular Z, que será el valor de probabilidad a probar en las tablas estadísticas para prueba de Wilcoxon y la U utilizada en el cálculo para el valor Z es la U de valor mínimo para los dos conjuntos de datos. La prueba determina si las medianas (Me) de los conjuntos son estadísticamente iguales o diferentes.

9.4 DIVERSIDAD MOLECULAR POR DISTANCIA EUCLIDIANA E ÍNDICE DE TANIMOTO

El concepto de diversidad es utilizado de manera general para medir la extensión de los objetos pertenecientes a un conjunto, de acuerdo con sus características diferenciales (Pérez, 2005). Evaluar la diversidad de un conjunto de datos permite evaluar, ya sea cualitativa o cuantitativamente, las probabilidades de encontrar nuevos miembros con características diferentes en un conjunto de datos, así como analizar aquellos datos con similitud entre ellos. En química, las propiedades moleculares y la actividad biológica de los compuestos pueden estar relacionados con su estructura molecular, por lo que el análisis de diversidad molecular se centra en la evaluación de la extensión de moléculas con características estructurales diferenciales, asumiendo, de acuerdo con el paradigma de estructura-actividad, que el proceso de análisis también distingue la diversidad de un conjunto de moléculas en términos de sus propiedades y actividad biológica (Pérez, 2005).

Se han descrito en la literatura diferentes procedimientos para medir diversidad o similitud molecular y la elección de uno u otro depende en gran medida de los objetivos del estudio y de la forma en que se describan las moléculas (Nikolova & Jaworska, 2003). En el presente estudio se emplea la medición de distancia Euclidiana y el índice de similitud de Tanimoto (Tanimoto, 1958), como métodos para medir la diversidad y la similitud respectivamente, entre los compuestos de las bases de datos. La distancia euclidiana puede verse como la distancia existente entre un par de compuestos en un espacio químico y es la desviación cuadrática media (RMSD) de una propiedad P, que se compara entre dos moléculas A y B, sumadas a las mediciones con los demás puntos de la cuadrícula. Por lo tanto, de acuerdo con la métrica de distancia Euclidiana, compuestos más alejados entre sí son más diversos.

$$rmsd = \sqrt{\frac{\sum (P_{Ai}^2 - P_{Bi}^2)}{N}}$$

El índice de similitud de Tanimoto permite evaluar la similitud entre objetos de acuerdo a descriptores asociados con ellos. El índice mide la relación del conjunto de características que tienen en común dos compuestos con respecto a las características totales de ambos compuestos (Tanimoto, 1958).

$$d_{AB} = \frac{C}{A+B-C}$$

Donde A representa las características del objeto A, de manera similar B representa las características del objeto B y C representa las características que A y B tienen en común (Tanimoto, 1958; Pérez, 2005). Los valores del índice de Tanimoto van del 0 al 1, donde cuanto más cercano sea el valor a 1, mayor similitud tienen los objetos comparados. En el presente estudio se utiliza esta métrica para medir la similitud entre propiedades fisicoquímicas y de huellas dactilares moleculares o fingerprints. En quimioinformática una huella dactilar molecular (Molecular Fingerprint), es una forma de simplificar la estructura de una molécula en un vector

lineal binario, donde cada posición del vector indica con un (0) la ausencia o con un (1) la presencia de una característica predeterminada. De esta manera, el índice de Tanimoto puede utilizarse de manera que A es el número de “1” en el vector representativo de la molécula A, B es el número de “1” en el vector de la molécula B y C es el número de “1” que poseen los vectores A y B de manera simultánea (Pérez, 2005).

9.5 CARACTERIZACIÓN DEL ESPACIO QUÍMICO Y ANÁLISIS QUIMIOINFORMÁTICO CON PLATFORM FOR UNIFIED MOLECULAR ANALYSIS (PUMA)

Los análisis de espacio químico, contenido de “*scaffolds*” y diversidad estructural son prácticas frecuentes en quimioinformática que encaminan la búsqueda de compuestos y la selección de bases de datos químicas para estudios de acoplamiento y optimización de moléculas (Medina-Franco, 2012). PUMA es un servidor de libre acceso desarrollado por la escuela de química de la Universidad Nacional Autónoma de México que permite calcular las propiedades moleculares, visualizar el espacio químico y calcular la diversidad estructural y química de bases de datos de compuestos (González-Medina & Medina-Franco, 2019). El servidor integra el cálculo de descriptores y la representación visual del espacio químico con base en esos descriptores (Medina-Franco *et al.*, 2021). Además, esta plataforma realiza seis análisis diferentes para la caracterización de las bases de datos: visualización de espacio químico, estadísticos de propiedades, propiedades de similitud y distancia, curvas CSR para recuperación de sistemas cíclicos, frecuencia de scaffolds y entropía de Shannon y función de distribución acumulativa (CDF) de valores de similitud por pares calculados con huellas dactilares moleculares o “molecular fingerprints” (González-Medina & Medina-Franco, 2017).

Para la caracterización y visualización del espacio químico, PUMA realiza un análisis de componentes principales (PCA) basado en seis propiedades fisicoquímicas de relevancia farmacéutica que a menudo evalúan el carácter drogable de los compuestos: MW, nHBD_{on}, nHB_{Acc}, TopoPSA, nRotB y ALogP) (González-Medina & Medina-Franco, 2017). Teóricamente el espacio químico

podría ser unidimensional, bidimensional o tridimensional y se representaría de manera sencilla usando diagramas de dispersión; no obstante, cuando se tienen más de 3 variables, es conveniente usar un enfoque matemático para reducir las dimensiones y poder visualizar los datos con mayor facilidad (Medina-Franco *et al.*, 2021). El PCA es un método matemático que reduce un espacio vectorial n -dimensional a un espacio de menor dimensión, en su mayoría dos o tres dimensiones, mientras mantiene las características que más contribuyen a la varianza de los datos. Los nuevos vectores base están formados por combinaciones lineales de los n vectores base del espacio vectorial anterior y los PC se eligen en orden de acuerdo con la variable que contribuya mayormente a la varianza de los datos, de esta manera PC1 será la variable que contribuye más a la varianza de los datos, PC2 será la segunda que más contribuye y así sucesivamente (Wetzel *et al.*, 2007). En PUMA, el usuario suministra a la plataforma la base de datos curada donde se encuentra el código SMILES de las moléculas, así como un identificador y la referencia a su base de datos y el servidor realiza los cálculos de las propiedades fisicoquímicas internamente (descriptores) y posteriormente el análisis de componentes principales. Para esto PUMA utiliza la función de R “prcomp” y escala las seis propiedades moleculares a una media de cero y una varianza de uno y calcula seis componentes principales, el usuario decide qué PCs graficar en dos o tres dimensiones y puede además descargar el resumen de los cálculos y la información respectiva al análisis (González-Medina & Medina-Franco, 2017).

9.6 ANÁLISIS DE SCAFFOLDS POR CURVAS DE RECUPERACIÓN DE SISTEMAS CÍCLICOS (CSR) Y ENTROPÍA DE SHANNON

El término “*scaffold*” que al español se traduce como andamio o marco molecular es un concepto utilizado en química medicinal para hacer referencia a la estructura central de un compuesto químico al que se le unen grupos funcionales (Hu *et al.*, 2016). El análisis de *scaffolds* permite la exploración de relaciones estructura-actividad, diversidad estructural y el cribado de compuestos con estructuras novedosas o de interés farmacéutico (Medina-Franco *et al.*, 2009). En el presente estudio se utilizan dos metodologías para medir la diversidad de *scaffolds* en las bases de datos utilizadas. La primera metodología es la generación de curvas de

recuperación de sistemas cíclicos (CSR), las cuales dan un indicativo de la distribución de moléculas sobre los diferentes *scaffolds*. En las curvas CSR, se traza la fracción de *scaffolds* (eje x), frente a la fracción de compuestos que contienen esos *scaffolds* (eje y). En casos extremos donde exista un conjunto de datos con la máxima diversidad, cada *scaffold* representaría o contendría el mismo número de compuestos que los otros *scaffolds* y el gráfico sería una diagonal que va del (0%, 0%) a (100%, 100%), por lo tanto, entre más se acerque una curva a la diagonal, los compuestos se distribuyen más uniformemente y son más diversos. La fracción de quimiotipos necesarios para recuperar el 50% de las moléculas (F50) y el área bajo la curva (AUC) son parámetros que ayudan a caracterizar e interpretar estas curvas. En el caso de máxima diversidad la curva CSR representaría una diagonal con un AUC de 0.5. (González-Medina & Medina-Franco, 2019; Langdon *et al.*, 2011). La segunda metodología es el uso de la métrica de Entropía de Shannon, la cual permite conocer la distribución específica de compuestos en el número n de quimiotipos o *scaffolds* más poblados. La entropía de Shannon (SE) de una población de P compuestos contenidos en N quimiotipos se define como:

$$SE = - \sum_{i=1}^n p_i \log_2 p_i; \quad p_i = c_i/P$$

donde p_i es la frecuencia relativa del quimiotipo i en una población de P compuestos que contienen n quimiotipos distintos; c_i corresponde al número absoluto de moléculas que contienen un quimiotipo particular i . Los valores de SE van desde 0 a $\log_2 n$ por lo que dependen de n (número de quimiotipos distintos) pero no explícitamente de P (número de compuestos). Si $SE = \log_2 n$, es decir, su valor máximo, los P compuestos están distribuidos uniformemente en los n quimiotipos, lo que representa una diversidad máxima de quimiotipos en la base de datos. Para normalizar los valores de SE para usar diferentes valores de n , se utiliza la entropía escalada de Shannon (SSE) que se define como:

$$SSE = \frac{SE}{\log_2 n}$$

De esta manera los valores de SSE varían entre el 0, donde todos los compuestos P están contenidos en un solo quimiotipo, y 1.0, donde cada quimiotipo contiene el mismo número de compuestos. Por lo tanto, los valores de SSE más cercanos a 1.0 indican una gran diversidad de quimiotipos o Scaffolds dentro de los n quimiotipos más poblados (Medina-Franco *et al.*, 2009).

9.7 ACOPLAMIENTO MOLECULAR CON GLIDE

El software *Glide* (<https://www.schrodinger.com/>) proporciona una gran versatilidad y velocidad de bajo costo computacional para realizar en primera instancia cribado virtual de alto rendimiento, que permite seleccionar en tiempo corto un número importante de compuestos *hits*. Estos *hits* son procesados después en el modo SP (precisión estándar) de *Glide* para seleccionar los compuestos que se acoplen a la proteasa con cierto grado de confiabilidad. La última etapa consiste en aplicar el modo XP (alta precisión) de *Glide*, por medio del cual se eliminan los falsos positivos al introducir análisis más restrictivos en los datos obtenidos (Friesner *et al.*, 2004). La suite de *Glide* incluye la interfase Maestro, el campo de fuerza OPLS3e, y los softwares *Protein Preparation Wizard*, *Epik* y *LigPrep*. El análisis de los datos obtenidos permitirá seleccionar a los mejores compuestos (*leads*) para ensayos de dinámica molecular.

9.8 GENERACIÓN DE MODELO DE FARMACÓFORO

Para generar el modelo de farmacóforo se utilizó la plataforma *Pharmit* (<http://pharmit.csb.pitt.edu>), la cual ofrece un entorno interactivo en línea para el cribado virtual de grandes bases de datos de compuestos mediante la generación de modelos de farmacóforo, estructura molecular y minimización de energía. Después de analizar las interacciones importantes entre los compuestos acoplados con mejor puntaje y el sitio activo de Mpro, se seleccionaron las partes más importantes del ligando responsables de su unión al receptor y se guardaron en un

archivo con formato .pdb. Dicho archivo se utilizó para especificar en pharmit las características relevantes para generar el modelo de farmacóforo, a partir del cual se cribaron las bases de datos ZINC (<https://zinc.docking.org>, 13,127,550 moléculas) y ChEMBL30 (<https://www.ebi.ac.uk/chembl/>, 2,186,411 moléculas). Los compuestos resultantes fueron descargados y acoplados con la proteína con el protocolo establecido anteriormente.

9.9 SIMULACIONES ATOMÍSTICAS DE DINÁMICA MOLECULAR DE LA INTERACCIÓN ENTRE LA PROTEASA PRINCIPAL M_{pro} DE SARS-COV-2 Y COMPUESTOS LEADS CON NAMD

Una vez obtenidos los mejores compuestos orgánicos *leads* se procederá a evaluar la estabilidad de su interacción con la proteasa principal M_{pro} de SARS-CoV-2 mediante la técnica computacional de simulaciones atomísticas de dinámica molecular (*all-atom MD simulations*) en un ambiente fisiológico relevante para el sistema. Para esto se usará el software *nanoscale molecular dynamics* (**NAMD**) (<https://www.ks.uiuc.edu/Research/namd/>) desarrollado por el Theoretical and Computational Biophysics Group y el Parallel Programming Laboratory así como el campo de fuerza **CHARMM** (<https://www.charmm.org/charmm/resources/charmm-force-fields/>). A partir de la estructura inicial del acoplamiento molecular, el complejo ligando-proteína será analizado y se asignará el estado de protonación más probable a los diferentes grupos químicos a un pH de 7.4. El sistema se solvatará en una caja de moléculas de agua usando el modelo TIP3P, el cual es compatible con el campo de fuerza CHARMM, y subsecuentemente se agregarán iones de sodio y cloro para replicar una concentración salina de 0.15 M. La temperatura y la presión del sistema se mantendrán constantes a 37 °C y 1 atm, respectivamente. El tamaño de paso de la simulación que se utilizará será de 2.0 fs obteniendo trayectorias de simulación de 150 ns (75,000,000 pasos). Los protocolos utilizados para realizar las simulaciones han sido descritos anteriormente (Rebolledo-Bustillo *et al.*, 2023; Dávila *et al.*, 2022; Rangel-Galván *et al.*, 2021). Los cálculos se realizarán en el Laboratorio de Supercómputo del Sureste de México (LNS-BUAP). Los resultados se visualizarán y analizarán con el programa VMD

(<https://www.ks.uiuc.edu/Research/vmd/>). Los resultados de estas simulaciones proporcionarán información muy detallada con respecto a la importancia de diferentes grupos funcionales de los ligandos, así como también la relevancia de algunos residuos del sitio activo en la formación del complejo.

10.- RESULTADOS

10.1 IDENTIFICACIÓN DE COMPUESTOS ORGÁNICOS INHIBIDORES DE Mpro Y PLpro REPORTADOS EN LA LITERATURA Y CREACIÓN DE QUIMIOTECA

El conjunto de moléculas a considerar para la formación de la base de datos química o quimioteca se obtuvo a través de una búsqueda sistemática en la literatura científica en los gestores de información PubMed, Mendeley y Scielo, con las palabras clave “SARS-CoV-2 inhibitors”, “SARS-CoV-2 Mpro inhibitors”, “SARS-CoV-2 PLpro inhibitors” y “SARS-CoV-2 protease inhibitors”. Además, se incluyó información sobre compuestos propuestos para la inhibición de las proteasas Mpro y PLpro de SARS-CoV (2003) y de fármacos reposicionados contra las proteasas de SARS-CoV-2 con las palabras clave “SARS-CoV protease inhibitors” y “SARS-CoV-2 Drug repurposing”, respectivamente. Como resultado de la búsqueda se obtuvieron 97 artículos científicos (Tabla 1, material suplementario), donde los criterios de inclusión fueron el uso de metodologías computacionales como cribado virtual y dinámica molecular, reposicionamiento de fármacos, así como la referencia para el rastreo de las moléculas en su formato SMILES. El formato SMILES del inglés, Simplified Molecular Input Line Entry Specification es una notación lineal que sirve para representar la estructura química de una molécula, utilizando cadenas de caracteres alfanuméricos. A partir de la literatura recopilada, se generó una base de datos que contiene el título del artículo, país de publicación, fecha, nombre e identificador del compuesto, origen natural o sintético, metodología computacional empleada en la publicación, quimioteca de la cual se extrajo, actividad biológica si el compuesto la posee, estructura química en formato SMILES y peso molecular. Como resultado de la revisión y previo al análisis, se recopilaron 965 compuestos que fueron sometidos a un proceso de curado y preparación, del cual se obtuvo un

total de 866 compuestos que forman la quimioteca (tabla 1). El curado de la quimioteca comprende la detección y eliminación de compuestos repetidos, remanentes salinos, compuestos con estructuras múltiples, con valencia inusual y/o con átomos no estándar que puedan significar toxicidad. El curado se realizó mediante la inspección manual guiada por el software Alva-Molecule para aquellas propiedades y características químicas y Windows Excel para la eliminación de los compuestos duplicados.

Tabla 1. Blanco biológico y número de compuestos obtenidos de la literatura científica

Blanco biológico	No. De compuestos
Mpro	576
PLpro	290

10.2 CÁLCULO DE PROPIEDADES FISICOQUÍMICAS DE INTERÉS FARMACEÚTICO DE LOS COMPUESTOS REPORTADOS

Con el objetivo de evaluar la diversidad química y el carácter “*Drug-likeness*” de los compuestos se obtuvo, mediante un análisis quimioinformático en la plataforma PUMA, información sobre propiedades fisicoquímicas de interés farmacéutico y su visualización en el espacio químico, análisis de similitud entre conjuntos, “*scaffoldings*” o estructuras base y huellas dactilares moleculares “*fingerprints*” (González-Medina & Medina-Franco, 2017). Como primer punto del análisis quimioinformático se calcularon seis propiedades fisicoquímicas frecuentemente asociadas con un buen perfil farmacéutico (MW, HBAcc, HBDon, nRotB, ALogP y TPSA). La medición de estas propiedades permite comparar nuevos compuestos contra fármacos aprobados y sugerir si estos podrían ser activos y administrados por vía oral (Bickerton *et al.*, 2012). La tabla 2 resume la distribución estadística de las seis propiedades calculadas para ambos conjuntos de compuestos, donde “*Data set*” se refiere al conjunto al que pertenecen las moléculas, dependiendo de su afinidad por la proteasa correspondiente (Mpro o PLpro), “*Min*” indica el valor mínimo, “*Max*” indica el valor máximo y “*1er Qu*” y “*3er Qu*” corresponden a los

valores del primer y tercer cuartil, respectivamente. En la figura 11 se observa además la distribución de estas propiedades en su histograma correspondiente.

Tabla 2. Distribución estadística de las propiedades MW, AlogP, nHBAcc, nHBDon, nRotB y TPSA para compuestos con afinidad por Mpro y PLpro.

Propiedad	Data set	Min	1er Qu	Mediana	Media	3er Qu	Max	Desv. Est
MW	Mpro	102	325	419	444	531	1700	170
	PLpro	125	304	390	402	454	888	144
AlogP	Mpro	-13	-1	0	0	1	9	2
	PLpro	-8	-1	0	0	2	5	2
nHBAcc	Mpro	0	4	6	7	9	46	5
	PLpro	0	4	5	6	8	16	3
nHBDon	Mpro	0	1	2	3	4	25	3
	PLpro	0	1	2	2	3	10	2
nRotB	Mpro	0	3	6	7	9	42	6
	PLpro	0	4	6	6	8	28	4
TPSA	Mpro	0	75	101	122	146	778	78
	PLpro	3	55	91	97	123	267	53

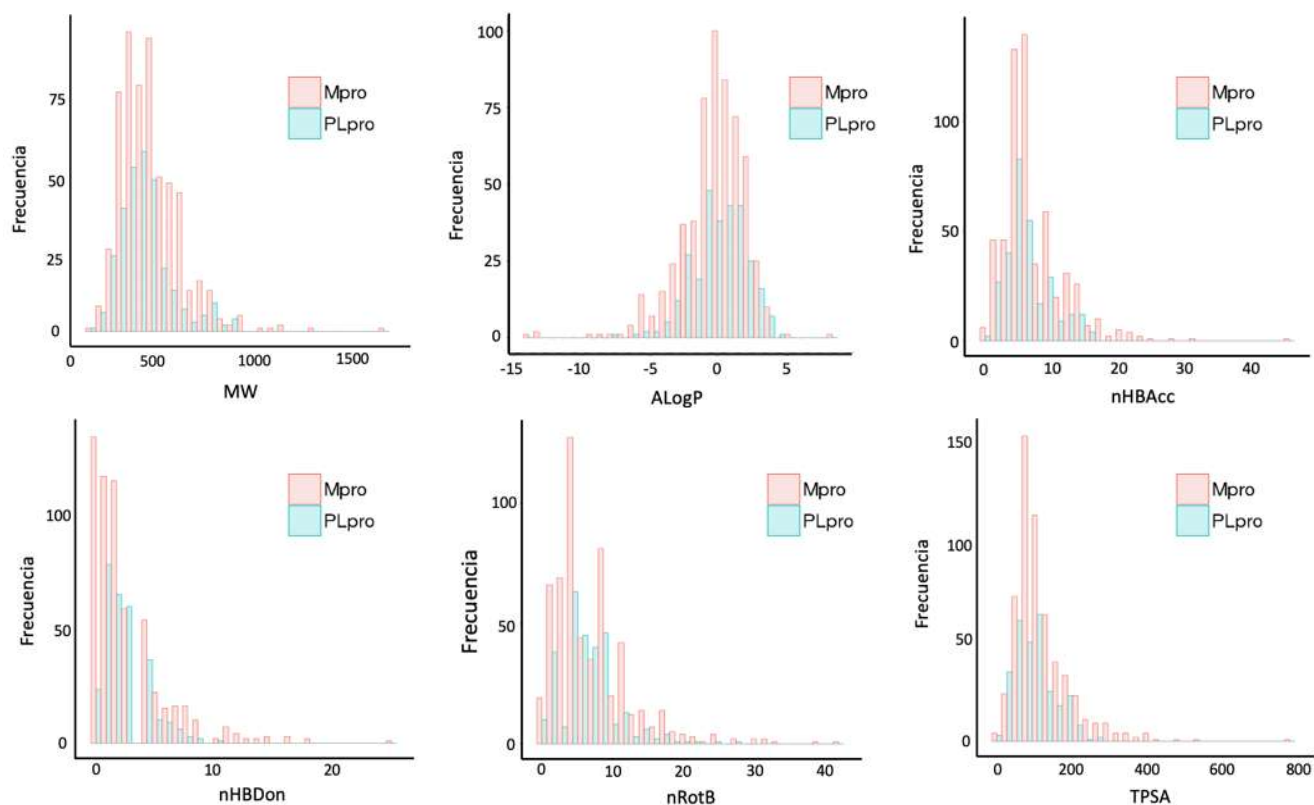
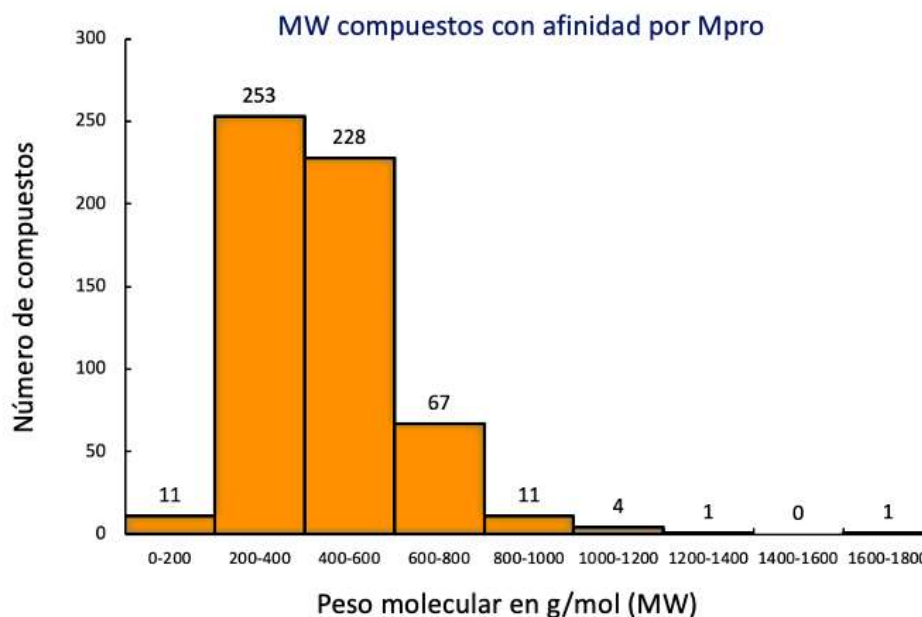


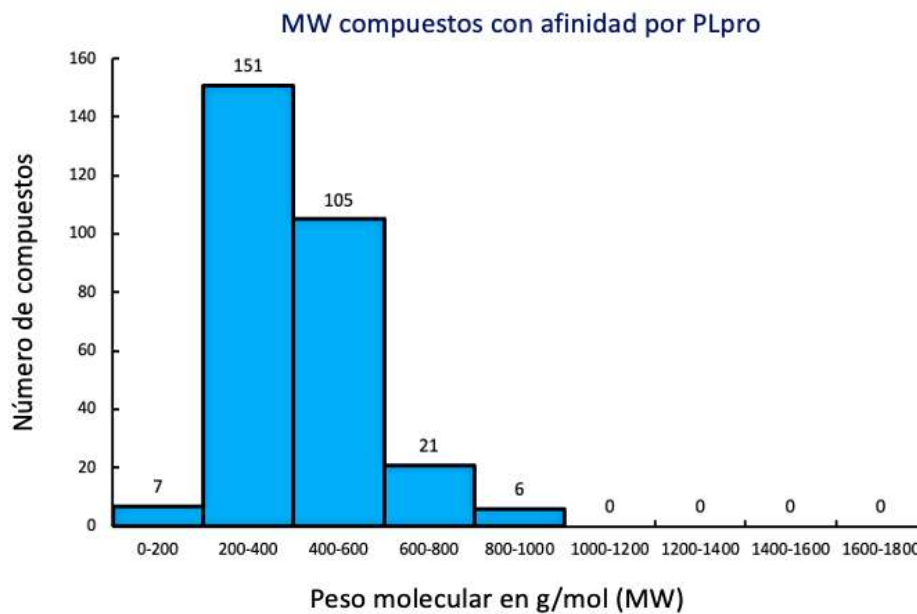
Figura 11. Histograma de seis propiedades fisicoquímicas de compuestos encontrados en la literatura con afinidad por las proteasas Mpro y PLpro.

El primer paso en el análisis de las seis propiedades fisicoquímicas consistió en realizar pruebas de normalidad para cada una de las propiedades y conjuntos de datos, lo que resultó en que ninguna de las propiedades para ambos conjuntos de datos sigue una distribución normal. Debido a esto, se realizó el test estadístico de suma de rangos de Wilcoxon, con el objetivo de determinar si existe una diferencia estadísticamente significativa para alguna de las propiedades fisicoquímicas entre ambos grupos de moléculas. Los resultados de las pruebas indican que las propiedades AlogP y nRotB no presentan diferencias estadísticamente significativas entre ambos sets de moléculas, mientras que las propiedades MW, nHBAcc, nHBDon y TPSA sí las presentan con valores más altos para los compuestos reportados con afinidad por la proteasa principal Mpro. Esto en general indica que los compuestos con afinidad por Mpro son más hidrofílicos que aquellos con afinidad por PLpro, ya que tienen mayor probabilidad de formar interacciones electrostáticas y por puentes de hidrógeno. La diferencia en tamaño de las moléculas, que resultó ser mayor para compuestos afines a Mpro, podría explicar estas diferencias. Además, los resultados muestran que existen moléculas reportadas con afinidad por Mpro que tienen valores de MW muy elevados, lo que los convierte en datos atípicos que contribuyen a la diferencia entre conjuntos de compuestos en los resultados obtenidos, no obstante, son considerados hits computacionales por los respectivos autores. Compuestos que tienen un $MW > 1000$ g/mol y se consideran datos atípicos son QST (MP205, $MW=1025.36$ g/mol), Caspofungina (MP533, $MW=1093.33$), Vapreotide (MP531, $MW=1131.39$), Colistina (MP525, $MW=1155.45$), Icatibant (MP527, $MW=1304.54$) y Ácido tánico (MP210, $MW=1701.2$, $IC_{50} = 2.1 \pm 0.2$ μ M) para Mpro. Mientras que los compuestos con mayor MW y afinidad por PLpro encontrados en esta revisión son Velpatasvir (PL290, $MW=883.016$) y Ledipasvir (PL251, $MW=889.014$). Para una mejor visualización de la distribución de los compuestos de acuerdo con su peso molecular se realizaron los gráficos 1 y 2 donde se observa que el 83% de los compuestos afines a Mpro se encuentran en un rango de 200 a 600 g/mol, mientras que el porcentaje en el mismo rango para compuestos afines a PLpro es de 88%. Para ambos conjuntos de moléculas se recupera el 95% de los compuestos en un rango

entre los 200 y 800 g/mol; no obstante, el número de compuestos que excede este rango es mayor para compuestos afines a Mpro, con datos atípicos que llegan hasta la categoría de 1600 a 1800 g/mol.



Gráfica 1. Peso molecular (MW) de moléculas encontradas en la literatura con afinidad por Mpro.



Gráfica 2. Peso molecular (MW) de moléculas encontradas en la literatura con afinidad por PLpro.

10.3 VISUALIZACIÓN DEL ESPACIO QUÍMICO BASADO EN LAS SEIS PROPIEDADES FISICOQUÍMICAS MW, HBAcc, HBDon, nRotB, ALogP y TPSA

La visualización del espacio químico se logró mediante un análisis de componentes principales (PCA) de las seis propiedades fisicoquímicas analizadas anteriormente para cada compuesto en la plataforma PUMA. El análisis por componentes principales permite reducir las dimensiones de las bases de datos para poderlas visualizar en dos o tres dimensiones, perdiendo la menor cantidad de información posible, de manera que lo que se obtiene es una visualización lo más representativa posible de la distribución de los compuestos de acuerdo con sus valores de propiedades fisicoquímicas. En teoría, compuestos más cercanos entre sí comparten una mayor similitud en sus propiedades. En la figura 12 se visualiza el espacio químico definido por los componentes principales PCa y PCb lo cuales son vectores formados resultantes de la combinación lineal de los vectores originales correspondientes a las seis propiedades fisicoquímicas medidas (MW, HBAcc, HBDon, nRotB, ALogP y TPSA). Estos vectores PCa y PCb recuperan el 86% de la covarianza de los datos y asignan un mayor peso a aquellas propiedades fisicoquímicas que son más importantes para graficar con mayor exactitud la variabilidad del conjunto de datos. En este caso, la tabla 3 indica que las propiedades MW, TPSA, nHBDon y nHBAcc tienen una mayor contribución en la formación del PCa, mientras que las propiedades nRotB y AlogP contribuyen con mayor importancia en la formación del PCb. En el caso de PCa, compuestos desplazados hacia valores negativos tienen mayor MW, TPSA, nHBDon y nHBAcc, lo cual es indicativo de hidrofiliidad, por otra parte, compuestos desplazados hacia valores negativos en PCb tienen mayor número de nRotB y AlogP lo que puede significar compuestos más flexibles y con mejor solubilidad. Se pueden observar algunos datos atípicos en el gráfico de espacio químico que particularmente corresponden a la base de datos de compuestos afines a Mpro. Su estructura y propiedades se muestran en la figura 13. Dentro de ellos se encuentra al ácido tánico (MP210), un compuesto polifenólico que se encuentra en las agallas formadas por insectos en las ramas de ciertos robles y que se usa tópicamente para el tratamiento de herpes labial, ampollas y dermatitis. Tiene un peso molecular de

1700 g/mol y de acuerdo con Coelho *et al.* (2020) mediante un ensayo de cribado de alto rendimiento (HTS) tiene un $IC_{50} = 2.1 \mu M$ contra Mpro de SARS-CoV-2. Los compuestos Icatibant (MP527), Colistina (MP525) y Caspofungina (MP533) resultaron candidatos para inhibir a Mpro después de un ensayo de cribado virtual de la base de datos “Drugbank” en un intento por reposicionar fármacos ya comercializados para tratar pacientes con COVID-19 (Liu & Wang, 2020). Icatibant es un fármaco peptidomimético de 10 aminoácidos y MW=1303.6 g/mol que actúa como antagonista de los receptores B2 de bradicinina, un péptido fisiológico que causa vasodilatación, por lo que Icatibant ha sido aprobado por la unión europea para su uso en angioedema hereditario. Colistina es un antibiótico polipeptídico cíclico de *Bacillus colistinus*, tiene un MW=1154.74 g/mol y se utiliza para tratar infecciones bacterianas causadas por bacterias Gram negativas. Caspofungina es un fármaco antimicótico y primer miembro de una nueva clase de fármaco denominado equinocandinas por Merck & Co., Inc., tiene un MW=1092.6 g/mol y se utiliza para tratar una variedad de infecciones fúngicas. El compuesto Q5T que tiene por nombre biotin-PEG(4)-Abu-Tle-Leu-Gln-vinilsulfona tiene un MW=1024.5 y fue reportado en una revisión de estructuras de SARS-CoV-2 Mpro en complejo con potenciales inhibidores como una molécula capaz de interactuar con el sitio activo de la enzima y que podría tener potencial actividad biológica (Mengist *et al.*, 2021). Heptafuhalol A (MP2), un compuesto natural marino de MW=918.1 g/mol proveniente de algas como *Sargassum spinuligerum* y *Carpophyllum angustifolium* resultó un hit computacional del cribado virtual de una quimioteca de compuestos naturales de origen marino “Marine Natural Product (MNP) library”, donde se utilizó un modelo de farmacóforo para guiar el cribado, además de ensayos de dinámica molecular para comprobar las interacciones entre los compuestos y el sitio activo de Mpro (Gentile *et al.*, 2020).

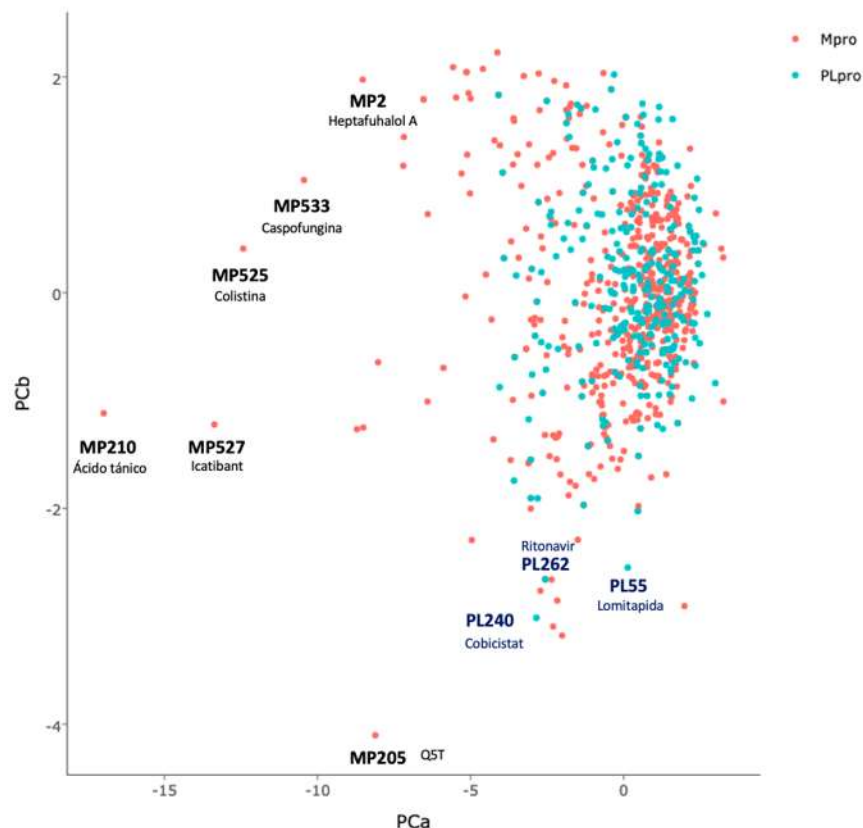


Figura 12. Representación visual del espacio químico para compuestos afines a Mpro y PLpro mediante PCA. Los componentes principales PCa y PCb recuperaron el 86% de la covarianza de los datos. Se observan etiquetados datos atípicos y sus estructuras químicas se muestran en las figuras 13 y 14.

Tabla 3. Contribución de las seis propiedades fisicoquímicas a cada PC.

	PCa	PCb	PCc	PCd	PCe	PCf
MW	-0.410	-0.385	0.380	-0.500	0.445	0.302
TPSA	-0.452	0.136	0.225	0.150	-0.667	0.509
nRotB	-0.324	-0.700	-0.507	0.384	-0.025	-0.021
nHBDon	-0.418	0.322	0.285	0.631	0.467	-0.157
nHBAcc	-0.462	0.018	0.080	-0.312	-0.310	-0.765
ALogP	0.367	-0.490	0.679	0.289	-0.205	-0.198

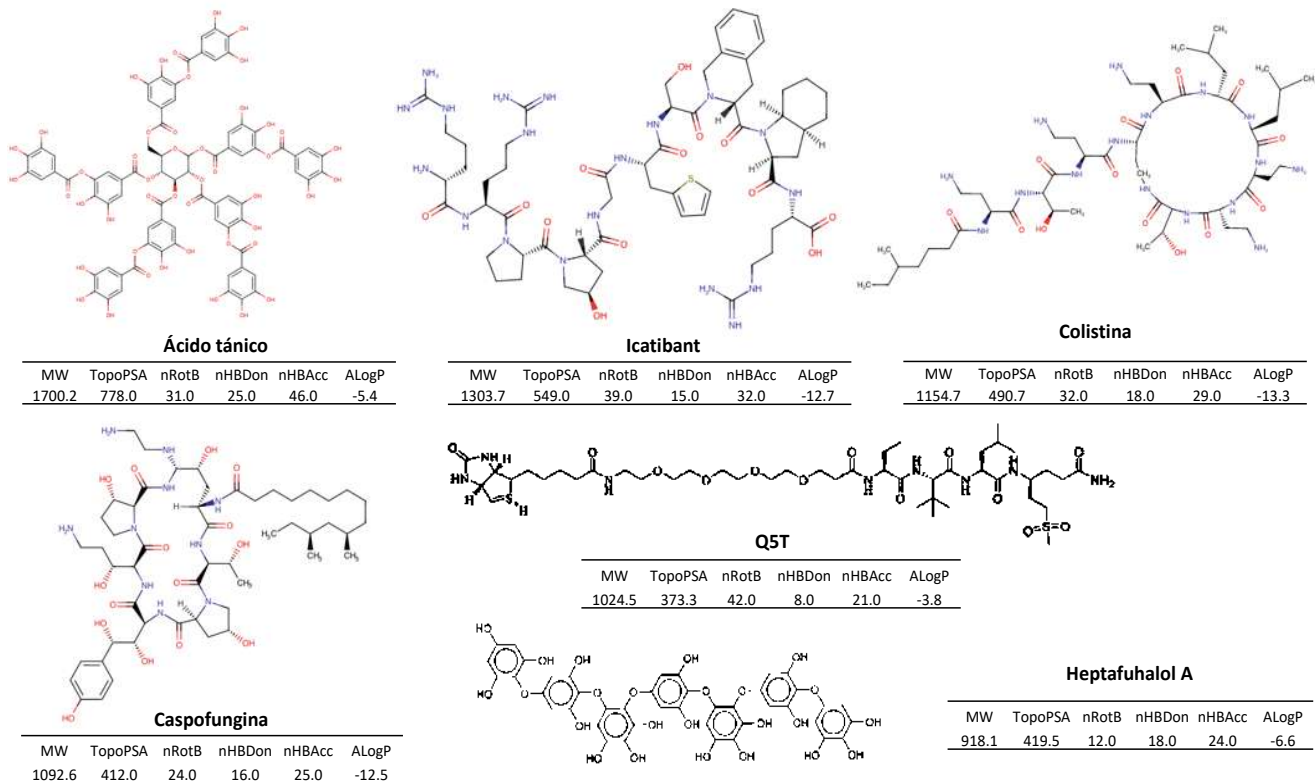


Figura 13. Estructura química de compuestos que representan datos atípicos en el gráfico de espacio químico para compuestos con afinidad por Mpro. Los valores de sus propiedades fisicoquímicas se indican debajo del nombre del compuesto.

Por otra parte, los compuestos afines a la proteasa PLpro parecen estar más próximos entre sí, con compuestos como Cobicistat (PL240), Ritonavir (PL262) y lomitapida (PL55) que aparecen más alejados de la mayoría (Figura 14). Cobicistat y Ritonavir resultaron hits computacionales del cribado virtual de la base de datos “*e-Drug3D*” que contiene fármacos aprobados por la FDA y metabolitos activos, utilizando la plataforma *DockThor*, un servidor web que permite hacer ensayos de acoplamiento molecular y ofrece una variedad de estructuras preparadas que son posibles blancos terapéuticos de SARS-CoV-2 (Guedes *et al.*, 2021). Tanto Cobicistat (MW=775.3) como Ritonavir (MW=720.3) son fármacos utilizados en combinación con otros antivirales en el tratamiento de la infección por VIH. Cobicistat no tiene actividad contra VIH, sino que actúa como un potenciador farmacocinético al inhibir las isoformas 3A del citocromo P450 (CYP3A) y, por lo tanto, aumenta la exposición sistémica de los agentes coadministrados que son metabolizados por enzimas CYP3A. Por otro lado, Ritonavir es un inhibidor de las

proteasas del VIH e interfiere con el ciclo reproductivo del virus. Lomitapida (MW=693.2) resultó un candidato computacional para inhibir a PLpro después de un ensayo de cribado de alto rendimiento de fármacos aprobados por la FDA (Huynh *et al.*, 2021). Es un inhibidor de la proteína de transferencia de triglicéridos (MTP) y se utiliza para reducir el colesterol asociado con la hipercolesterolemia familiar homocigota (HoFH), lo que reduce el riesgo de eventos cardiovasculares como infarto al miocardio y accidente cerebrovascular.

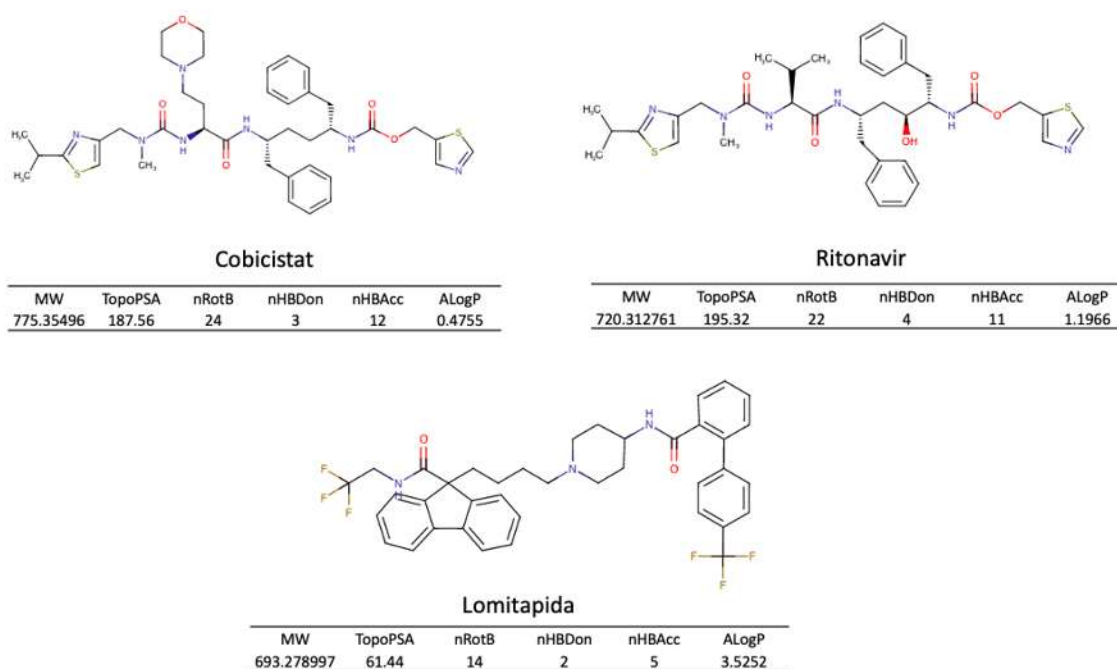


Figura 14. Estructura química de compuestos que representan datos atípicos en el gráfico de espacio químico para compuestos con afinidad por PLpro. Los valores de sus propiedades fisicoquímicas se indican debajo del nombre del compuesto.

10.4 ANÁLISIS DE DIVERSIDAD MOLECULAR BASADO EN COMPARACIONES SISTEMÁTICAS POR PARES

Existen diferentes procedimientos para medir la diversidad química entre grupos de moléculas los cuales dependen de la forma en la que los compuestos sean descritos. Medir la diversidad de un conjunto de compuestos implica el cálculo de la similitud o la disimilitud entre todos los pares de moléculas del conjunto, lo que puede verse en un gráfico de espacio químico como qué tan lejos están los puntos entre sí de acuerdo con un criterio de similitud. Para este análisis se computó la

distancia existente entre pares de compuestos de acuerdo a sus valores de propiedades fisicoquímicas calculadas anteriormente utilizando la métrica de distancia Euclídeana. Esta medida permite comparar la disimilitud que existe internamente en un grupo de compuestos así como la disimilitud existente entre diferentes grupos. Recordando que compuestos cercanos entre sí son de acuerdo a sus propiedades fisicoquímicas similares, entre mayor sea la distancia Euclídeana calculada para dos grupos menor será la similitud que exista entre ellos. Otra métrica utilizada a manera de comprobación es el cálculo del índice de similitud de Tanimoto, el cual mide la similitud existente entre pares de compuestos dando valores de 0 a 1 y que contrario a la métrica de distancia Euclídeana, entre más cercano a 1 sea el valor del índice calculado para dos grupos, mayor será la similitud que exista entre ellos. La figura 15 muestra las matrices de distancia y similitud generadas a partir de la media de las distancias y los índices calculados internamente y entre diferentes conjuntos.

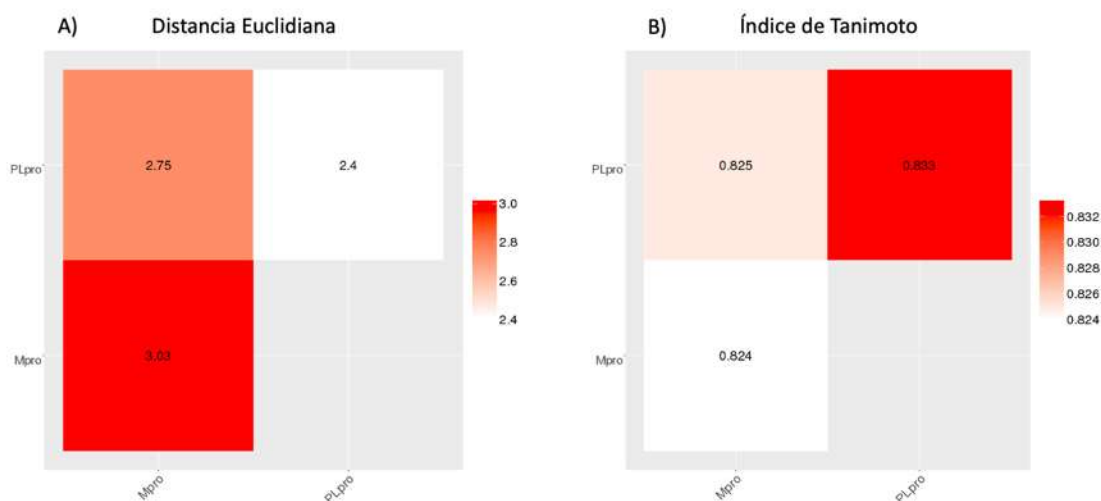


Figura 15. Matrices de (dis)similitud A) Distancia Euclídeana y B) Índice de similitud de Tanimoto basados en el cálculo de seis propiedades fisicoquímicas de interés farmacéutico. Para A) los valores más altos de distancia se muestran en un gradiente de color de rojo (valores altos) a blanco (valores bajos), para B) se utiliza el mismo gradiente de colores que van de rojo (el valor más cercano a 1) a blanco (valores menos cercanos a 1).

De acuerdo a los valores en las matrices de la figura 15, los compuestos pertenecientes al grupo de moléculas con afinidad por Mpro tienen un valor de distancia más alto (3.03) y un valor de índice de similitud más bajo entre ellas (0.824)

que el resto de valores resultantes de las comparaciones entre grupos, lo que sugiere que este grupo de compuestos tiene una mayor diversidad química, resultado que va en acuerdo con el gráfico de espacio químico de la figura 12.

10.5 ANÁLISIS DE LA DIVERSIDAD DE SCAFFOLDS MEDIANTE EL USO DE CURVAS DE RECUPERACIÓN DE SISTEMAS CÍCLICOS (CSR) Y ENTROPÍA DE SHANNON

Como parte del análisis de los grupos de moléculas nos interesó conocer también su diversidad estructural en términos de estructuras núcleo o “*Scaffolds*” y su distribución en las bases de datos. El término *Scaffold* hace referencia a la estructura “*base*” de un compuesto, definido por Bemis y Murcko como todos sus sistemas de anillos y los átomos que los conectan entre sí. Para analizar la distribución de los *Scaffolds*, se generaron curvas de recuperación de sistemas cíclicos (CSR) correspondientes para cada conjunto de datos en la plataforma PUMA. Estas curvas indican la fracción de moléculas del conjunto de datos representada en el eje y, contenida en una fracción dada de sistemas cíclicos o quimiotipos representados en el eje x. Las curvas obtenidas se caracterizaron además cuantitativamente utilizando la fracción de quimiotipos necesarios para recuperar el 50% de los compuestos (F_{50}) y el área bajo la curva (AUC). En este enfoque, cuanto menor sea el área bajo la curva CSR para un conjunto de compuestos, mayor será la diversidad de sus quimiotipos y con respecto al F_{50} , cuanto mayor sea su valor, la diversidad estructural de los compuestos será también más alta. Un conjunto de datos con máxima diversidad contendría un quimiotipo diferente para cada molécula de la quimioteca y la curva CSR sería una diagonal con una AUC de 0,5 (Medina-Franco *et al.*, 2009). En el gráfico 3 se muestran las curvas CSR para los conjuntos de datos y la tabla 4 resume información sobre el número de compuestos (M), número de Scaffolds o de quimiotipos únicos (N), la fracción de quimiotipos únicos relativa al número de moléculas en el conjunto de datos (N/M), el número de quimiotipos que contienen un solo compuesto (NSING), las fracciones de NSING relativas a M y a N (NSING/M y NSING/N,

respectivamente), el área bajo la curva (AUC) y la fracción de quimiotipos necesarios para recuperar el 50% de los compuestos (F_{50}).

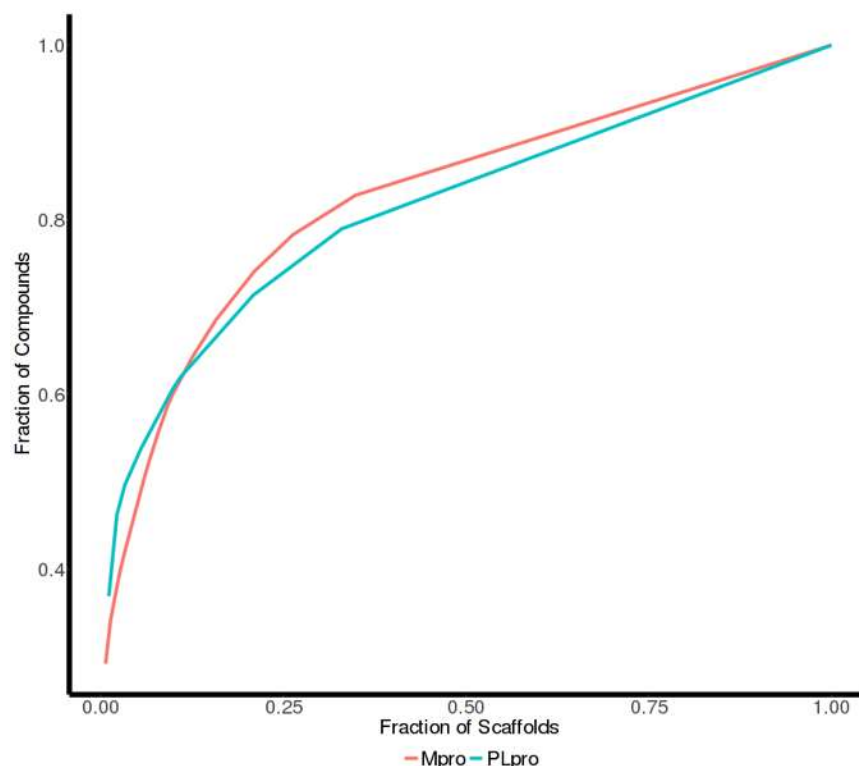


Gráfico 3. Curvas de recuperación de sistemas cíclicos (CSR) para compuestos reportados afines a las proteasas Mpro y PLpro.

Tabla 4. Recuento de Scaffolds, AUC y F_{50} obtenidos para ambos conjuntos de datos. M: número de moléculas; N: número de Scaffolds; NSING: número de Scaffolds con solo un compuesto; AUC: área bajo la curva; F_{50} : Fracción de quimiotipos requerida para recuperar el 50% de las moléculas.

	M	N	N/M	NSING	NSING/M	NSING/N	AUC	F_{50}
Mpro	576	152	0.26	99	0.17	0.65	0.82	0.06
PLpro	290	91	0.31	61	0.21	0.67	0.81	0.03

Según los criterios mencionados, la base de datos de compuestos afines a Mpro debería ser más diversa en quimiotipos que la base de datos de PLpro, de acuerdo al valor de F_{50} (0.06 y 0.03, respectivamente), no obstante, se puede observar en las curvas CSR de la gráfica 3 que aunque la curva naranja correspondiente a los compuestos afines a Mpro alcanza un valor un poco más alto en la fracción de quimiotipos (eje x) al llegar al 0.5 de los compuestos (eje y), esta situación cambia a medida que se aumenta el número de compuestos en el análisis, por lo que al

final el AUC de los compuestos afines a PLpro es más baja que la de Mpro (0.81 y 0.82, respectivamente) sugiriendo una mayor diversidad en quimiotipos únicos. Los valores de N/M, NSING/M y NSING/N van de acuerdo con esta última lectura, aunque en términos generales parece no existir una diferencia en la diversidad de quimiotipos entre los conjuntos de datos.

Además de la generación de curvas CSR se cuantificó la distribución específica de los compuestos en los n quimiotipos únicos más poblados utilizando una métrica de información basada en la entropía (Medina-Franco *et al.*, 2009). Las curvas CSR describen la diversidad de andamios en función de todo el conjunto de datos, mientras que la entropía de Shannon utiliza solo los quimiotipos únicos más frecuentes para tales cálculos (Saldivar-González *et al.*, 2019). Recordando que la entropía escalada de Shannon (SSE) mide la distribución específica de moléculas en un determinado número de quimiotipos, los valores de SSE más cercanos a 1.0 indican que las moléculas están distribuidas más equitativamente en los sistemas cíclicos, por lo tanto, es un indicador de gran diversidad. A su vez, valores más pequeños de SSE indican que la mayoría de las moléculas se distribuyen en pocos sistemas cíclicos, lo que indica una menor diversidad (Medina-Franco *et al.*, 2009). El valor de SSE máximo sucederá solo cuando todos los quimiotipos contengan el mismo número de compuestos o cuando cada quimiotipo contenga un solo compuesto (González-Medina *et al.*, 2017). La tabla 5 resume los valores de SSE para los quimiotipos más poblados (10 a 60 quimiotipos), mientras que las figuras 16 y 17 muestran la distribución de los 10 y 20 quimiotipos más frecuentes para cada base de datos, respectivamente. De acuerdo con los valores de SSE en la tabla 5, el conjunto de moléculas afines a la proteasa Mpro tienen una mayor diversidad en sus quimiotipos únicos, con valores de SSE que van desde 0.976 a 0.962. En los histogramas de la figura 16 se observa que el número de quimiotipos únicos para la base de datos de Mpro se distribuye más uniformemente, comparada con los quimiotipos de compuestos afines a PLpro, esto es aún más notable en la figura 17 donde se muestra el histograma de los 20 quimiotipos más poblados en ambas bases de datos. El quimiotipo más frecuente para la base de Mpro lleva por

identificador ID:43 con 13 compuestos (2.25%), mientras que el quimiotipo único más frecuente para la base de datos de PLpro es ID:249 con 18 moléculas (6.20%). Las estructuras químicas, el número de identificador y la frecuencia de los quimiotipos únicos más frecuentes se ilustra en las figuras 18 y 19, donde además se indica que para ambas bases de datos los compuestos acíclicos tuvieron lugar dentro de los diez quimiotipos más poblados en ambos sets con 6 compuestos para cada enzima respectivamente. Dentro de los quimiotipos más frecuentes de la base de datos Mpro se encuentran 5 scaffolds con átomos de azufre (ID: 43, 174, 51, 32 y 137) siendo muy similares ID43 Y ID51 con un átomo de azufre de diferencia. Por otra parte las estructuras químicas de los quimiotipos para PLpro revelan dos estructuras muy similares (ID 255 y 258) donde la diferencia parece ser la orientación del par de anillos de naftaleno. Entre ambos sets de quimiotipos se comparte como quimiotipo frecuente, el anillo de benceno (ID:22), con cinco y quince compuestos para Mpro y PLpro respectivamente y se observa también una gran similitud entre los quimiotipos ID287 de Mpro con ID11 de PLpro con la orientación espacial de diferencia.

Tabla 5. Valores de SSE de los 10-60 quimiotipos únicos más poblados.

Data set	SSE10	SSE20	SSE30	SSE40	SSE50	SSE60
Mpro	0.976	0.976	0.971	0.968	0.963	0.962
PLpro	0.939	0.904	0.903	0.890	0.884	0.884

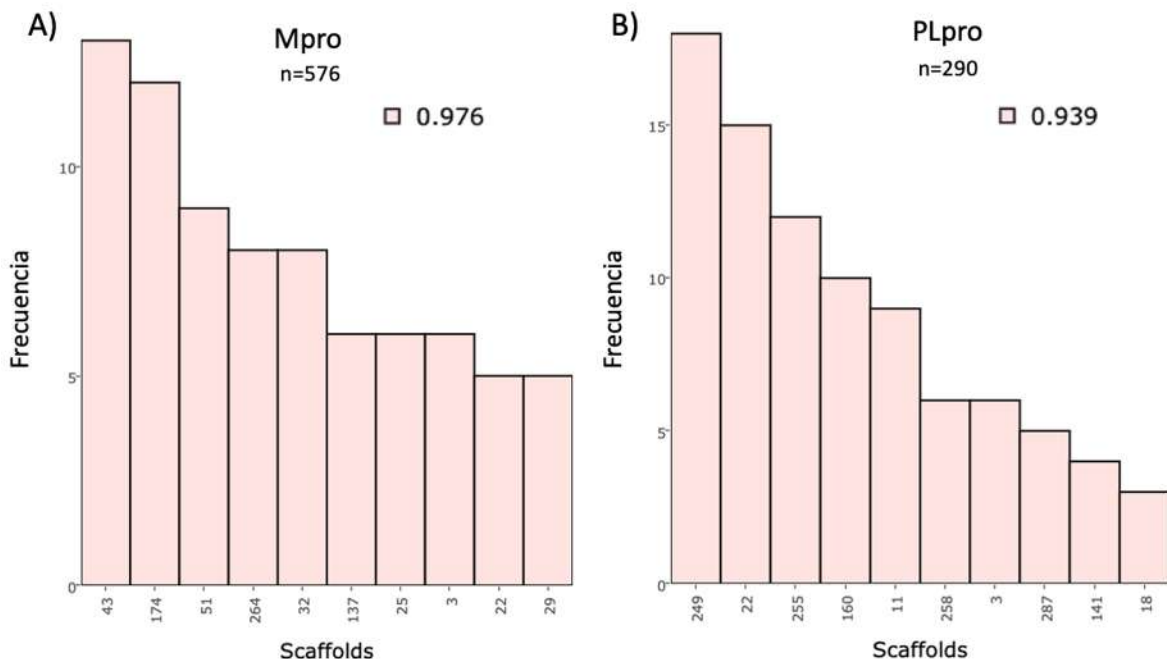


Figura 16. Distribución de las diez estructuras de scaffolds más frecuentes para A) compuestos con afinidad por Mpro y B) compuestos con afinidad por PLpro. El valor de SSE10 para los diez scaffolds más frecuentes se encuentra de lado del recuadro rojo. Los números del eje X representan un ID para cada estructura de scaffold.

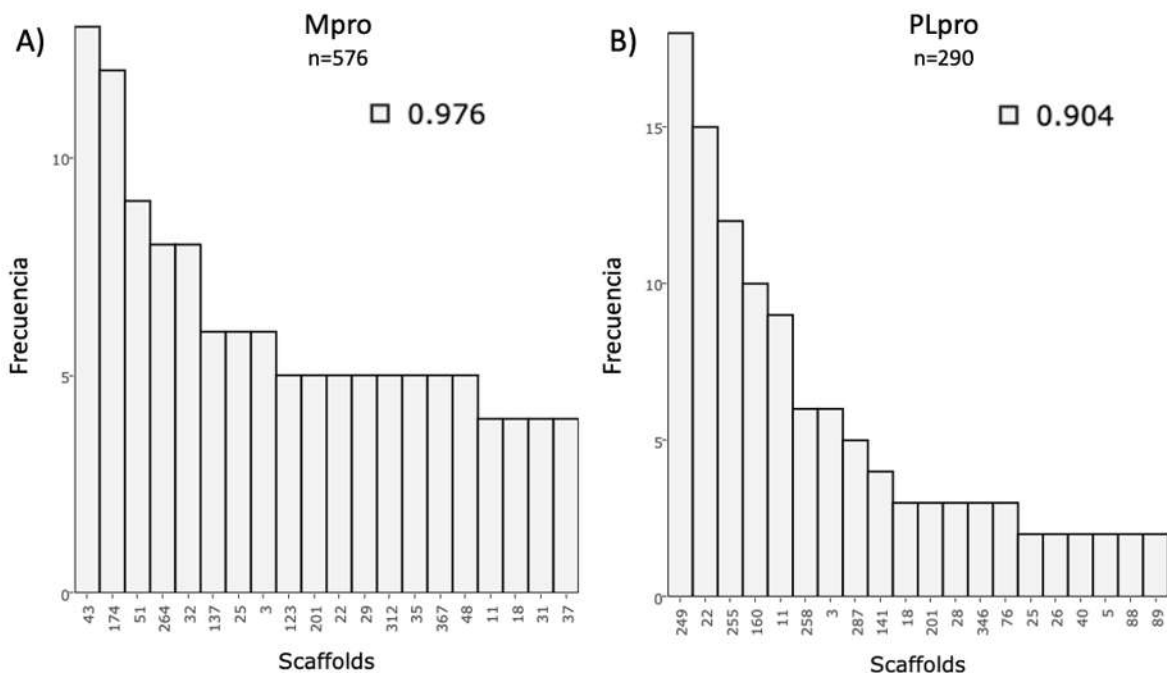


Figura 17. Distribución de las veinte estructuras de scaffolds más frecuentes para A) compuestos con afinidad por Mpro y B) compuestos con afinidad por PLpro. El valor de SSE20 para los veinte scaffolds más frecuentes se encuentra de lado del recuadro verde. Los números del eje X representan un ID para cada estructura de Scaffold.

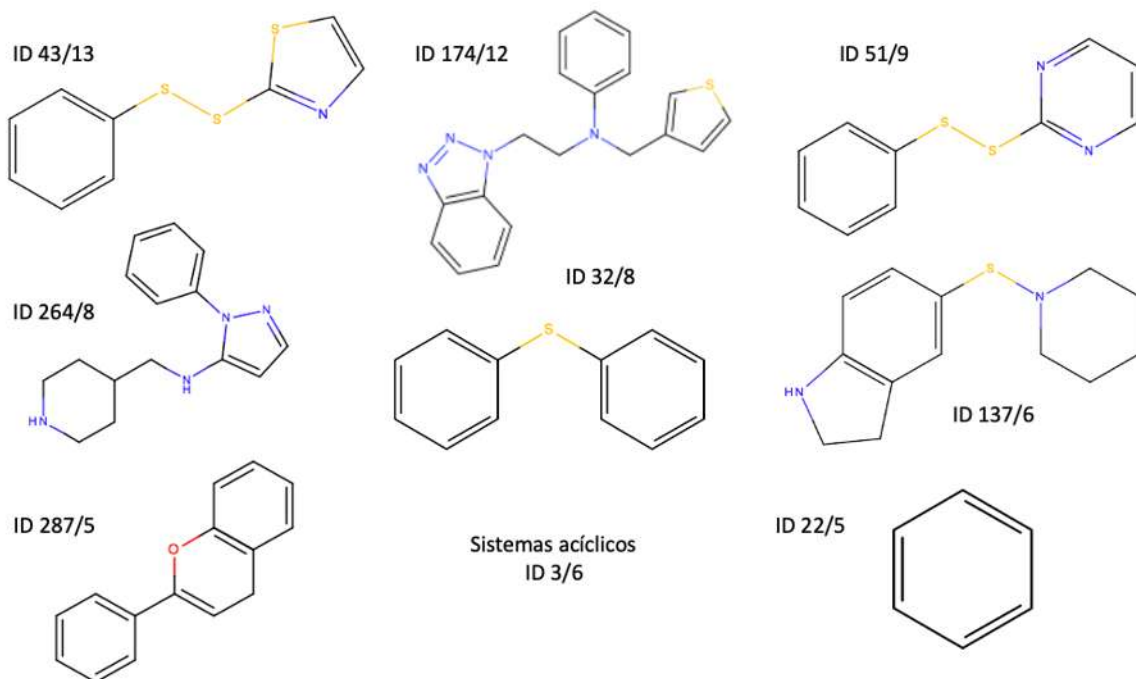


Figura 18. *Scaffolds* o quimiotipos más frecuentes en la quimioteca de compuestos reportados con afinidad por Mpro (n=576).

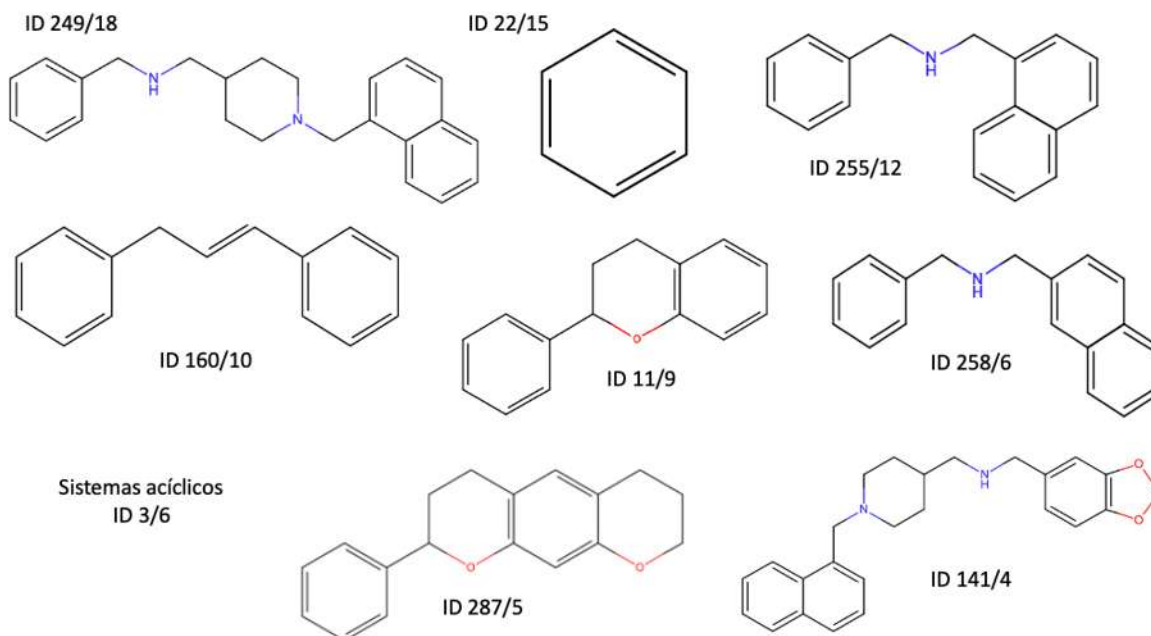


Figura 19. *Scaffolds* o quimiotipos más frecuentes en la quimioteca de compuestos reportados con afinidad por PLpro (n=290).

10.6 DIVERSIDAD ESTRUCTURAL DE MOLÉCULAS COMPLETAS UTILIZANDO HUELLAS DACTILARES MOLECULARES (STRUCTURAL MOLECULAR FINGERPRINTS) E ÍNDICE DE SIMILITUD DE TANIMOTO

Para comparar la diversidad estructural de las moléculas enteras, se generaron en la plataforma PUMA, curvas de función de distribución acumulativa (CDF) basadas en el análisis de tres huellas dactilares moleculares "*Molecular Fingerprints*", utilizando el índice de similitud de Tanimoto para medir la relación del conjunto de características estructurales que tienen en común dos compuestos con respecto a las características estructurales totales de ambos compuestos. En quimioinformática las huellas dactilares moleculares son una de las representaciones más comunes para compuestos químicos. Estas representaciones se observan como cadenas lineares binarias que simplifican la información química de acuerdo a la ausencia o presencia de fragmentos estructurales predeterminados (bits) en un compuesto, donde cero (0) indica ausencia y uno (1) presencia del fragmento en la estructura química del compuesto (Fernández-De Gortari *et al.*, 2017). En este estudio se utilizaron los fingerprints "*Molecular ACCess System (MACCS) keys (166-Bits)*" "*Pubchem Fingerprint (881-Bits)*" y "*Extended Conectivity Fingerprint con un diametro de 4 (ECFP_4)*" el cual a diferencia con los primeros dos, toma en consideración el ambiente molecular que rodea a cada átomo pesado, revisando de manera circular la periferia del átomo de referencia y dando un identificador para dicho átomo de manera actualizada después de cada ciclo de revisión. Los resultados en forma de curvas CDF, muestran la similitud molecular que existe entre los conjuntos de moléculas tomando en cuenta su estructura completa. La figura 20 ilustra que ambos grupos de moléculas son muy similares y al utilizar el fingerprint MACCS no existe una diferencia en el desplazamiento de las curvas. Respecto a los otros dos fingerprints (Pubchem y ECFP_4), las curvas para los compuestos afines a PLpro están ligeramente desplazados a valores más cercanos a 1 en el eje x, lo que indica mayor similitud o bien, menor diversidad entre los compuestos de ese grupo. Estos resultados concuerdan con los obtenidos en el análisis de diversidad de quimiotipos

únicos, por lo que se puede concluir que los compuestos afines a Mpro y a PLpro son altamente similares entre si.

Tabla 6. Distribución estadística de similitud por pares utilizando las huellas dactilares moleculares MACCS, Pubchem y ECFP_4 y el índice de similitud de Tanimoto.

Fingerprint	Data set	Min	1er Qu	Mediana	Media	3er Qu	Max	Desv. Est.
MACCS (161-Bits)	Mpro	0.00	0.27	0.36	0.37	0.46	1.00	0.14
	PLpro	0.00	0.28	0.36	0.38	0.46	1.00	0.14
Pubchem (886-Bits)	Mpro	0.02	0.29	0.40	0.40	0.50	1.00	0.15
	PLpro	0.03	0.32	0.41	0.42	0.51	1.00	0.16
ECFP_4	Mpro	0.00	0.07	0.09	0.10	0.11	1.00	0.05
	PLpro	0.00	0.08	0.10	0.11	0.12	0.83	0.07

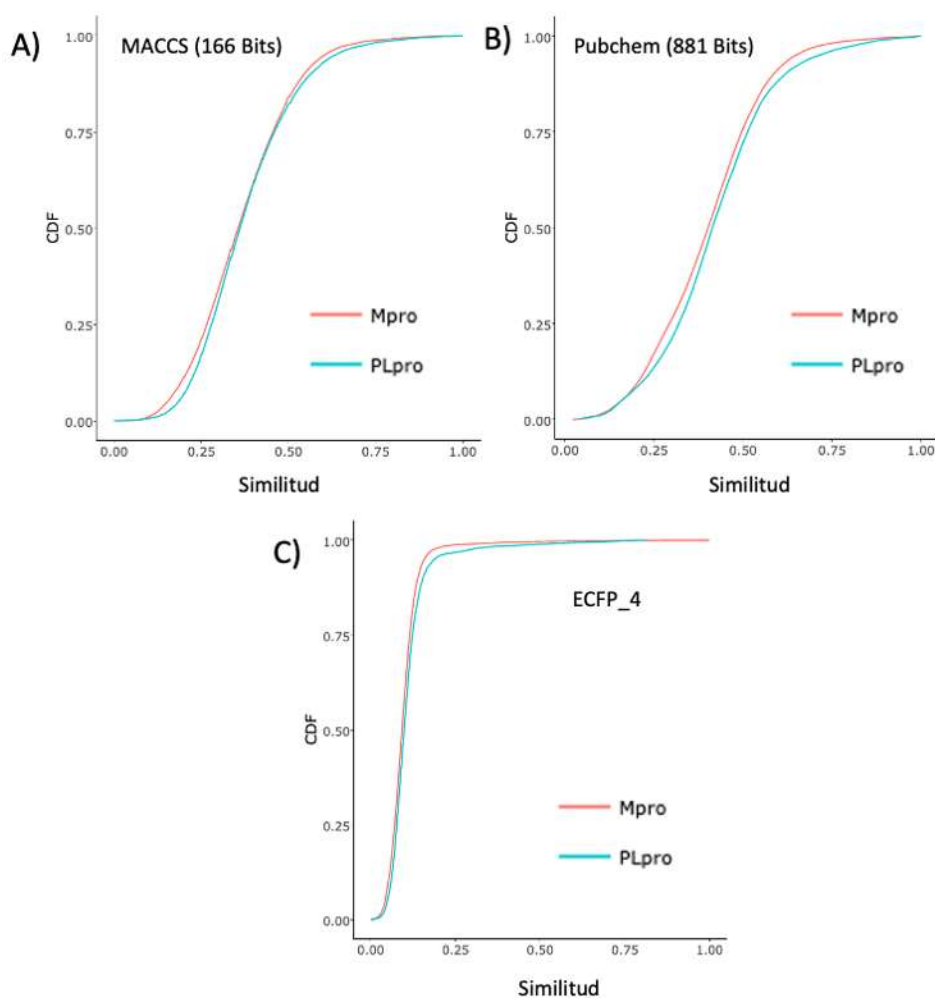


Figura 20. Curvas de función de distribución acumulativa (CDF) para la similitud por pares utilizando huellas dactilares moleculares (*Fingerprints*), MACCS (166-Bits), Pubchem (881-Bits) y ECFP_4.

10.7 VALIDACIÓN DEL PROTOCOLO PARA ENSAYOS DE ACOPLAMIENTO MOLECULAR CON GLIDE DE SCHRÖDINGER

Los ensayos de acoplamiento molecular permiten simular la interacción molecular y predecir el modo de unión y la afinidad entre receptores y ligandos. Esto puede proporcionar información importante en forma de un puntaje de acoplamiento o “score” sobre la posible interacción entre una enzima y una gran cantidad de compuestos en relativamente poco tiempo. Además, es posible observar las interacciones entre los compuestos y el blanco biológico, analizando diferentes poses o conformaciones que puede adoptar el ligando en la cavidad seleccionada y, junto con el puntaje servir como un filtro para la selección de compuestos con mayor probabilidad de interacción.

Con la finalidad de validar un protocolo computacional para el acoplamiento de los compuestos encontrados con la enzima correspondiente (p.ej. compuestos reportados con afinidad a Mpro), se seleccionaron de la literatura artículos de alta relevancia en la caracterización de la estructura de las enzimas por métodos de cristalografía por difracción de rayos X, así como en la interacción de estas enzimas con compuestos reportados como buenos candidatos para su inhibición. El proceso de validación comenzó con la descarga de dichas estructuras cristalográficas en complejo con sus inhibidores de “*Protein Data Bank*” (PDB) y, posteriormente se ajustaron en el software *Glide* de Schrödinger, Inc. las condiciones para replicar los resultados de acoplamiento reportadas por los autores correspondientes. Se prepararon las estructuras proteicas (Mpro y PLpro) removiendo el agua y los remanentes de solvente que se usan para la cristalización, y se separó el ligando de la proteína, dejando la proteína libre para acoplarse con el mismo ligando utilizando el protocolo de acoplamiento diseñado. Con este enfoque lo que se pretende es que al reacomodar *in silico* el ligando con la proteína, el protocolo establecido replique los resultados experimentales (cristalográficos) publicados por los autores, generando poses altamente similares a la del ligando original. Los acoplamientos de validación se corrieron por triplicado para asegurar la reproducibilidad del protocolo y se calcularon los valores de RMSD para las mejores poses, obteniendo los resultados siguientes.

10.8 VALIDACIÓN PARA LA PROTEASA PRINCIPAL Mpro EN COMPLEJO CON MASITINIB

Para la validación del protocolo de acoplamiento de la proteasa principal Mpro con Masitinib (PubChem CID: 10074640), se utilizó como referencia el cristal de Mpro con PDB ID: 7JU7 correspondiente al artículo publicado por Drayman *et al.* (2021) “*Masitinib is a broad coronavirus 3CL inhibitor that blocks replication of SARS-CoV-2*” donde los autores cribaron de manera virtual una quimioteca de 1900 medicamentos seguros contra OC43, un beta coronavirus humano que causa resfriado común, y evaluaron a los mejores candidatos contra SARS-CoV-2. Masitinib resultó el inhibidor más potente de acuerdo con las pruebas bioquímicas y de cristalografía por rayos X con un valor de $IC_{50}=2.5 \mu\text{M}$ y un mecanismo de inhibición competitiva. Ratones infectados con SARS-CoV-2 y luego tratados con masitinib mostraron una reducción >200 veces en los títulos virales en los pulmones y nariz, así como una reducción de la inflamación pulmonar. Estos autores publicaron para uso libre de la comunidad científica el cristal 7JU7 correspondiente a la estructura de la proteasa principal Mpro en complejo con Masitinib a una resolución de 1.60 Å, la cual se utilizó como punto de partida para la validación del protocolo del presente estudio.

Después de preparada y separada la proteína del ligando co-cristalizado (Masitinib), se realizó el reacoplamiento molecular de Masitinib en el sitio activo de Mpro. Los resultados mostraron una excelente superposición espacial del ligando acoplado computacionalmente con respecto a las coordenadas de los átomos del complejo Mpro-Masitinib reportado experimentalmente. Los cálculos de RMSD para las tres mejores conformaciones se muestran en la tabla 7 y la mejor pose con un valor de 0.6640 Å se muestra en la Fig. 20. Además, se analizaron las interacciones intermoleculares de la mejor pose de Masitinib reacoplado con la proteína y se observó que se establecieron las interacciones más importantes, por ejemplo, las interacciones con los residuos His41 y Cys145 se conservaron (Fig. 21). Estos resultados sugieren que el protocolo de acoplamiento funciona y puede utilizarse para acoplar, bajo las mismas condiciones, los compuestos encontrados en la

literatura y contenidos en la quimioteca para predecir su afinidad con el sitio activo de Mpro.

Tabla 7. Valores de RMSD en Å para los mejores tres cálculos de acoplamiento al replicar la orientación experimentalmente co-cristalizada de Masitinib con Mpro.

Compuesto	Mejor acoplamiento	2do mejor acoplamiento	3er mejor acoplamiento
Masitinib	RMSD= 0.6640 Å	RMSD= 1.223 Å	RMSD= 1.4147 Å

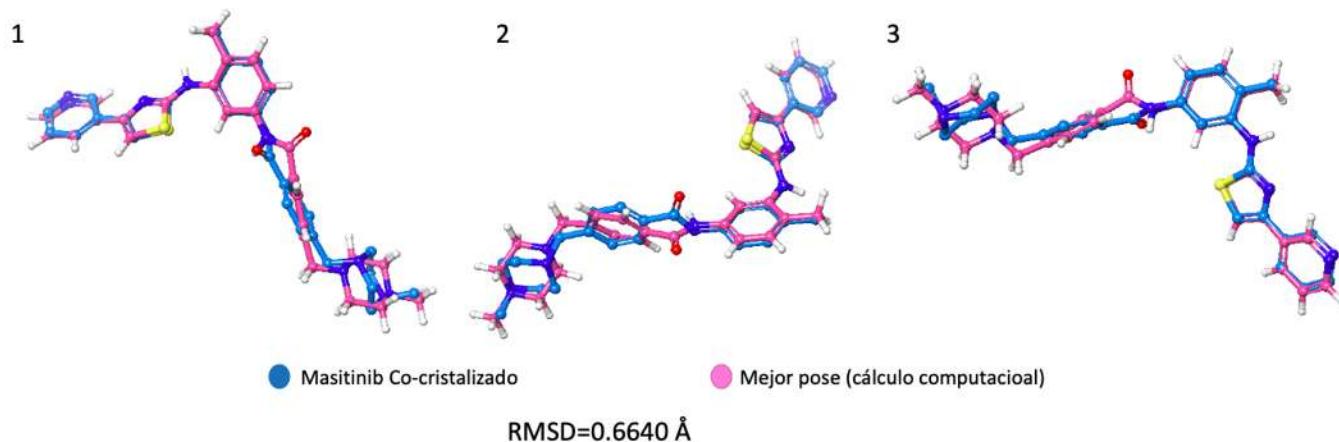


Figura 21. Tres visualizaciones diferentes del mejor cálculo de acoplamiento al reproducir la orientación de Masitinib con Mpro reportada experimentalmente (PDB ID: 7JU7). En azul se muestra Masitinib en su orientación original co-cristalizado y en rosa el mejor cálculo de acoplamiento con un RMSD=0.6640 Å.

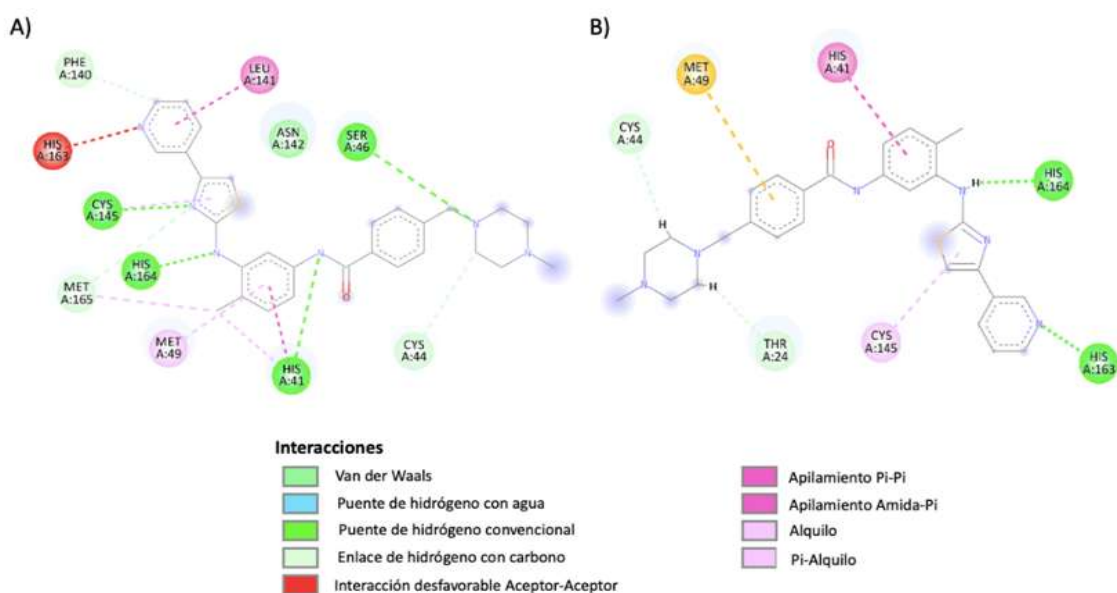


Figura 22. Reproducción de las interacciones intermoleculares más relevantes entre Masitinib y Mpro. A) Masitinib en orientación original co-cristalizado con Mpro (PDB ID: 7JU7), B) Mejor cálculo de acoplamiento Mpro-Masitinib con Glide de Schrödinger.

10.9 VALIDACIÓN PARA LA PROTEASA DE TIPO PAPAÍNA PL_{pro} EN COMPLEJO CON EL INHIBIDOR GRL-0617

Para el caso de la proteasa de tipo papaína PL_{pro} se utilizó la estructura cristalográfica PDB ID: 7JRN, correspondiente a la estructura de PL_{pro} en complejo con el inhibidor GRL-0617, a una resolución de 2.48 Å. Entre los inhibidores de PL_{pro} de SARS-CoV y SARS-CoV-2, GRL-0617 es uno de los compuestos más potentes con un IC₅₀= 2.1 μM (Fu *et al.*, 2021). Esta reportado que GRL-0617 actúa mediante un mecanismo de inhibición competitiva con la proteína del producto génico estimulado por interferón 15 (ISG15). Las modificaciones por ISG15 (ISGilación) consisten en la conjugación covalente de la proteína ISG15 (MW=17.1 KDa) a sustratos proteicos y este proceso inhibe la replicación viral (Perng & Lenschow, 2018; Hermann & Bogunovic, 2017). PL_{pro} es capaz de revertir este proceso de ISGilación, evitando la inhibición de la replicación del virus. El mecanismo de inhibición de PL_{pro} mediado por GRL-0617 sugiere la competencia entre el compuesto con ISG15 por el sitio de unión en PL_{pro}. Esto permite el bloqueo de la unión de ISG15 con la enzima, permitiendo el proceso de ISGilación y la inhibición de la replicación viral. GRL-0617 forma interacciones importantes con los aminoácidos Tyr268, Gln269, Asp164 y Tyr264 dentro del bolsillo de unión, el cual ha sido propuesto como un sitio crítico para el descubrimiento de fármacos antivirales y se ha utilizado como sitio de unión en el presente estudio. De manera similar que la validación de Masitinib con M_{pro}, se preparó la proteína y el ligando GRL-0617 para reacomplarlo con *Glide* de Schrödinger, Inc. a la proteasa PL_{pro}. La tabla 8 muestra los valores de RMSD para los mejores tres cálculos de acoplamiento y la figura 23 ilustra, desde tres ángulos diferentes, la mejor pose de validación con un RMSD=0.3690 Å, donde se logra observar la superposición entre el ligando naranja (cálculo computacional) y el ligando verde (GRL-0617 co-cristalizado). La figura 24 muestra también que las interacciones intermoleculares entre GRL-0617 con PL_{pro} se conservan en el mejor cálculo de acoplamiento, manteniendo ocho de las nueve interacciones resueltas en el PDB original 7JRN.

Tabla 8. Valores de RMSD en Å para los mejores tres cálculos de acoplamiento al replicar la orientación experimentalmente co-cristalizada de GRL-0617 con PLpro

Compuesto	Mejor acoplamiento	2do mejor acoplamiento	3er mejor acoplamiento
GRL-0617	RMSD= 0.3690 Å	RMSD= 0.7123 Å	RMSD= 0.7890 Å

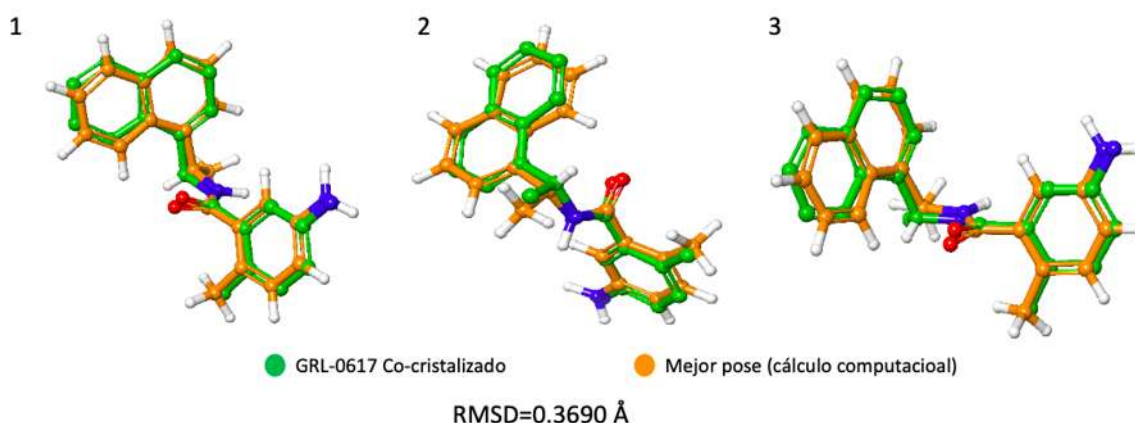


Figura 23. Tres visualizaciones diferentes del mejor cálculo de acoplamiento al reproducir la orientación de GRL-0617 con PLpro reportada experimentalmente (PDB:7JRN). En verde se muestra GRL-0617 en su orientación original co-cristalizado y en naranja el mejor cálculo de acoplamiento con un RMSD=0.3690 Å.

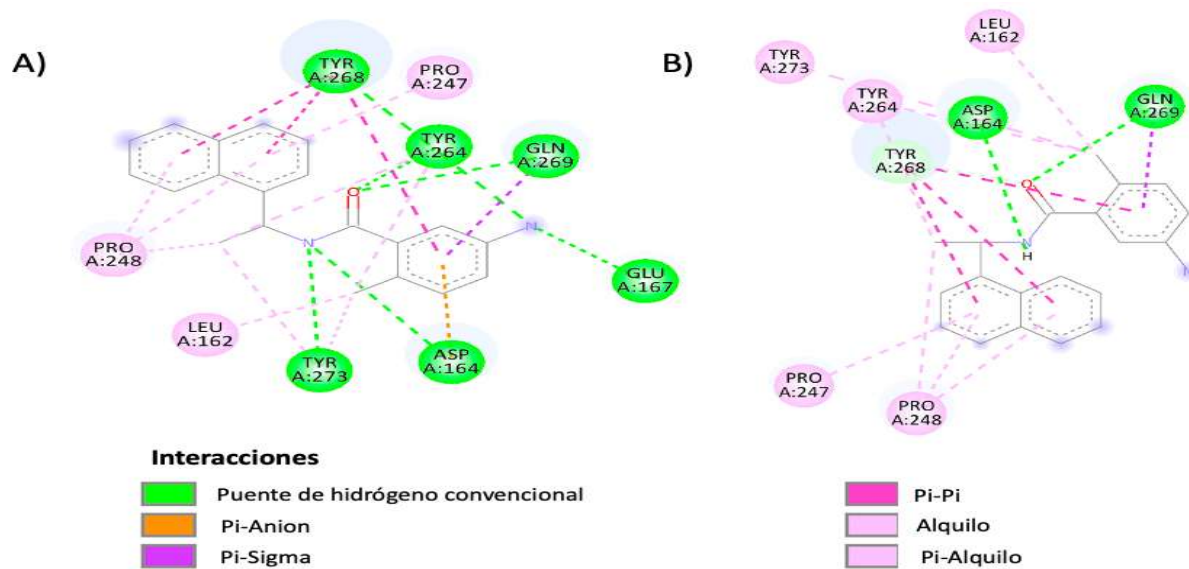


Figura 24. Reproducción de las interacciones intermoleculares más relevantes entre GRL-0617 y PLpro. A) GRL-0617 con orientación original co-cristalizado (PDB ID: 7JRN), B) Mejor cálculo de acoplamiento con Glide de Schrödinger.

10.10 ANÁLISIS DE PROPIEDADES FÍSICOQUÍMICAS DE INTERÉS FARMACÉUTICO DE LOS COMPUESTOS REPORTADOS CON AFINIDAD POR LA PROTEASA PRINCIPAL MPRO

Con el objetivo de hacer un análisis extensivo sobre la diversidad de los compuestos dentro de los diferentes grupos se hizo una búsqueda de valores IC_{50} , medida que indica la concentración a la que un fármaco es capaz de inhibir un proceso biológico concreto en un 50 % para aquellos compuestos con dicho valor reportado. Mediante la revisión en la literatura se diferenciaron en subgrupos aquellos compuestos que tienen su actividad reportada con un IC_{50} medido con la abreviatura “ACT”, compuestos que presentan afinidad por la proteasa Mpro “Mpro” y fármacos aprobados que han sido mencionados en estudios de reposicionamiento de fármacos “RepMP”. Este análisis permitirá comparar dentro del mismo grupo de compuestos afines a Mpro, aquellos que tienen actividad contra aquellos que no la tienen reportada aún. Además, se incluyen en el análisis dos compuestos que sirven como referencia debido a su gran importancia en la búsqueda de inhibidores de la enzima, el compuesto Masitinib con un IC_{50} reportado de 2.5 μ M y referente de la interacción con Mpro en su sitio activo y el compuesto Paxlovid, el único fármaco aprobado por la FDA hasta ahora como inhibidor de Mpro que se administra por vía oral. En la tabla 10 se encuentra el resumen de resultados graficados en las figuras 25 y 26, recordando que “Data set” hace referencia al subconjunto al que pertenecen las moléculas, “Min” indica el valor mínimo, “Max” indica el valor máximo y “1er Qu” y “3er Qu” corresponden a los valores del primer y tercer cuartil, respectivamente. Al realizar las pruebas de normalidad para las diferentes propiedades no se encontró ninguna que siguiera una distribución normal para ninguno de los tres subgrupos, por lo que el test de rangos de Wilcoxon arrojó los resultados siguientes respectivo a las diferencias entre grupos. Al comparar los compuestos con afinidad por Mpro “Mpro” contra aquellos que tienen su actividad reportada “ACT” se encontró una diferencia estadísticamente significativa en todos los parámetros excepto el número de enlaces rotables nRotB. Esto puede observarse en los histogramas de la figura 25 como un desplazamiento a valores más altos en los parámetros o bien en los gráficos de cajas de la figura 26, donde los compuestos

con ACT tienen un MW menor con un promedio de 384.75 g/mol y por lo tanto valores también de TPSA, nHBAcc y nHBDon menores, al ser compuestos más pequeños y con valores para reglas de Lipinski aceptables. En promedio son compuestos que tienen una mejor solubilidad, lo que puede observarse en el parámetro AlogP más alto con respecto a los otros dos grupos. Al comparar los compuestos con afinidad a Mpro contra los fármacos reposicionados las diferencias disminuyen, siendo diferentes únicamente en tres de las seis propiedades, las cuales son MW, nRotB y nHBAcc, mientras que para TPSA, HBDon y ALogP no existieron diferencias significativas. Al comparar el grupo de moléculas con actividad reportada "ACT" contra los fármacos reposicionados "RepMP" se encontró diferencia en todas las propiedades, siendo los compuestos "ACT" aquellos con valores menores de MW, TPSA, nRotB, nHBDon, nHBAcc y mayores en ALogP. Masitinib y Paxlovid se encuentran en valores intermedios, donde los compuestos con afinidad por Mpro "Mpro" tienen la mayoría de valores para sus propiedades fisicoquímicas. Los fármacos reposicionados presentan la ventaja de haber pasado por pruebas clínicas avanzadas y tener un perfil farmacéutico favorable por lo que el tiempo de desarrollo puede potencialmente acortarse; por otro lado, los compuestos pequeños del grupo "ACT" parecen tener propiedades que les aportan buena solubilidad e interacción con el blanco biológico. El grupo de moléculas afines "Mpro", al ser un grupo intermedio, presenta un excelente espacio de búsqueda de compuestos con mejor afinidad y actividad contra la proteasa principal Mpro.

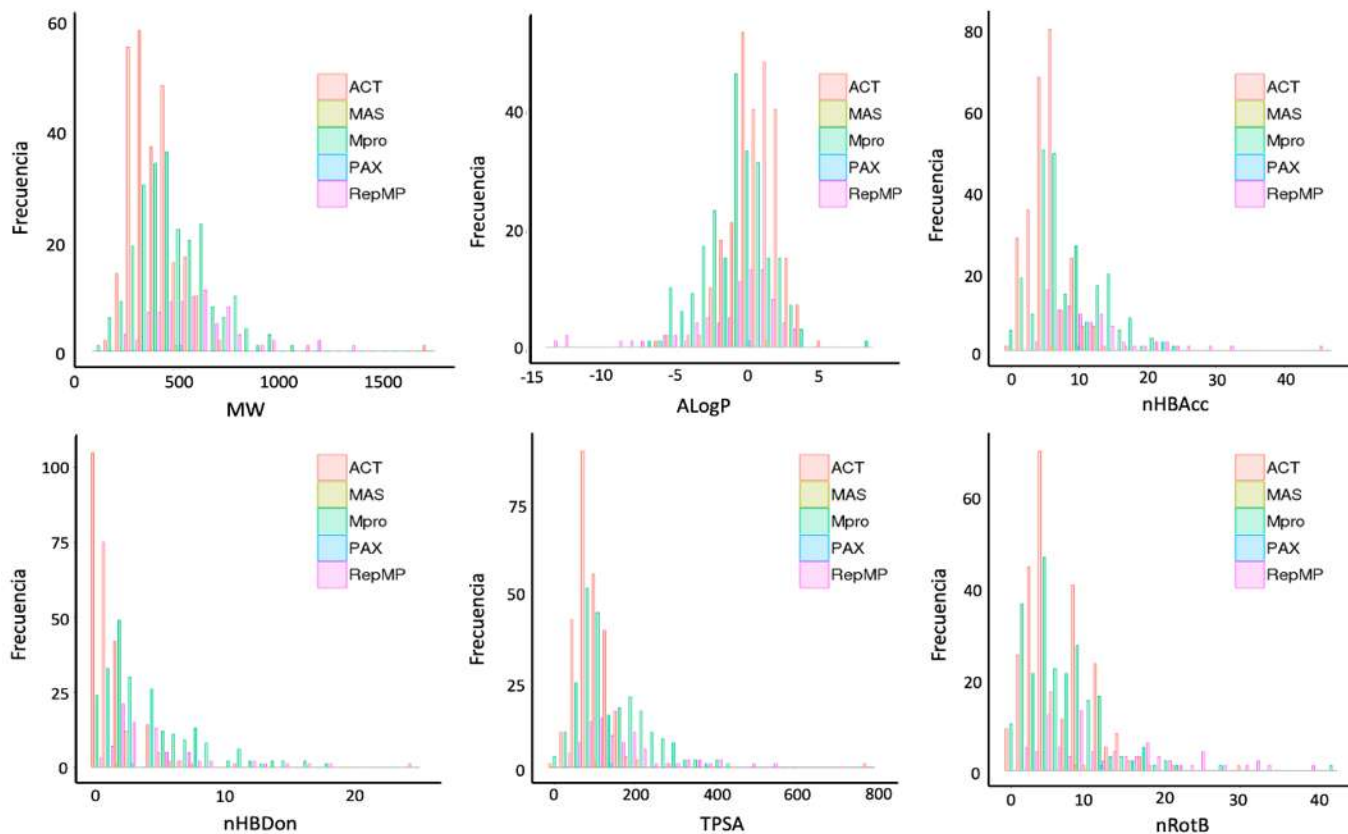


Figura 25. Histograma de seis propiedades fisicoquímicas de compuestos encontrados en la literatura con afinidad por la proteasa Mpro. División en subgrupos “ACT” (actividad biológica reportada), “Mpro” (Afinidad probada por métodos computacionales) y “RepMp” (fármacos reposicionados).

Tabla 9. Propiedades fisicoquímicas de Masitinib y Paxlovid.

Molécula	MW	ALogP	nHBAcc	nHBDdon	TPSA	nRotB
Masitinib	498.22	1.18	7.00	2.00	97.63	8.00
Paxlovid	499.24	0.10	9.00	3.00	131.40	11.00

Tabla 10. Distribución estadística de las propiedades MW, AlogP, nHBAcc, nHBDon, nRotB y TPSA para compuestos con afinidad por Mpro divididos en subgrupos “Mpro”, “ACT” y “RepMP”.

Propiedad	Data set	Min	1er Qu	Mediana	Media	3er Qu	Max	Desv. Est
MW	Mpro	102.03	354.15	444.20	466.76	578.14	1024.50	167.94
	ACT	188.04	300.78	359.98	384.75	446.36	1700.17	130.91
	RepMP	203.09	435.07	543.17	564.65	659.38	1303.66	209.51
AlogP	Mpro	-6.60	-2.16	-0.48	-0.71	0.73	8.62	2.19
	ACT	-6.02	-0.16	0.76	0.69	1.87	5.28	1.67
	RepMP	-13.32	-2.56	-0.37	-1.28	0.86	3.27	3.40
nHBAcc	Mpro	0.00	5.00	7.00	8.15	11.00	24.00	4.90
	ACT	0.00	4.00	5.00	5.72	7.00	46.00	3.71
	RepMP	3.00	6.00	9.00	10.47	12.00	32.00	6.02
nHBDon	Mpro	0.00	2.00	3.00	3.95	6.00	18.00	3.57
	ACT	0.00	0.00	1.00	1.33	2.00	25.00	2.17
	RepMP	0.00	2.00	3.00	4.17	5.00	18.00	3.57
TPSA	Mpro	0.00	76.02	110.49	135.88	186.37	419.52	81.55
	ACT	0.00	71.00	91.03	98.61	117.64	777.98	56.08
	RepMP	35.91	91.34	133.60	159.08	195.32	549.02	104.08
nRotB	Mpro	0.00	3.00	6.00	6.48	9.00	42.00	5.11
	ACT	0.00	3.00	5.00	6.23	9.00	31.00	4.34
	RepMP	1.00	5.00	9.00	12.04	17.00	39.00	8.58

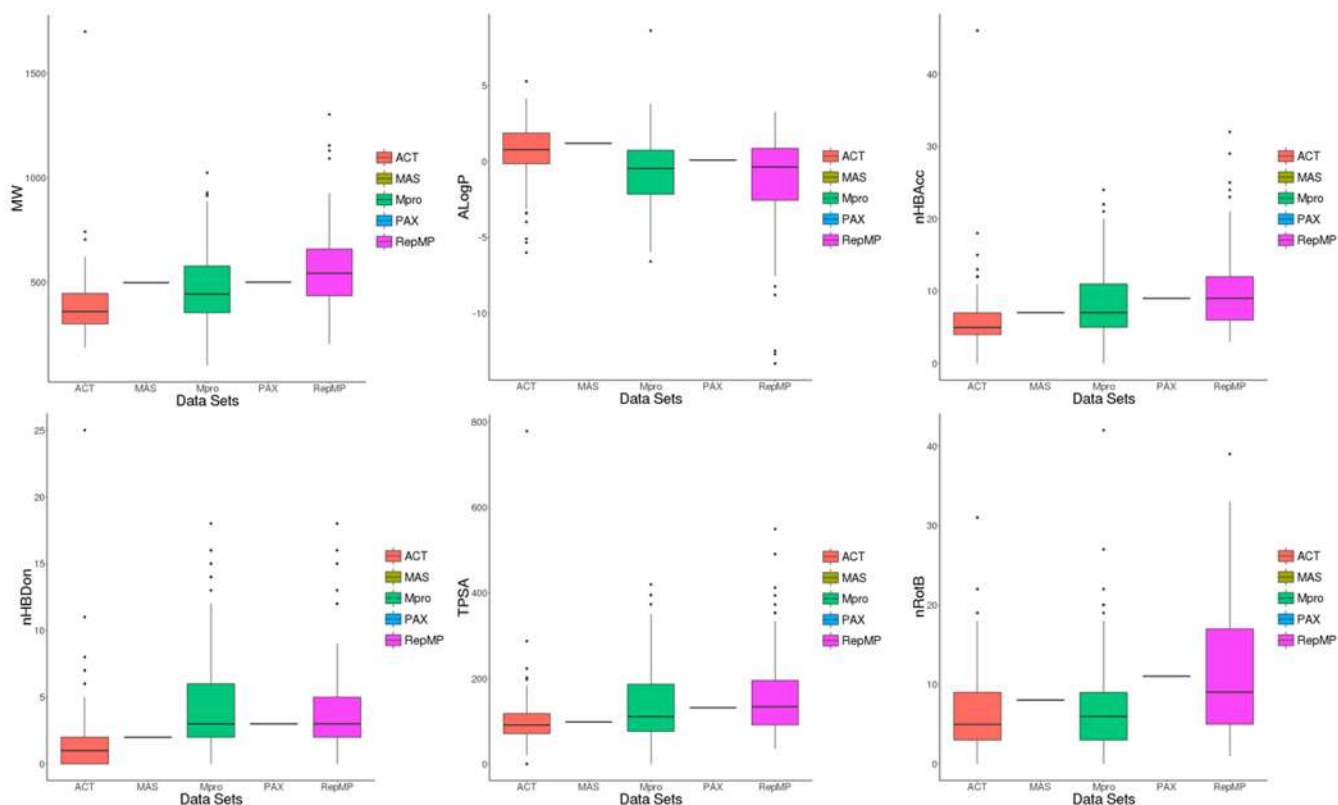


Figura 26. Distribución por gráficos de cajas para las seis propiedades de interés farmacéutico de compuestos con afinidad por Mpro divididos en subgrupos “ACT”, “Mpro” y “RepMP”.

10.11 ANÁLISIS DE DIVERSIDAD MOLECULAR BASADO EN COMPARACIONES SISTEMÁTICAS POR PARES

Los valores en las matrices de la figura 27 indican valores de distancia euclidiana mayores entre los fármacos reposicionados y los otros grupos de compuestos “ACT” y “Mpro” (3.37 y 3.48 respectivamente), pero además presentan un valor de distancia mayor respectivo a su propio grupo (3.68) lo que indica una gran diversidad química o una distancia más grande entre los puntos en el espacio químico representativo de compuestos fármacos reposicionados. El grupo con menor distancia Euclidiana entre sus compuestos es el de compuestos con actividad biológica “ACT”, lo que puede sugerir que los compuestos tienen en general estructuras con propiedades similares que pueden ser importantes en la actividad biológica. Los resultados, de acuerdo con el índice de similitud de Tanimoto, confirman lo anterior con un valor de 0.88 para la comparación de los compuestos “ACT” dentro de su mismo grupo. Las comparaciones entre Paxlovid y Masitinib muestran valores de distancia euclidiana pequeños al compararse entre ellos (0.965) y de índice de similitud de Tanimoto (1) debido a que tienen estructuras y propiedades fisicoquímicas muy similares, las cuales se resumen en la tabla 9.

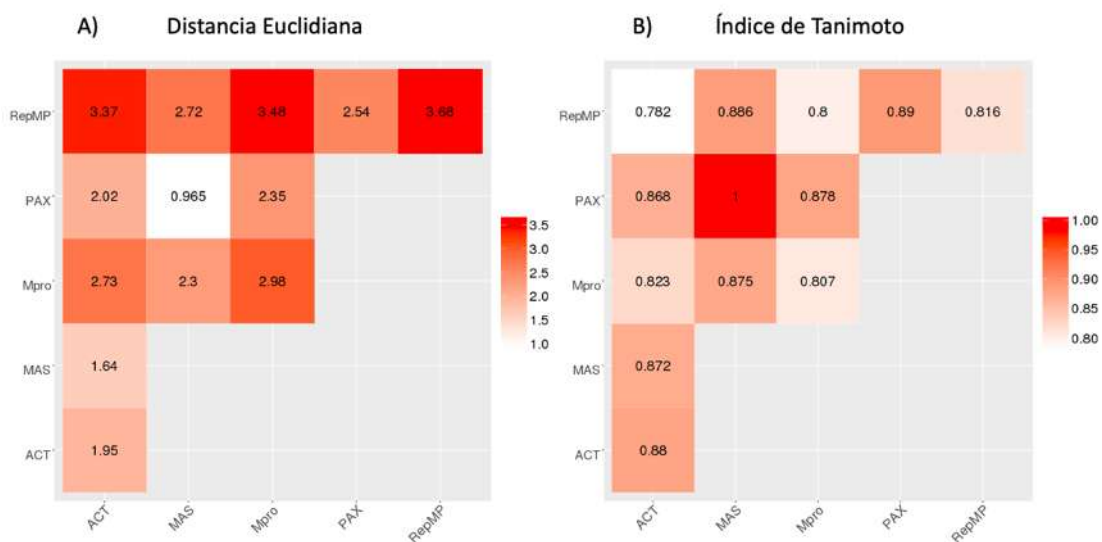


Figura 27. Matrices de (dis)similitud A) Distancia Euclidiana y B) Índice de similitud de Tanimoto basados en el cálculo de seis propiedades fisicoquímicas de interés farmacéutico. Para A) se utilizó un gradiente de color de rojo (valores altos de distancia) a blanco (valores bajos de distancia), para B) se utiliza el mismo gradiente de colores que van de rojo (el valor más cercano a 1) a blanco (valores menos cercanos a 1).

10.12 ANÁLISIS DE LA DIVERSIDAD DE SCAFFOLDS MEDIANTE EL USO DE CURVAS DE RECUPERACIÓN DE SISTEMAS CÍCLICOS (CSR) Y ENTROPÍA DE SHANNON PARA COMPUESTOS CON AFINIDAD POR MPRO

Recordando de las curvas CSR obtenidas para la comparación entre grupos de compuestos afines a Mpro y PLpro, las curvas CSR indican en el eje (y) la fracción de moléculas del conjunto de datos, mientras que en el eje (x) se grafica la fracción de sistemas cíclicos que las contienen. El valor de F_{50} indica la fracción de quimiotipos necesarios para recuperar el 50% de los compuestos y el área bajo la curva (AUC) es un indicativo de la distribución de los compuestos de acuerdo a sus quimiotipos. Valores altos de F_{50} sugieren una mayor diversidad en los quimiotipos encontrados para un grupo de datos mientras que una curva más cercana a ajustarse a una diagonal en el gráfico, es decir, con una AUC menor es indicativo de mayor diversidad de quimiotipos. La tabla 11 resume la información sobre el número de compuestos (M), número de *Scaffolds* o de quimiotipos únicos (N), la fracción de quimiotipos únicos relativa al número de moléculas en el conjunto de datos (N/M), el número de quimiotipos que contienen un solo compuesto (NSING), las fracciones de NSING relativas a M y a N (NSING/M y NSING/N respectivamente), el área bajo la curva (AUC) y la fracción de quimiotipos necesarios para recuperar el 50 %de los compuestos (F_{50}). Para los compuestos con afinidad por la proteasa Mpro, las curvas CSR de la gráfica 4 indican que los compuestos fármacos repositionados (RepMP) representados con el color púrpura tienen una mayor diversidad en sus estructuras núcleo o Scaffolds, con valores de $F_{50}=0.13$ y $AUC=0.73$. Además, los valores de N/M, un indicativo de aquellos quimiotipos con un con un solo compuesto es mucho menor para los compuestos “ACT” con un valor de 0.23 contra 0.41 y 0.47 para “Mpro” y “RepMP” respectivamente, lo que va de acuerdo a los resultados anteriores donde se sugiere una alta similitud dentro de este grupo de compuestos.

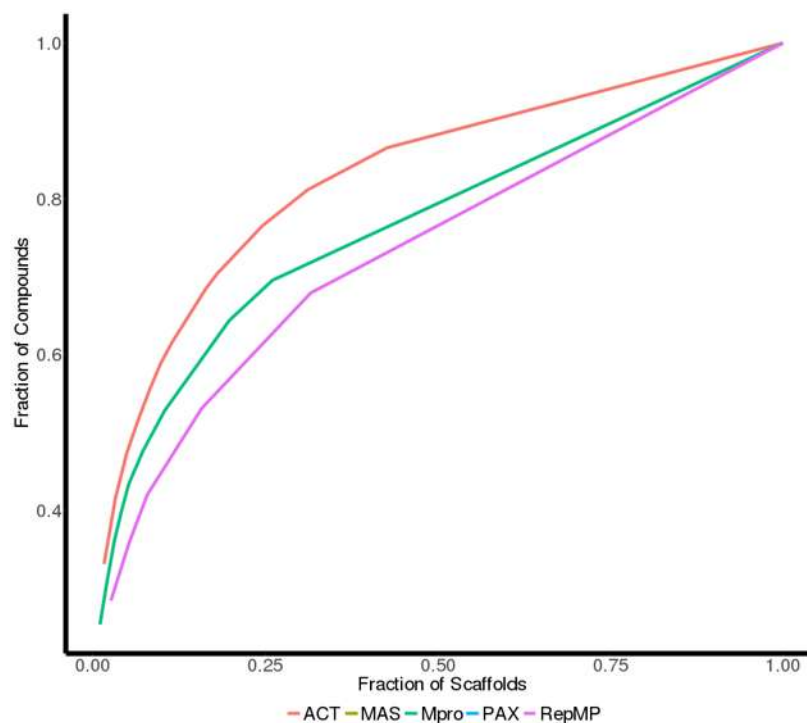


Gráfico 4. Curvas de recuperación de sistemas cíclicos (CSR) para subgrupos de compuestos reportados con afinidad por la proteasa principal Mpro. Compuestos con actividad biológica reportada (ACT), con afinidad sugerida por métodos computacionales (Mpro) y fármacos reposicionados (RepMP).

Tabla 11. Recuento de Scaffolds, AUC y F50 obtenidos para los conjuntos de datos. M: número de moléculas; N: número de Scaffolds; NSING: número de Scaffolds con solo un compuesto; AUC: area debajo de la curva; F₅₀: Fracción de quimiotipos requerida para recuperar el 50% de las moléculas.

	M	N	N/M	NSING	NSING/M	NSING/N	AUC	F50
ACT	260.00	61.00	0.23	35.00	0.13	0.57	0.82	0.07
Mpro	233.00	96.00	0.41	71.00	0.30	0.74	0.76	0.08
RepMP	81.00	38.00	0.47	26.00	0.32	0.68	0.73	0.13

Adicionado al análisis de Scaffolds por la curva CSR de la gráfica 4, se cuantificó por entropía escalada de Shannon (SSE) la distribución específica de los compuestos en los quimiotipos o scaffolds más poblados para cada conjunto de datos. La tabla 12 muestra los valores de SSE para los quimiotipos más poblados (10-60 quimiotipos), recordando que los valores de SSE más cercanos a 1.0 indican que las moléculas están distribuidas más equitativamente en los sistemas de quimiotipos únicos por lo que son más diversas. De acuerdo a los valores de la

tabla, los compuestos de la base de datos de fármacos reposicionados siguen con la tendencia mostrada por la curva CSR de la gráfica 4, donde se sugirió que este grupo es el más diverso en sus estructuras núcleo con respecto a los otros dos, con valores de SSE de 0.977 a 0.986. Se muestran además en las figuras 28, 29 y 30, la distribución de los diez quimiotipos más poblados así como la estructura de los tres scaffolds más poblados. Se observa en los gráficos que el histograma para los quimiotipos más poblados de fármacos reposicionados que los compuestos están distribuidos más equitativamente en las barras, lo cual es un indicador de diversidad.

Tabla 12. Valores de SSE de los 10-60 quimiotipos únicos más poblados.

Data set	SSE10	SSE20	SSE30	SSE40	SSE50	SSE60
ACT	0.964	0.955	0.946	0.947	0.936	0.93
Mpro	0.966	0.965	0.952	0.948	0.949	0.951
RepMP	0.977	0.969	0.975	0.98	0.983	0.986

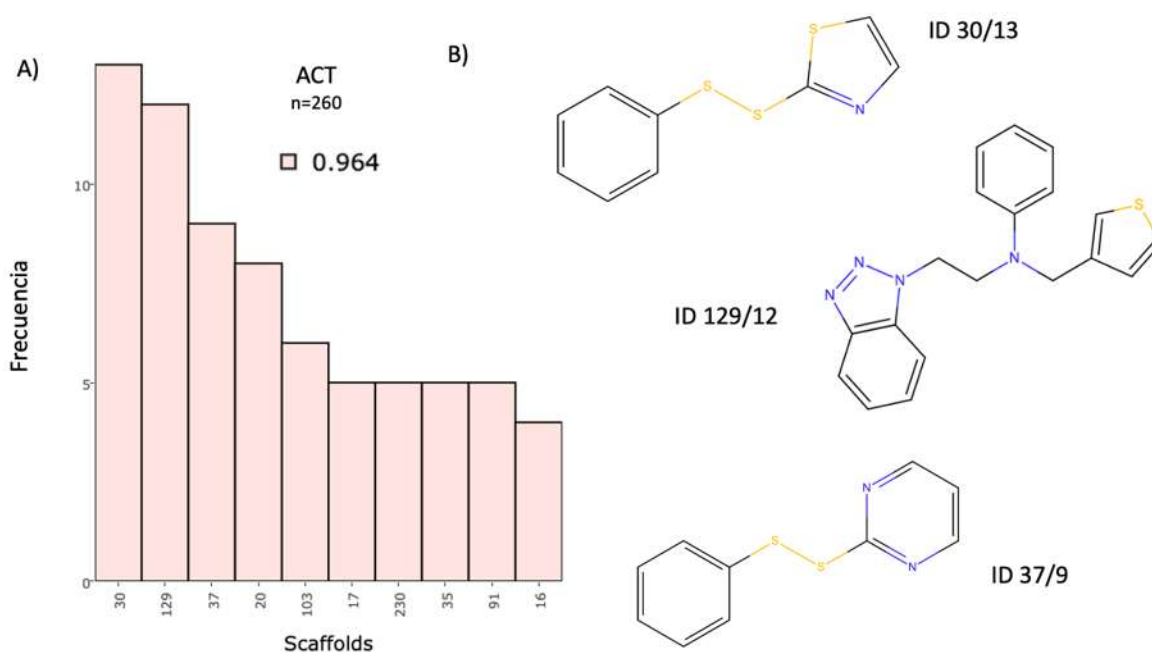


Figura 28. Distribución y estructura de *Scaffolds* o quimiotipos más frecuentes en la quimioteca de compuestos con actividad biológica reportada contra la proteasa Mpro (n=260).

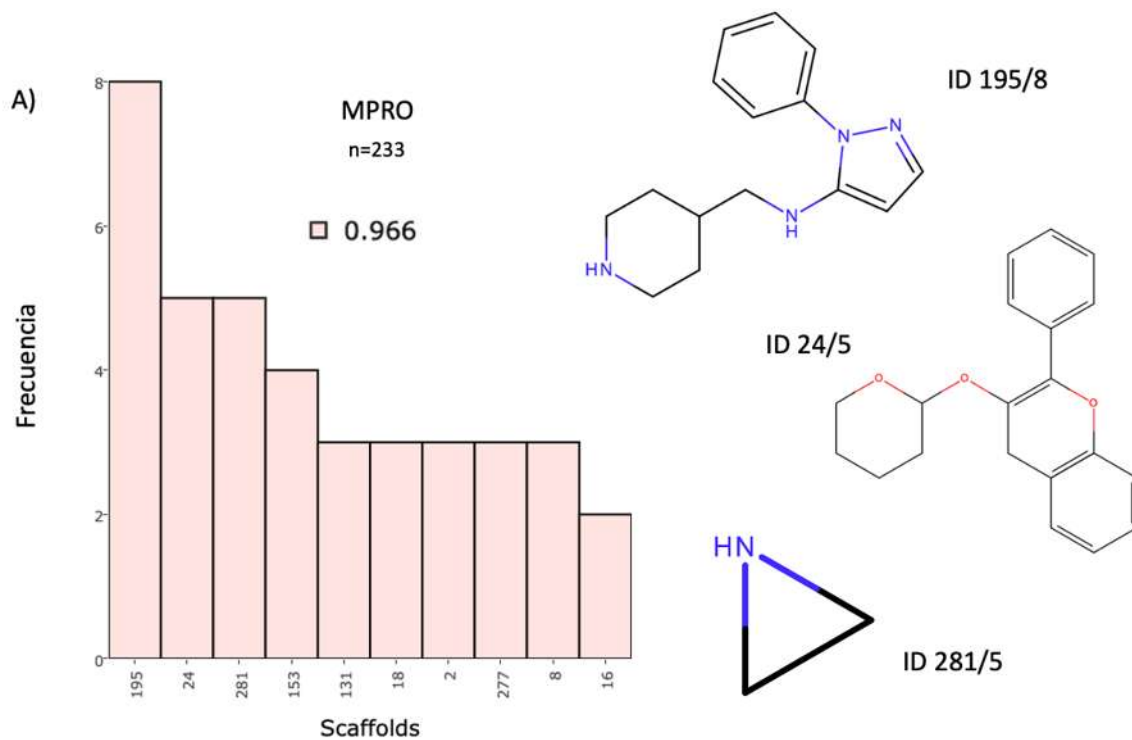


Figura 29. Distribución y estructura de *Scaffolds* o quimiotipos más frecuentes en la quimioteca de compuestos con afinidad por la proteasa Mpro de acuerdo a metodologías computacionales (n=233).

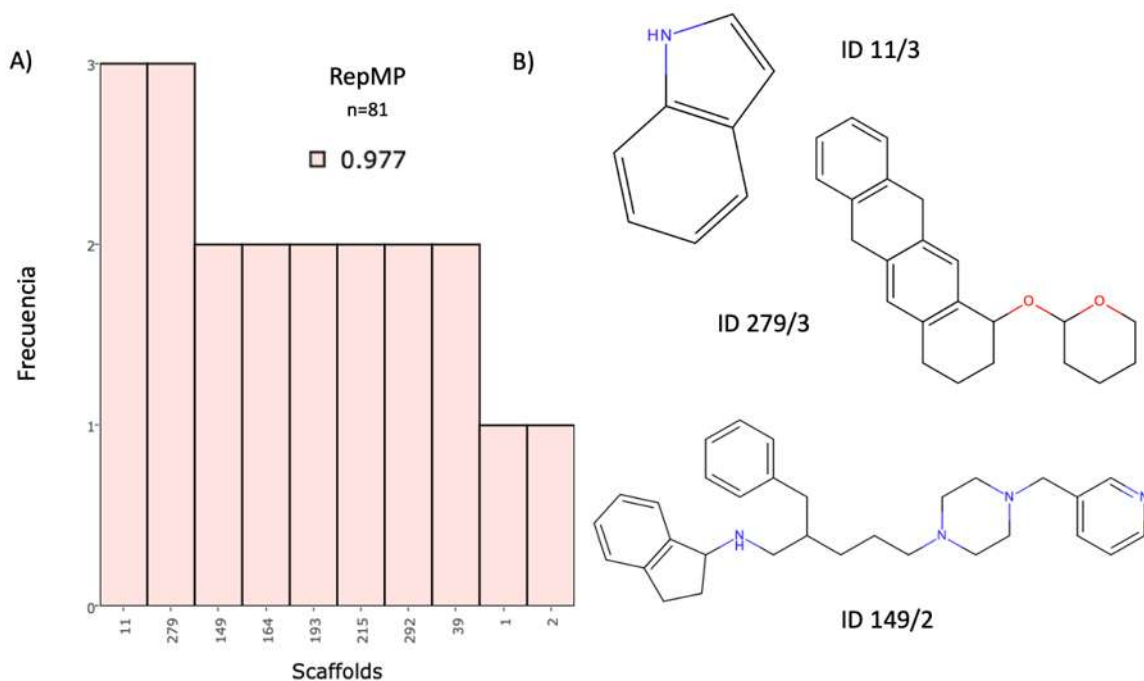


Figura 30. Distribución y estructura de *Scaffolds* o quimiotipos más frecuentes en la quimioteca de fármacos reposicionados contra la proteasa Mpro (n=81).

10.13 ANÁLISIS DE PROPIEDADES FÍSICOQUÍMICAS DE INTERÉS FARMACEÚTICO PARA LOS COMPUESTOS REPORTADOS CON AFINIDAD POR LA PROTEASA DE TIPO PAPAÍNA PLPRO

De manera similar al análisis para los compuestos afines a la proteasa Mpro, se calcularon en la plataforma PUMA las seis propiedades de interés farmacéutico (MW, ALogP, nHBAcc, nHBDOn, TPSA, nRotB) para los compuestos reportados con afinidad por la proteasa de tipo papaína PLpro. Las pruebas de normalidad indican que ninguna de las propiedades para los grupos presenta una distribución normal por lo que el test de rango de Wilcoxon se utilizó para determinar la existencia de diferencias estadísticamente significativas entre los grupos. Las figuras 31 y 32 muestra la distribución para las seis propiedades calculadas a manera de histogramas y gráfico de cajas respectivamente, mientras que la tabla 13 resume los resultados de las propiedades para los diferentes grupos. De la misma manera que para el análisis de Mpro, los compuestos se nombraron “ACT” (con actividad biológica reportada), “PLpro” (con afinidad por PLpro de acuerdo a métodos computacionales) y RepPL (fármacos reposicionados contra PLpro), se incluye además el compuesto GRL-0617 (“GRL”) que se utilizó para validar el protocolo de Docking como un compuesto de referencia. Al comparar los compuestos que tienen actividad biológica reportada “ACT” con los afines por la proteasa “PLpro” y fármacos reposicionados “RepPL” no se encontraron diferencias en cuanto a su MW y nRotB, lo que sugiere que los compuestos “ACT” comparten características de tamaño y flexibilidad con los demás grupos. Por otra parte, las propiedades TPSA, nHBDOn, nHBAcc y AlogP si presentan diferencias contra ambos grupos “PLpro” y “RepPL” donde los compuestos “ACT” parecen tener estructuras menos polares con valores de TPSA, nHBAcc y nHBDOn menores. La comparación de compuestos “PLpro” contra fármacos reposicionados “RepPL” presenta diferencias en las propiedades TPSA, nRotB, nHBAcc y ALogP, mientras que MW y nHBDOn son estadísticamente iguales, esto podría indicar una mayor polaridad y flexibilidad de los fármacos reposicionados “RepPL” con valores de nRotB y TPSA más altos.

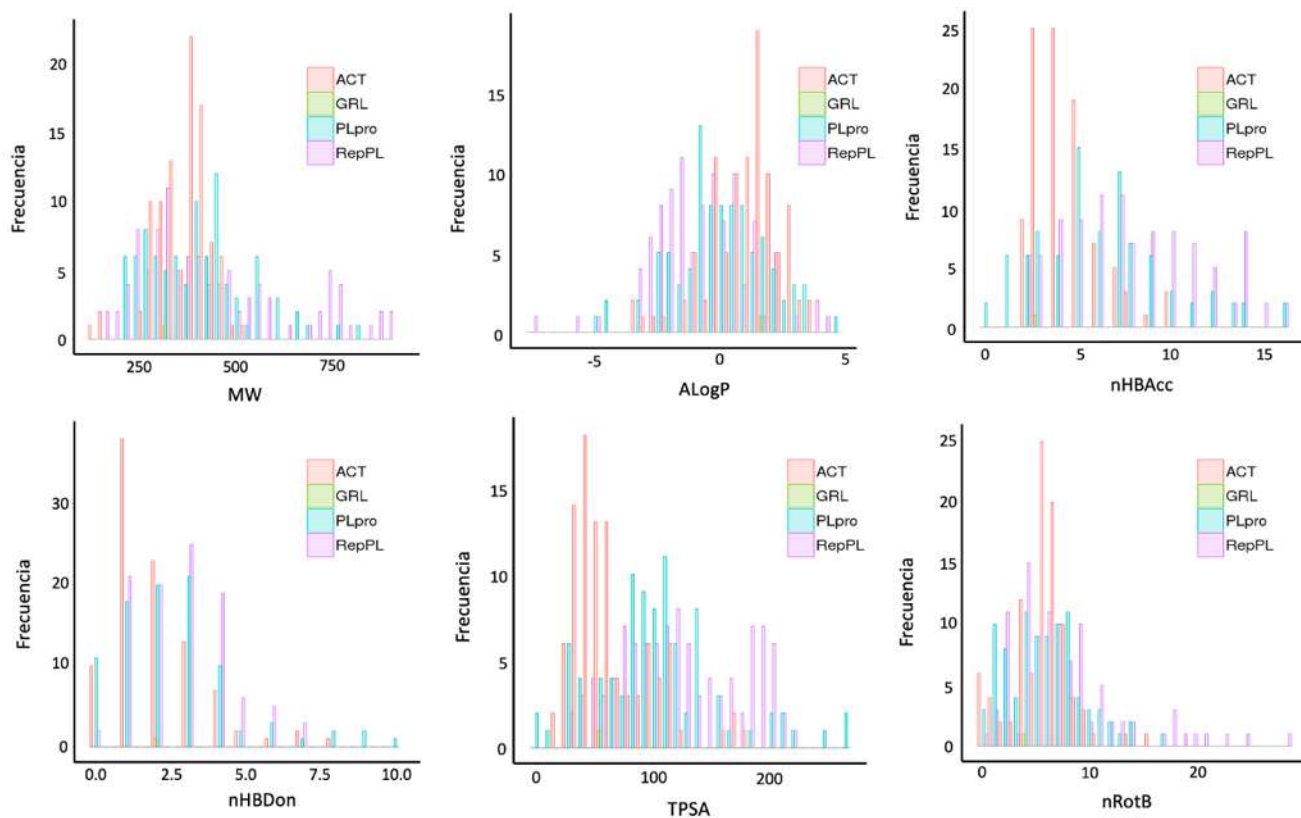


Figura 31. Histograma de seis propiedades fisicoquímicas de compuestos encontrados en la literatura con afinidad por la proteasa PLpro. División en subgrupos “ACT” (actividad biológica reportada), “PLpro” (Afinidad probada por métodos computacionales) y “RepPL” (fármacos reposicionados).

Tabla 13. Distribución estadística de las propiedades MW, AlogP, nHBacc, nHBDon, nRotB y TPSA para compuestos con afinidad por PLpro divididos en subgrupos “PLpro”, “ACT” y “RepPL”.

Propiedad	Data set	Min	1er Qu	Mediana	Media	3er Qu	Max	Desv. Est
MW	PLpro	200.13	288.06	398.13	397.57	457.77	822.40	132.12
	ACT	125.05	328.13	391.21	373.38	418.23	538.09	71.45
	RepPL	157.11	301.17	378.12	435.06	540.31	888.41	193.03
AlogP	PLpro	-4.98	-0.97	-0.12	-0.05	1.12	4.51	1.86
	ACT	-3.48	0.17	1.17	1.08	2.16	3.80	1.49
	RepPL	-7.68	-2.16	-0.75	-0.82	0.48	4.16	1.95
nHBacc	PLpro	0.00	4.00	6.00	6.20	8.00	16.00	3.58
	ACT	2.00	3.00	4.00	4.41	5.00	10.00	1.82
	RepPL	2.00	5.00	7.00	7.91	11.00	16.00	3.73
nHBDon	PLpro	0.00	1.00	2.00	2.64	3.00	10.00	2.14
	ACT	0.00	1.00	2.00	1.95	3.00	8.00	1.58
	RepPL	0.00	2.00	3.00	2.90	4.00	7.00	1.60
TPSA	PLpro	3.24	73.72	98.29	102.67	122.89	267.04	53.55
	ACT	20.31	41.57	55.12	65.08	80.87	180.30	35.05
	RepPL	27.63	81.78	114.59	121.87	166.06	220.15	51.01
nRotB	PLpro	0.00	3.00	6.00	5.70	8.00	16.00	3.54
	ACT	0.00	4.00	6.00	5.86	7.00	15.00	2.76
	RepPL	0.00	4.00	6.00	7.58	9.00	28.00	5.23

Tabla 14. Propiedades fisicoquímicas de GRL-0617.

Molécula	MW	ALogP	nHBAcc	nHBDon	TPSA	nRotB
GRL-0617	304.157563	1.5758	3	2.00	55.12	4

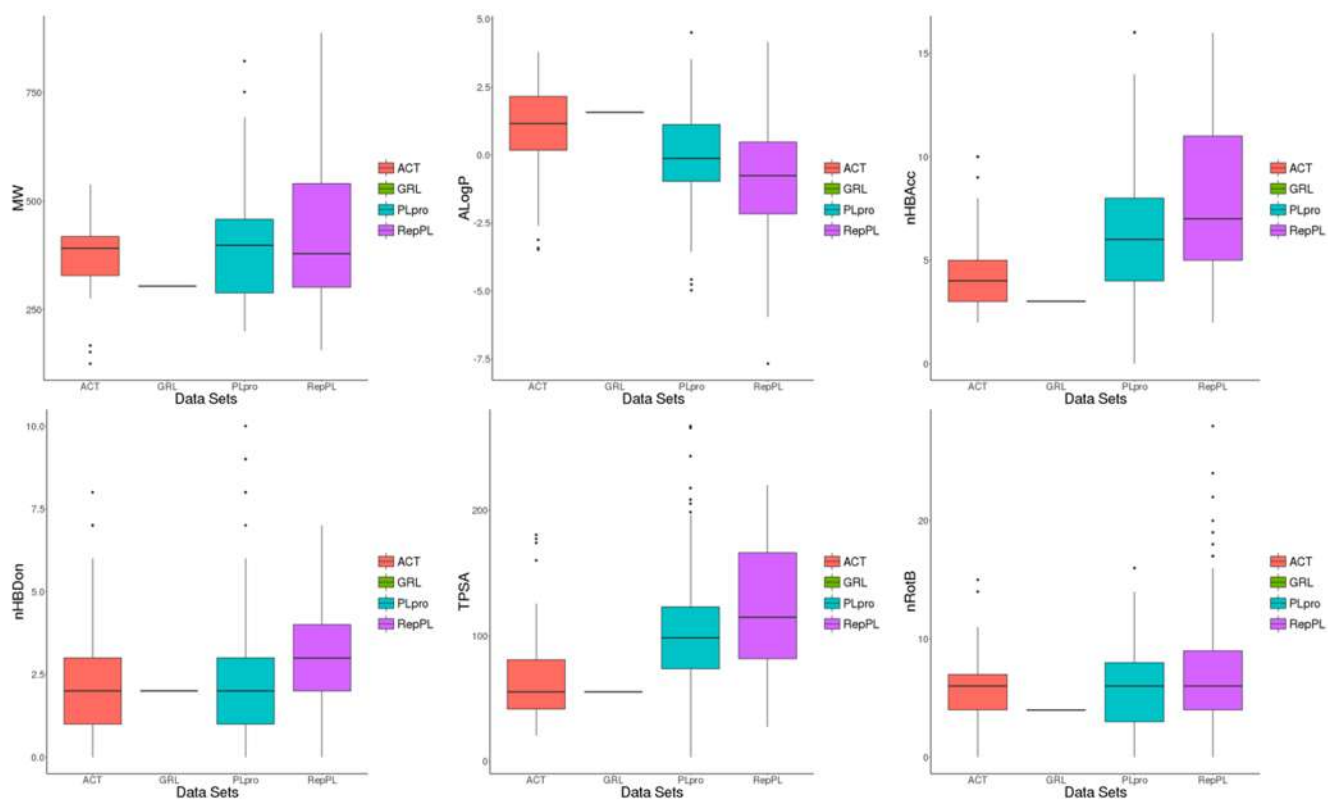


Figura 32. Distribución por gráficos de cajas para las seis propiedades de interés farmacéutico de compuestos con afinidad por PLpro divididos en subgrupos “ACT”, “Mpro” y “RepMP”.

10.14 ANÁLISIS DE DIVERSIDAD MOLECULAR BASADO EN COMPARACIONES SISTEMÁTICAS POR PARES

En el análisis de distancia euclidiana y similitud por índice de Tanimoto se encontró que el grupo de fármacos reposicionados “RepPL” tiene los valores más altos respecto a los otros grupos, lo que indica una mayor distancia y una menor similitud con respecto a los grupos “ACT” y “PLpro”, además el valor de distancia con respecto a otros compuestos de su mismo grupo (3.47) es el más grande con respecto a los otros valores de compuestos dentro de un mismo grupo (3.09 y 2.04 para compuestos “PLpro” y “ACT” respectivamente), lo que indica que el grupo “RepPL” es muy diverso con puntos en el espacio químico más separados entre sí y con respecto a los otros grupos, esto se comprueba con un valor de índice de similitud de Tanimoto de 0.777, el cual es el valor más bajo en la matriz B (Fig. 33). El grupo con menor distancia entre sus compuestos y mayor similitud es el de compuestos con actividad biológica reportada “ACT” (2.04 y 0.917 respectivamente) lo que de manera análoga al análisis de Mpro indica que estos compuestos comparten propiedades fisicoquímicas en común que pueden estar relacionadas a su estructura y a su actividad.

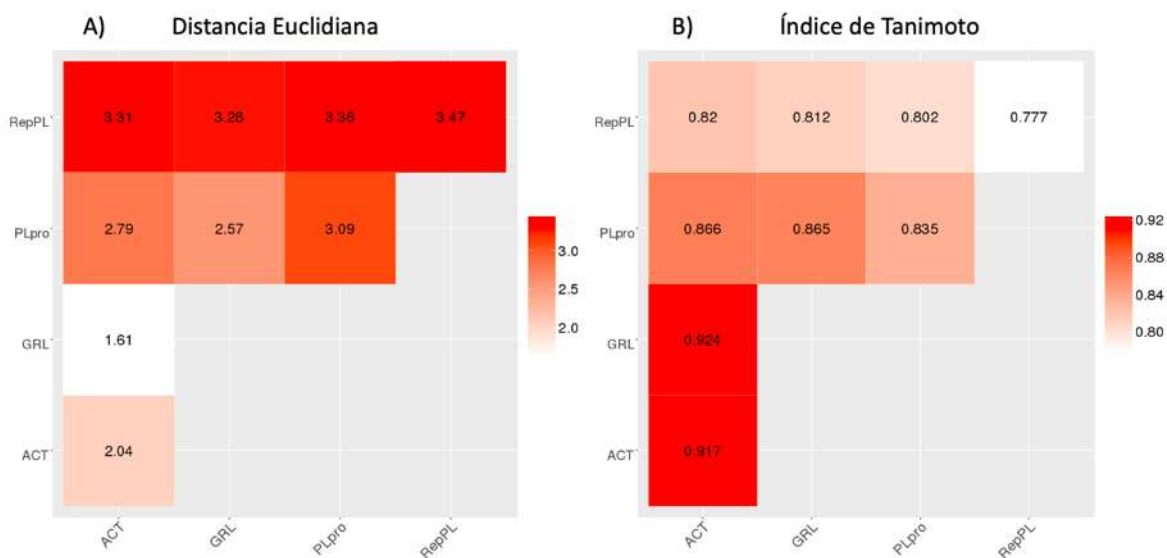
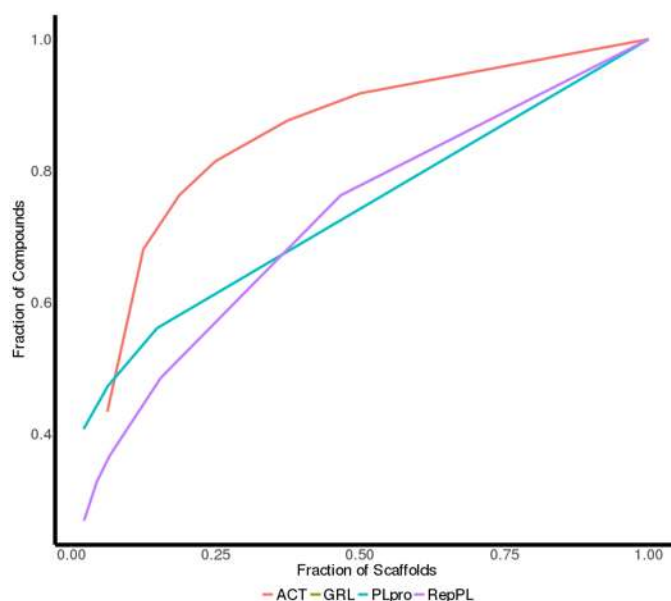


Figura 33. Matrices de (dis)similitud A) Distancia Euclidiana y B) Índice de similitud de Tanimoto basados en el cálculo de seis propiedades fisicoquímicas de interés farmacéutico. Para A) los valores más altos de distancia se muestran en color rojo fuerte y los más bajos Para A) se utilizó un gradiente de color de rojo (valores altos de distancia) a blanco (valores bajos de distancia), para B) se utiliza el mismo gradiente de colores que van de rojo (el valor más cercano a 1) a blanco (valores menos cercanos a 1).

10.15 ANÁLISIS DE LA DIVERSIDAD DE ESCAFOLDES MEDIANTE EL USO DE CURVAS DE RECUPERACIÓN DE SISTEMAS CÍCLICOS (CSR) Y ENTROPÍA DE SHANNON PARA COMPUESTOS CON AFINIDAD POR PLPRO

Para los compuestos con afinidad por la proteasa de tipo papaína PLpro, las curvas CSR, indicativas de la diversidad de quimiotipos únicos en los diferentes grupos, muestran una mayor diversidad de Scaffolds o estructuras núcleo para los compuestos de los grupos “RepPL” y “PLpro” con valores de F_{50} de 0.18 y 0.9 respectivamente, se observa en las curvas de la gráfica 5, que los compuestos “RepPL” tienen un valor más alto de F_{50} pero a medida que la fracción de compuestos aumenta, la curva se empalma con la correspondiente a los compuestos “PLpro”, por otra parte, la curva para los compuestos “ACT” tiene una mayor AUC (0.83) y valores de N/M (fracción de quimiotipos únicos respecto a número de compuestos) y NSING (quimiotipos con un solo compuesto) más bajos respecto a los otros dos grupos. Esto junto con los resultados de similitud de Tanimoto y distancia Euclidiana sugieren una alta similitud de propiedades y baja diversidad de quimiotipos únicos en los compuestos con actividad biológica reportada “ACT”.



Gráfica 5. Curvas de recuperación de sistemas cíclicos (CSR) para subgrupos de compuestos reportados con afinidad por la proteasa de tipo papaína PLpro. Compuestos con actividad biológica reportada (ACT), con afinidad sugerida por métodos computacionales (PLpro) y fármacos reposicionados (RepPL).

Tabla 15. Recuento de *Scaffolds*, AUC y F50 obtenidos para los conjuntos de datos. M: número de moléculas; N: número de *Scaffolds*; NSING: número de *Scaffolds* con solo un compuesto; AUC: area debajo de la curva; F₅₀: Fracción de quimiotipos requerida para recuperar el 50% de las moléculas.

	M	N	N/M	NSING	NSING/M	NSING/N	AUC	F₅₀
PLpro	91.00	47.00	0.52	40.00	0.44	0.85	0.73	0.09
ACT	97.00	16.00	0.16	8.00	0.08	0.50	0.83	0.06
RepPL	101.00	45.00	0.45	24.00	0.24	0.53	0.72	0.18

Además de las curvas CSR, se identificaron mediante entropía escalada de Shannon (SSE) los quimiotipos más frecuentes para cada base de datos junto con sus valores de SSE. La tabla 16 resume los valores de SSE10-60 para los diferentes grupos, donde se observan valores de SSE más altos para los compuestos con afinidad por PLpro “PLpro”, lo que indica que en cuanto a quimiotipos únicos, la base de datos “PLpro” presenta mayor diversidad, se observa también que para los compuestos “ACT” no se tienen más de treinta quimiotipos únicos y se tienen los valores de SSE más bajos, un indicativo de menor diversidad respecto a los otros grupos, esto puede verse en los histogramas de las figuras 34, 35 y 36, donde se ilustra la distribución de los quimiotipos más poblados así como la estructura química de los tres quimiotipos más frecuentes. En los histogramas se observa que los quimiotipos de los grupos “RepPL” y “PLpro” tienen una distribución más uniforme, lo que denota diversidad, mientras que el histograma para los compuestos ACT, muestra poca diversidad con una distribución más variada, con *Scaffolds* como ID 88 con 18 compuestos (18.5%) y ID 94 con 11 compuestos (11.32%). En resumen los resultados indican que en cuanto a propiedades los fármacos reposicionados son más diversos, mientras que en cuanto a quimiotipos únicos los compuestos con afinidad por PLpro parecen tener más diversidad, por otra parte, los compuestos con actividad biológica “ACT” son similares tanto en propiedades como en quimiotipos únicos.

Tabla 16. Valores de SSE de los 10-60 quimiotipos únicos más poblados.

Data set	SSE10	SSE20	SSE30	SSE40	SSE50	SSE60
PLpro	0.942	0.957	0.969	0.976	0.981	0.984
ACT	0.913	0.873	0.858	NA	NA	NA
RepPL	0.901	0.897	0.913	0.927	0.938	0.946

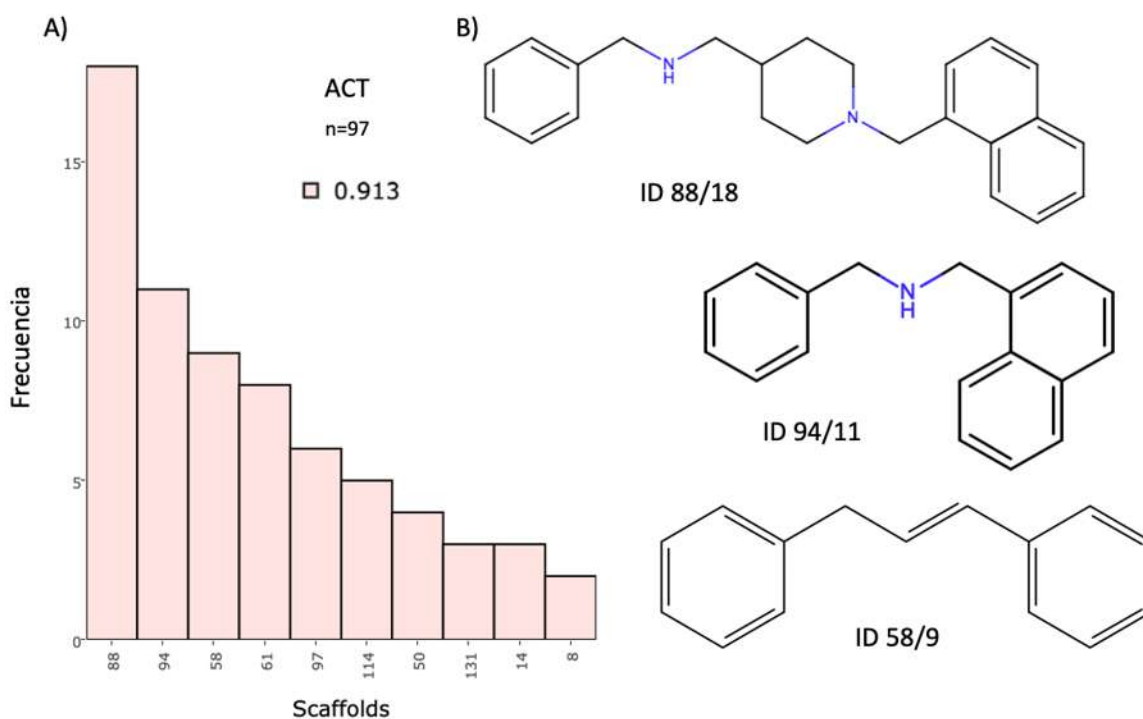


Figura 34. Distribución y estructura de *Scaffolds* o quimiotipos más frecuentes en la quimioteca de compuestos con actividad biológica reportada contra la proteasa PLpro (n=97).

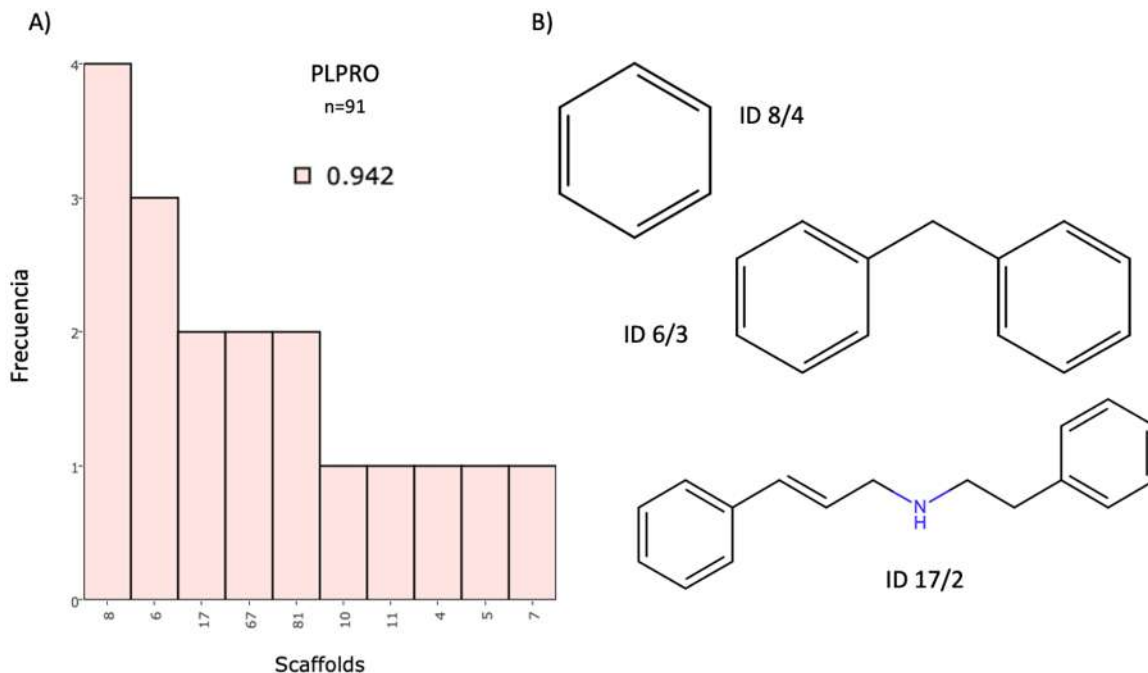


Figura 35. Distribución y estructura de *Scaffolds* o quimiotipos más frecuentes en la quimioteca de compuestos con afinidad por la proteasa PLpro de acuerdo a metodologías computacionales (n=91).

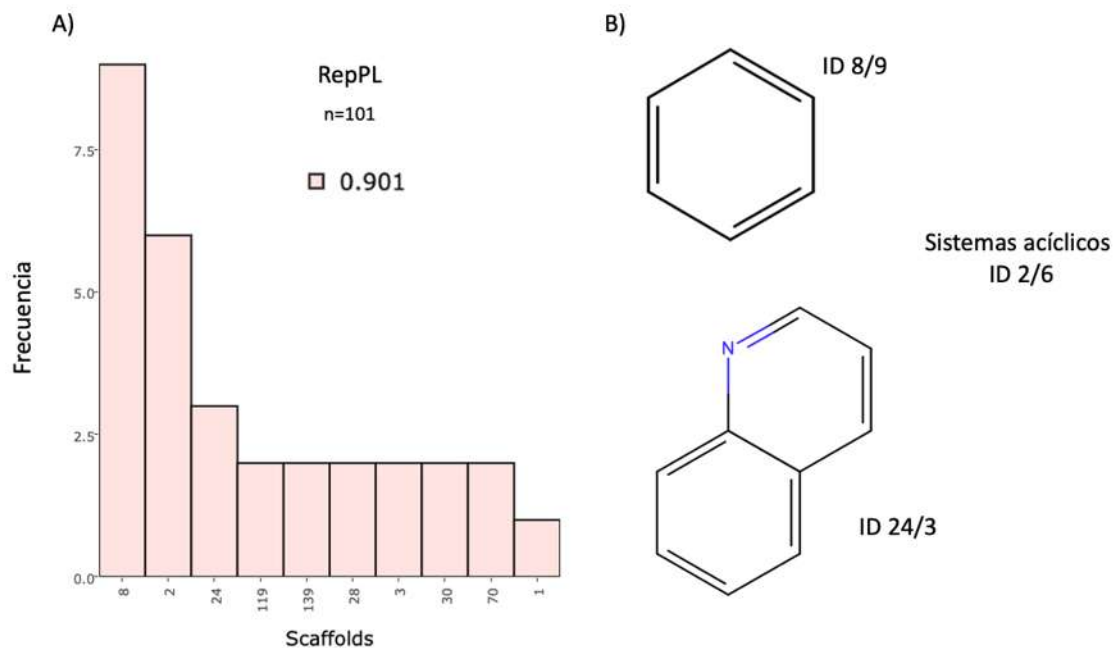


Figura 36. Distribución y estructura de *Scaffolds* o quimiotipos más frecuentes en la quimioteca de fármacos reposicionados contra la proteasa PLpro (n=101).

10.16 VISUALIZACIÓN DEL ESPACIO QUÍMICO BASADO EN LAS SEIS PROPIEDADES FISICOQUÍMICAS MW, HBAcc, HBDon, nRotB, ALogP y TPSA DE COMPUESTOS CON AFINIDAD POR LA PROTEASA PRINCIPAL MPRO

Se utilizaron las seis propiedades fisicoquímicas mencionadas anteriormente (MW, HBAcc, HBDon, nRotB, ALogP y TPSA) para mediante un análisis de componentes principales (PCA) visualizar el espacio químico de los compuestos reportados con afinidad por la proteasa Mpro. Los componentes PCa y PCb recuperaron el 87% de la covarianza de los datos y se utilizaron como base para graficar en el espacio los diferentes compuestos (Fig. 37). Se incluyen además en el espacio químico los compuestos Masitinib y Paxlovid, ambos compuestos de referencia para la búsqueda de moléculas con propiedades fisicoquímicas similares. De acuerdo con la tabla 17, los descriptores MW, TPSA, nRotB, nHBDon y nHBAcc contribuyen con mayor peso a la variabilidad representada por el PCa y tienen una correlación inversa con respecto al ALogP. Esto indica que compuestos posicionados en valores más negativos en el eje del PCa tienen valores de MW, TPSA, nRotB, mHBDon y nHBAcc mayores y menor ALogP. El nRotB tiene mayor peso en la variabilidad del PCb, compuestos posicionados a valores negativos de PCb tienen mayor número de enlaces rotables.

Con respecto a los compuestos de referencia, Masitinib es un fármaco inhibidor de tirosina-cinasa oralmente disponible que actúa a través de un mecanismo de acción competitiva. Se ha demostrado que ratones infectados con SARS-CoV-2 tratados con Masitinib redujeron significativamente la inflamación pulmonar, el compuesto tiene un $IC_{50} = 2.5 \mu M$ (Drayman *et al.*, 2021). Por otro lado, Nirmatrelvir/ritonavir (Paxlovid, Pfizer) es una combinación de inhibidores de Mpro que se ha demostrado bloquea la replicación del SARS-CoV-2, reduciendo el riesgo de hospitalización y muerte entre los pacientes con COVID-19 de leve a moderado, con riesgo de progresión a enfermedad grave. La FDA autorizó en diciembre de 2021 el tratamiento con Paxlovid para pacientes adultos y pediátricos (12 años de edad y mayores que pesen al menos 40 Kg) con COVID-19, siendo al momento el único fármaco inhibidor de Mpro que cuenta con esta aprobación (Owen *et al.*, 2021).

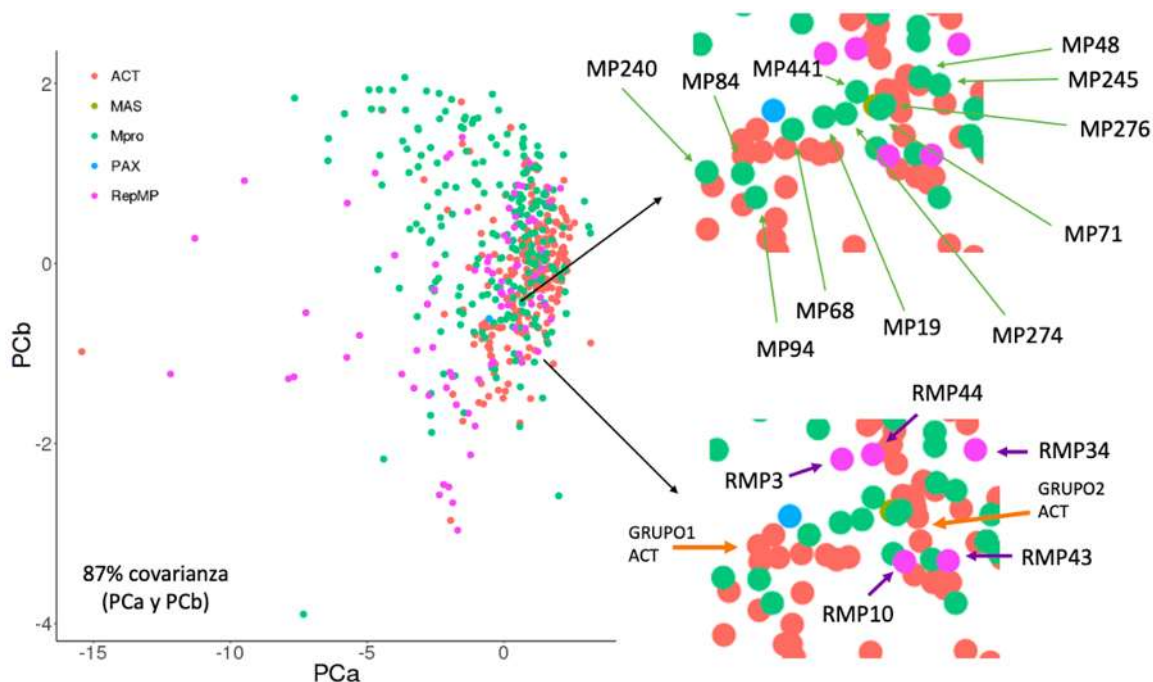
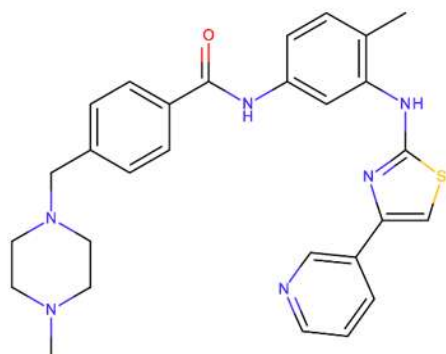


Figura 37. Representación visual del espacio químico para la base de datos de compuestos afines a Mpro mediante análisis de PCA. Los componentes principales PCa y PCb recuperaron el 87% de la covarianza de los datos. Los compuestos están divididos en compuestos reportados con afinidad por la proteasa (esferas verdes), compuestos con actividad biológica (esferas naranja) y fármacos reposicionados (esferas moradas). Se incluyen los compuestos de referencia Masitinib (esfera verde militar) y Paxlovid (esfera azul), sus estructuras se muestran en la figura 38.

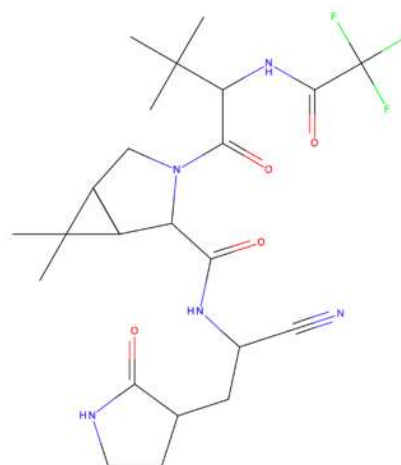
Tabla 17. Contribución de las seis propiedades fisicoquímicas a cada PC en el espacio químico de Mpro.

	PCa	PCb	PCc	PCd	PCe	PCf
MW	-0.417	-0.322	0.383	-0.582	0.345	-0.342
TPSA	-0.444	0.131	0.239	0.296	-0.646	-0.471
nRotB	-0.324	-0.743	-0.447	0.371	0.059	0.023
nHBDon	-0.420	0.332	0.276	0.528	0.568	0.183
nHBAcc	-0.456	0.011	0.063	-0.292	-0.349	0.761
AlogP	0.370	-0.463	0.718	0.265	-0.119	0.216



Masitinib

MW	TPSA	nRotB	nHBDdon	nHBAcc	ALogP
498.22	97.63	8	2	7	1.1757



Paxlovid

MW	TPSA	nRotB	nHBDdon	nHBAcc	ALogP
499.24	131.4	11	3	9	0.0983

Figura 38. Estructura química de Masitinib y Paxlovid. Los valores de sus propiedades fisicoquímicas se indican debajo del nombre del compuesto.

Partiendo de la idea de que compuestos agrupados con cercanía en el espacio químico tienen propiedades fisicoquímicas similares se hizo una colección de moléculas que se encontraban cercanas a los compuestos Masitinib y Paxlovid en el espacio químico, pensando en que ambos compuestos siendo ya inhibidores conocidos y que se unen con buena afinidad en el sitio activo de la proteasa podrían agruparse con otros compuestos con propiedades similares. Para esto se graficó también el espacio químico en tercera dimensión, ya que existen compuestos que en una imagen bidimensional pueden estar superpuestos. Los gráficos generados son los siguientes:

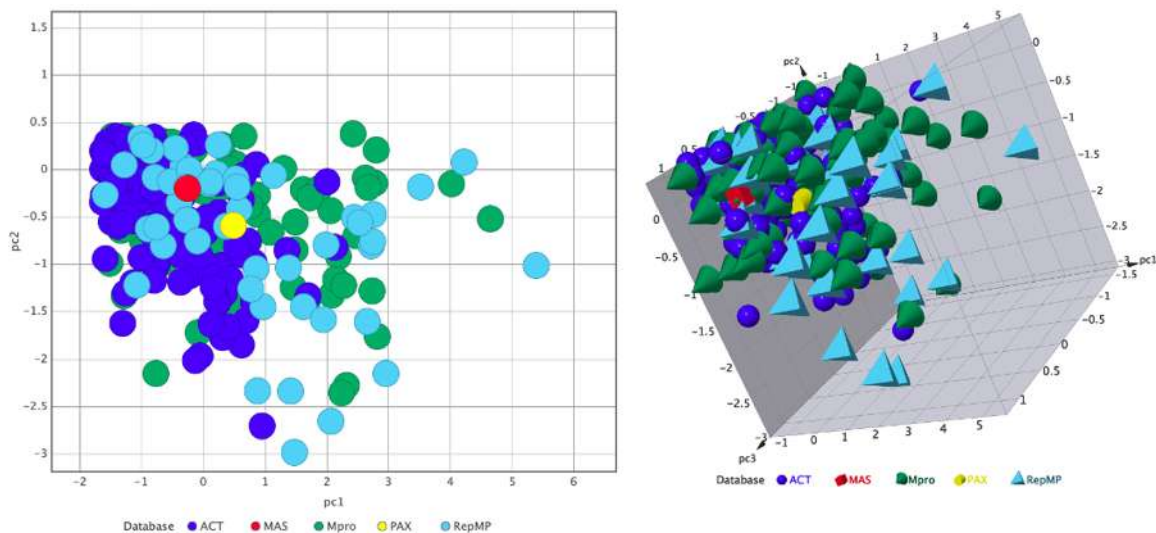


Figura 39. Representación visual 2D y 3D del espacio químico para la base de datos de compuestos afines a Mpro mediante PCA en Datawarrior. Los componentes principales PC1, PC2 y PC3 recuperaron el 89% de la covarianza de los datos. Los compuestos de referencia Masitinib (esfera roja) y Paxlovid (esfera amarilla).

En la figura 39 se observan las representaciones en dos y tres dimensiones del espacio químico obtenidas del software DataWarrior. Los diferentes colores y figuras en el gráfico denotan los diferentes grupos de moléculas en la base de datos. Los conos de color verde representan a los compuestos que tienen afinidad por la proteasa Mpro, las esferas de color azul representan a los compuestos que tienen actividad biológica reportada, y las pirámides de color azul claro representan a los fármacos reposicionados. Para tener más información en el espacio químico, se graficaron de manera independiente los tres grupos de moléculas por separado, añadiendo en la representación una escala de color, de acuerdo con el score de acoplamiento que obtuvieron en pruebas de acoplamiento molecular. Los gráficos son los siguientes:

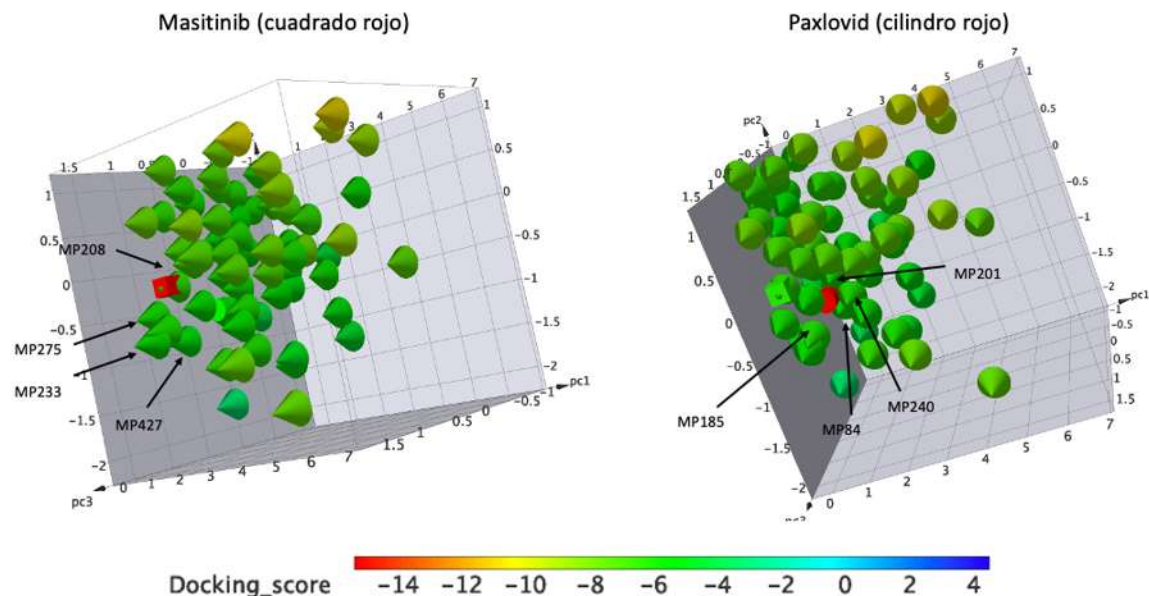


Figura 40. Representación 3D del espacio químico del grupo de compuestos con afinidad por la proteasa Mpro, mediante PCA obtenido en Datawarrior. La escala de color es representativa del score de acoplamiento obtenido en ensayo de acoplamiento molecular. El primer gráfico muestra los compuestos cercanos a Masitinib (cuadrado rojo) mientras que el segundo muestra compuestos cercanos a Paxlovid (cilindro rojo).

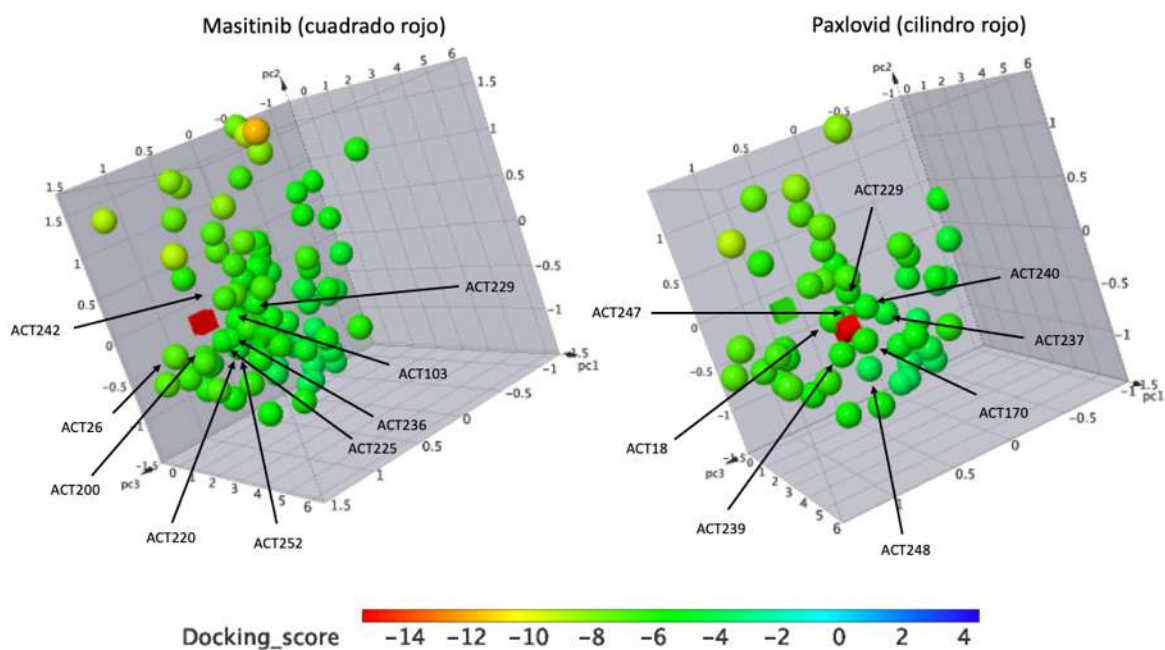


Figura 41. Representación 3D del espacio químico del grupo de compuestos con afinidad y actividad biológica reportada por la proteasa Mpro, mediante PCA obtenido en Datawarrior. La escala de color es representativa del score de acoplamiento obtenido en ensayo de acoplamiento molecular. El primer gráfico muestra los compuestos cercanos a Masitinib (cuadrado rojo) mientras que el segundo muestra compuestos cercanos a Paxlovid (cilindro rojo).

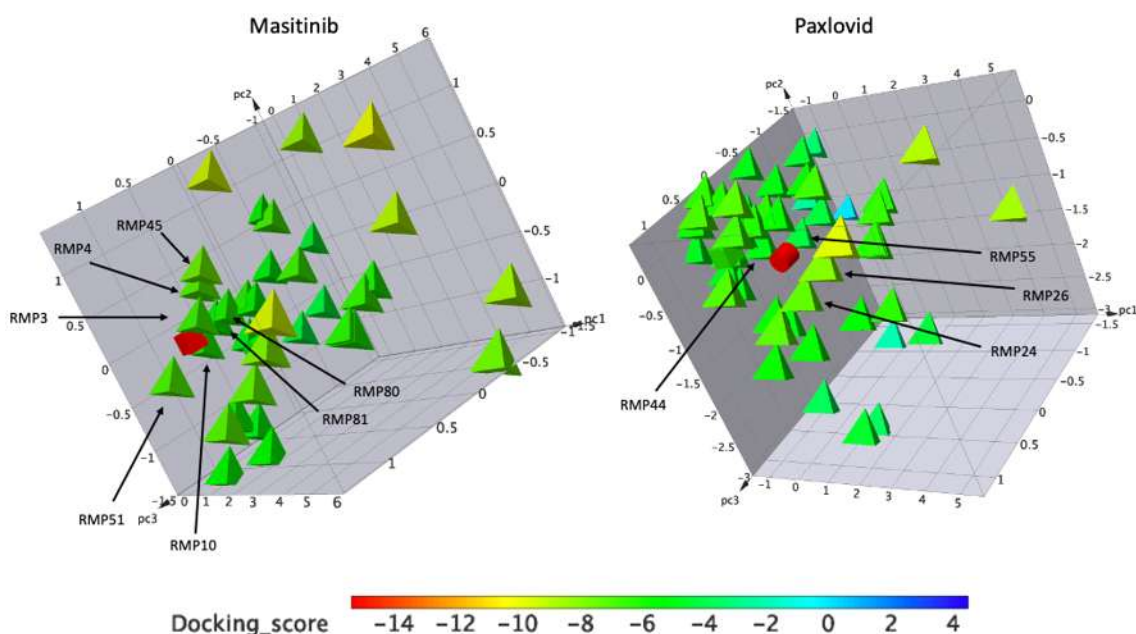


Figura 42. Representación 3D del espacio químico del grupo de compuestos reposicionados contra la proteasa Mpro, mediante PCA obtenido en Datawarrior. La escala de color es representativa del score de acoplamiento obtenido en ensayo de acoplamiento molecular. El primer gráfico muestra los compuestos cercanos a Masitinib (cuadrado rojo) mientras que el segundo muestra compuestos cercanos a Paxlovid (cilindro rojo).

Se identificaron 56 compuestos cercanos a Masitinib y Paxlovid de los diferentes grupos en el espacio químico. Esta nueva colección de compuestos se sometió a pruebas de acoplamiento molecular para evaluar de manera extensiva la afinidad de los compuestos y su acomodo con el sitio activo de Mpro. Los compuestos con mejor acoplamiento fueron los compuestos COMP11 (MP201) y COMP53 (RMP26) con puntajes de acoplamiento de -8.78 kJ y -8.62 kJ, respectivamente. El compuesto MP201 fue reportado por Mengist *et al.* (2021) en el artículo “Structural basis of potential inhibitors targeting SARS-CoV-2 main protease” donde se discute la base estructural de algunos compuestos como posibles inhibidores de Mpro. Por otra parte, el compuesto RMP26 es un analgésico de naturaleza opioide poco estudiado que causa una depresión transitoria de la presión arterial y la frecuencia cardiaca. Este compuesto fue reportado por Liu *et al.* (2020) en el artículo “computational evaluation of the COVID-19 3C-like protease inhibition mechanism, and drug

repurposing screening”. Las poses de acoplamiento de ambos compuestos se muestran en los gráficos siguientes.

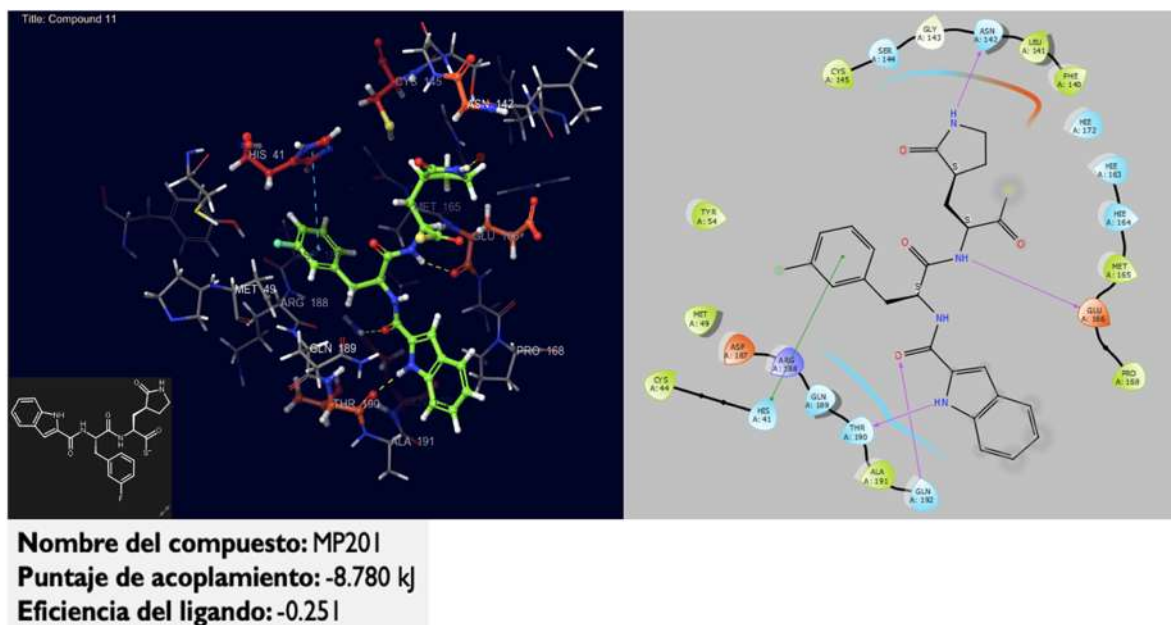
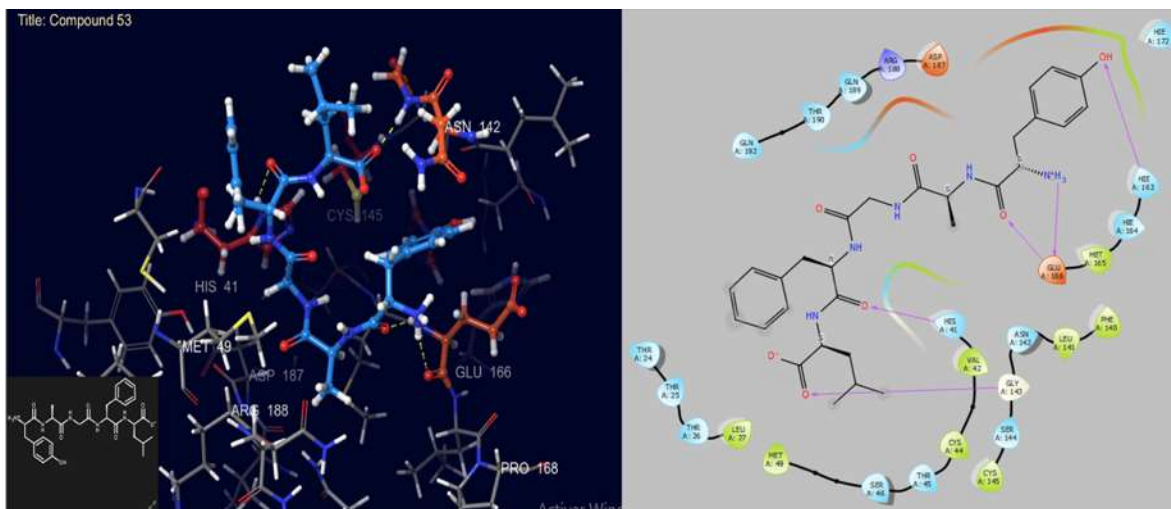


Figura 43. Acoplamiento del compuesto MP201 en el sitio activo de Mpro. En la parte izquierda se muestra la interacción tridimensional del compuesto con los residuos de aminoácidos en el sitio activo de Mpro. Las líneas amarillas indican interacciones por puentes de hidrógeno, las azules indican interacciones de apilamiento π - π . En la parte derecha se muestran las interacciones más importantes con los residuos del sitio activo.

Ambos compuestos MP201 y RMP26 tienen un acomodo tridimensional aceptable en el sitio activo de Mpro. MP201 obtuvo un puntaje de acoplamiento de -8.780 kJ y forma interacciones de tipo puente de hidrógeno con los residuos Thr190, Glu166, Gln192 y Asn142, y una interacción de apilamiento π - π entre el anillo aromático con un átomo de Flúor del compuesto y el anillo de His41 del sitio activo de la enzima. Por otro lado, RMP26 tuvo un puntaje de -8.626 kJ y forma interacciones por puente de hidrógeno con los residuos Glu166, His163, Gly143, y con His41. El acomodo de los compuestos en el sitio activo representado como superficie puede observarse en la figura 1 del material suplementario.



Nombre del compuesto: RMP26
Puntaje de acoplamiento: -8.626 kJ
Eficiencia del ligando: -0.197

Figura 44. Acoplamiento del compuesto RMP26 en el sitio activo de Mpro. En la parte izquierda se muestra la interacción tridimensional del compuesto con los residuos de aminoácidos en el sitio activo de Mpro. Las líneas amarillas indican interacciones por puentes de hidrógeno, las azules indican interacciones de apilamiento π - π . En la parte derecha se muestran las interacciones más importantes con los residuos del sitio activo.

Las estructuras de los ligandos muestran similitud con la estructura base más frecuente en la base de datos de compuestos con actividad reportada para Mpro. El scaffold se muestra en la figura 45 con el ID129 y tiene doce apariciones en su colección de compuestos. Este scaffold es proveniente de un grupo de compuestos reportados por Turlington *et al.*, en 2013 para inhibir a la proteasa principal Mpro de SARS-CoV-1 por ensayos enzimáticos y de acoplamiento molecular. Estos compuestos tienen actividad biológica contra Mpro de SARS-CoV-1 y tienen un acomodo aceptable en el sitio activo de la enzima por lo que se coleccionaron en una base de datos (26 compuestos) y se acoplaron al sitio activo de Mpro por ensayos de acoplamiento molecular para identificar su afinidad e interacciones importantes entre los átomos de los compuestos y los residuos de aminoácido del sitio activo de la enzima. Los mejores acoplamientos resultaron de los compuestos MP462, MP463 y MP470 (Fig. 46, 47 y 48).

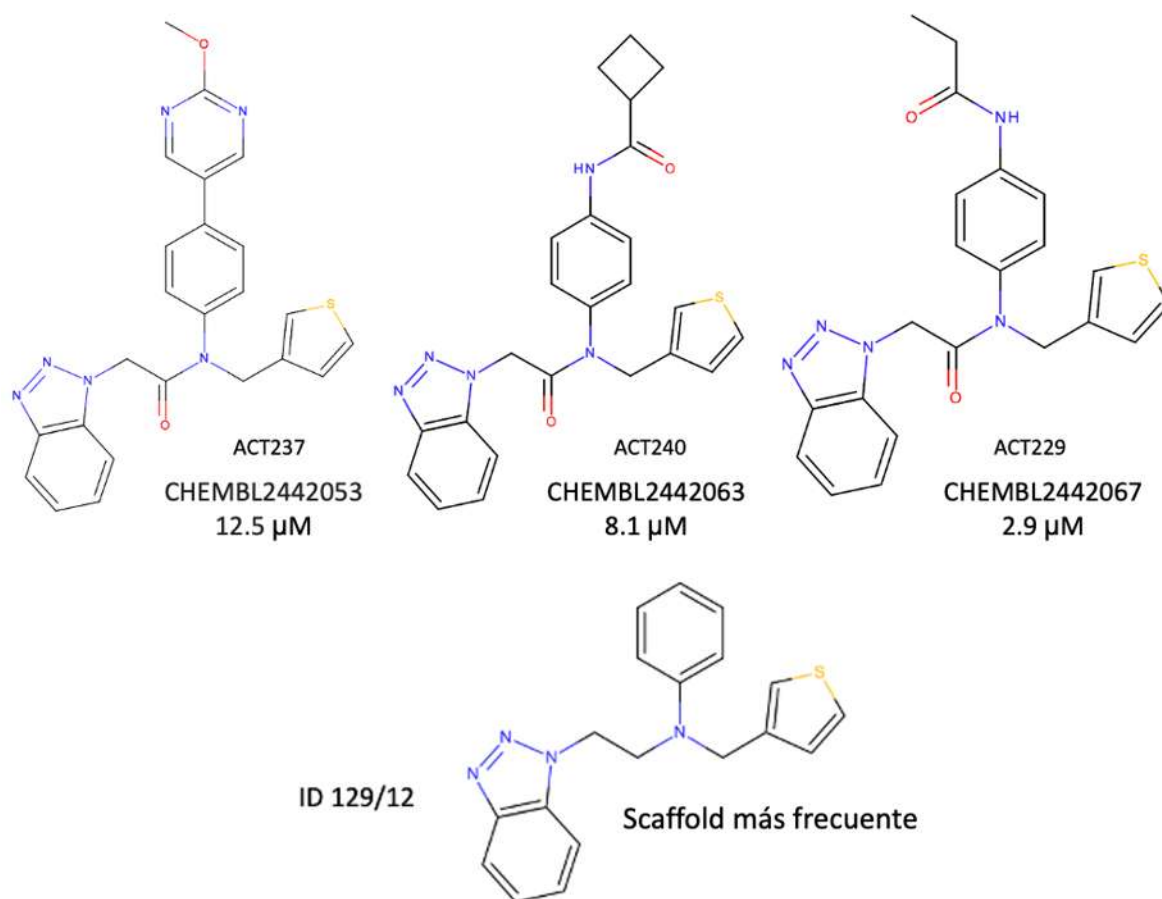
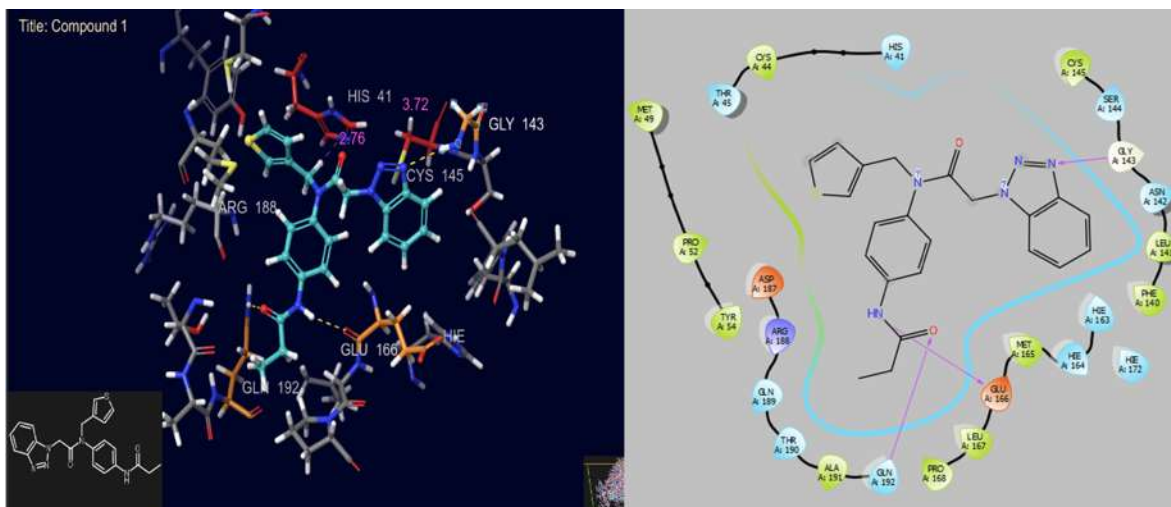


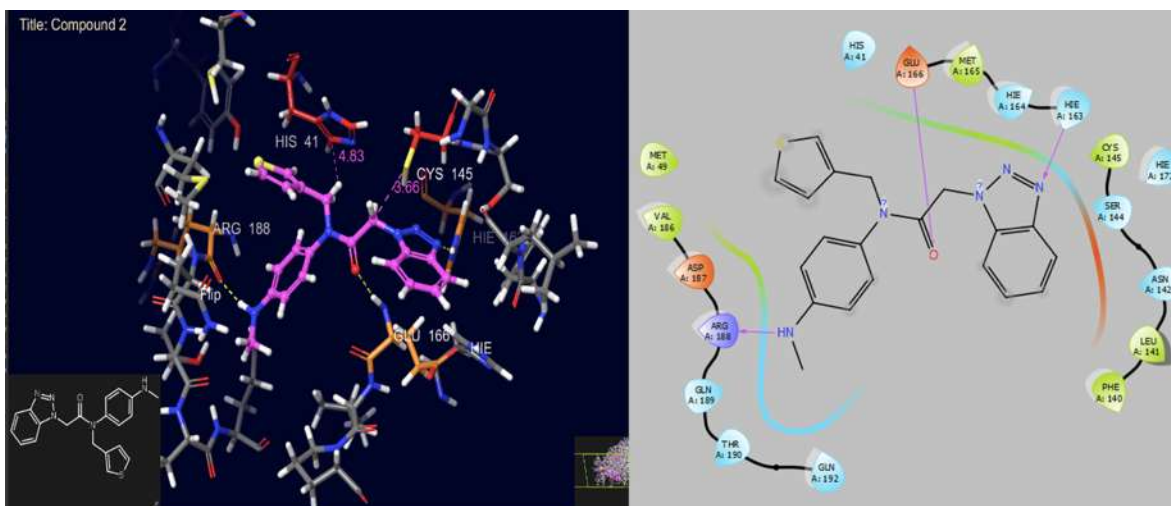
Figura 45. Familia de compuestos inhibidores de Mpro de SARS-CoV provenientes del trabajo de Turlington *et al.* (2013) y scaffold más frecuente en la base de datos de compuestos con afinidad por Mpro.

Después del análisis del acoplamiento de los 26 compuestos que contienen el scaffold ID129, las mejores poses correspondieron a los compuestos MP462, MP463 y MP470 con puntajes de acoplamiento de -7.028 kJ, -7.192 kJ y -6.968 kJ, respectivamente. Los tres compuestos presentan una orientación y conformación apropiada en el sitio activo de Mpro, lo cual puede observarse en los gráficos de superficie (figura 2 del material suplementario). El compuesto MP462 forma interacciones por puente de hidrógeno con los residuos Gly143, Glu166 y Gln192. El compuesto MP463 forma el mismo tipo de interacciones, pero con los residuos His163, Glu166 y Arg188, mientras que, por otro lado, el compuesto MP470 interacciona con el sitio activo, formando un puente de hidrógeno con Gln192 y otro puente de hidrógeno junto con una interacción de apilamiento π - π con His41.



Nombre del compuesto: MP462
Puntaje de acoplamiento: -7.028 kJ
Eficiencia del ligando: -0.234

Figura 46. Acoplamiento del compuesto MP462 en el sitio activo de Mpro. En la parte izquierda se muestra la interacción tridimensional del compuesto con los residuos de aminoácidos en el sitio activo de Mpro. Las líneas amarillas indican interacciones por puentes de hidrógeno. En la parte derecha se muestran las interacciones más importantes con los residuos del sitio activo.



Nombre del compuesto: MP463
Puntaje de acoplamiento: -7.193 kJ
Eficiencia del ligando: -0.266

Figura 47. Acoplamiento del compuesto MP463 en el sitio activo de Mpro. En la parte izquierda se muestra la interacción tridimensional del compuesto con los residuos de aminoácidos en el sitio activo de Mpro. Las líneas amarillas indican interacciones por puentes de hidrógeno. En la parte derecha se muestran las interacciones más importantes con los residuos del sitio activo.

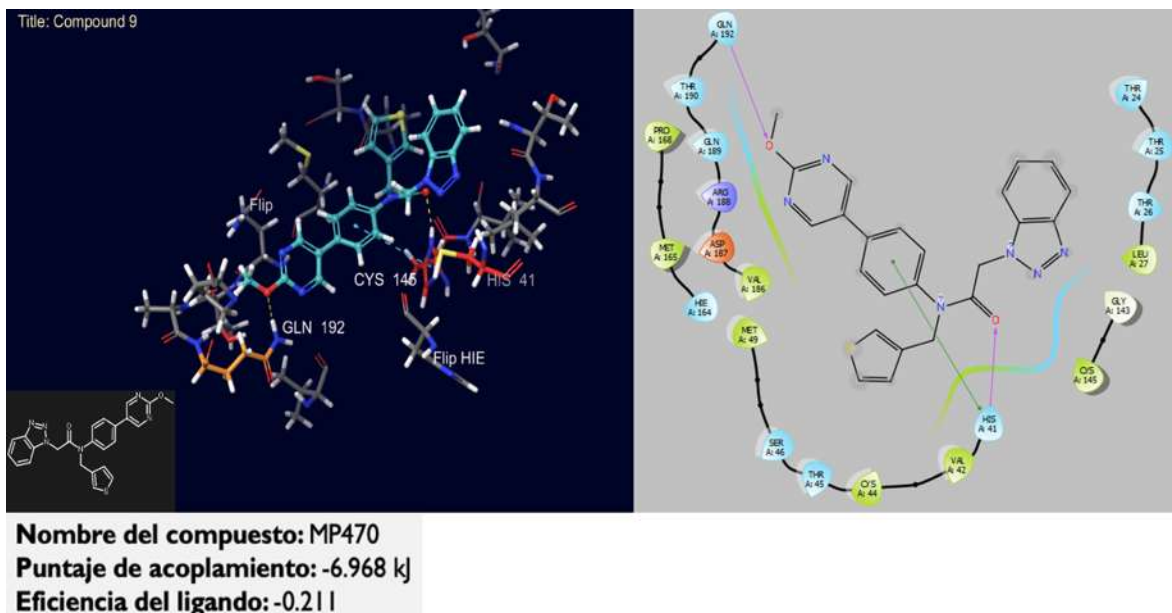


Figura 48. Acoplamiento del compuesto MP470 en el sitio activo de Mpro. En la parte izquierda se muestra la interacción tridimensional del compuesto con los residuos de aminoácidos en el sitio activo de Mpro. Las líneas amarillas indican interacciones por puentes de hidrógeno, las azules indican interacciones de apilamiento π - π . En la parte derecha se muestran las interacciones más importantes con los residuos del sitio activo.

10.17 MODELO DE FARMACÓFORO BASADO EN CARACTERÍSTICAS MOLÉCULARES DE LOS MEJORES CON MEJOR ACOPLAMIENTO CON Mpro

La orientación y conformación de los compuestos con el scaffold más frecuente (ID129), proveniente de la colección de compuestos reportada por Turlington *et al.* (2013), junto con el análisis de las mejores poses de acoplamiento de los compuestos MP462, MP463 y MP470, se tomaron en cuenta para recolectar características moleculares a incluir en un modelo de farmacóforo. De acuerdo con la IUPAC, un farmacóforo es “*un conjunto de rasgos estéricos y electrónicos necesarios para asegurar las óptimas interacciones moleculares con un blanco biológico específico y activar o bloquear su respuesta biológica*”. Por tanto, una vez acoplados los compuestos, se recopilaron las características farmacofóricas que se muestran en la figura 49. El par de anillos aromáticos en los extremos y otro dentro de la molécula aportan al compuesto volumen y un acomodo estérico favorable, mientras que las esferas verdes indican características de hidrofobicidad y las esferas amarillas indican regiones o características de átomos capaces de aceptar interacciones por puentes de hidrógeno.

Características para el modelo de farmacóforo

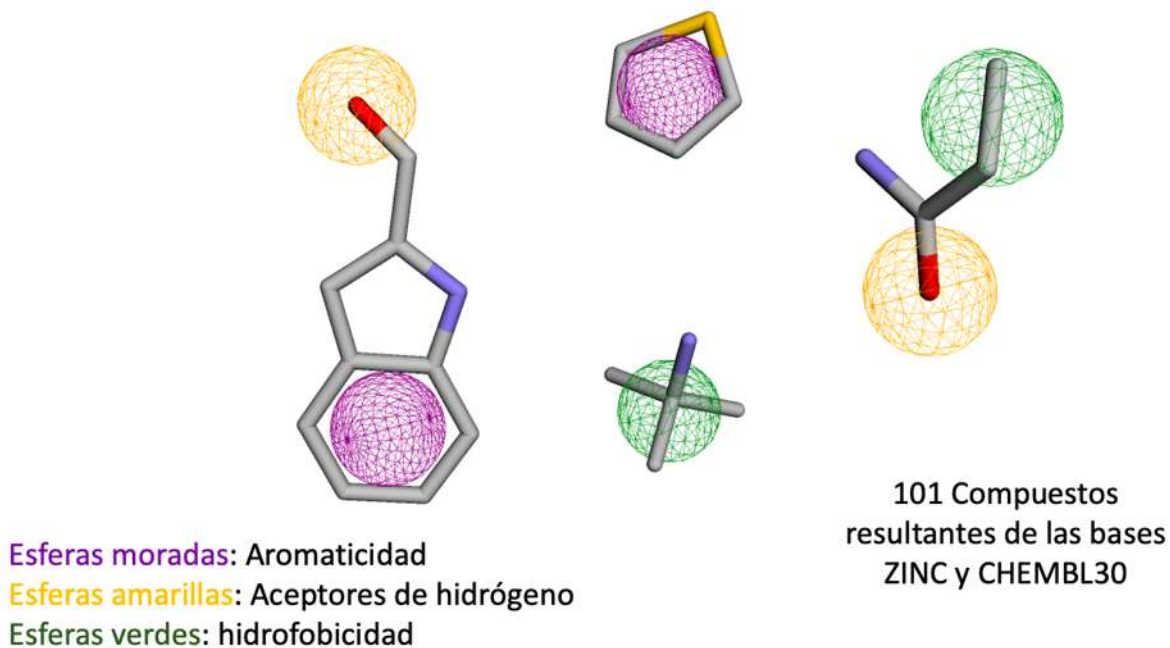
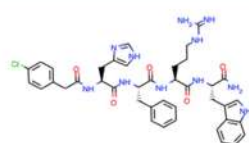
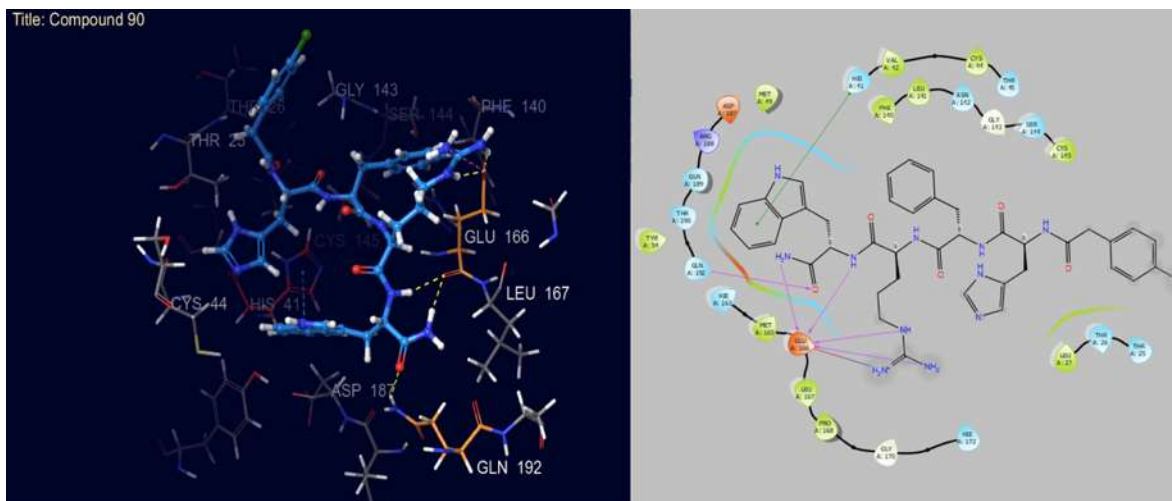


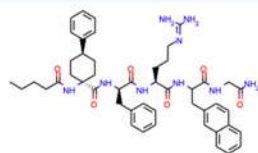
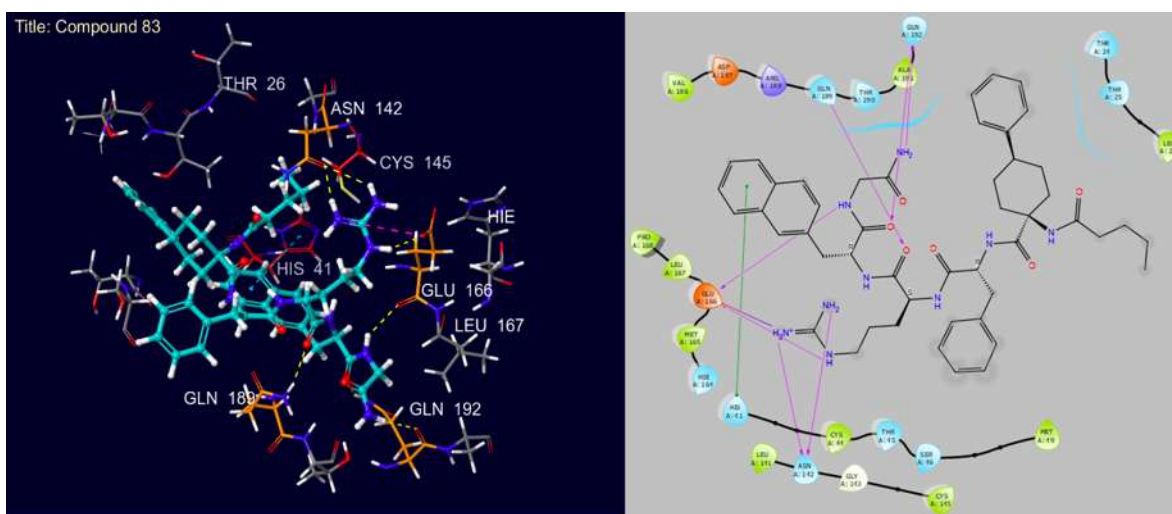
Figura 49. Modelo de farmacóforo propuesto con base en las características moleculares provenientes de los compuestos que presentaron un mejor acoplamiento con el sitio activo de Mpro. Las características se consideraron importantes por su aporte en las interacciones con residuos del sitio activo y por la estereoquímica del compuesto.

Con el modelo de farmacóforo generado se cribaron las bases de datos ZINC (<https://zinc.docking.org>) y ChEMBL30 (<https://www.ebi.ac.uk/chembl/>) para la búsqueda de compuestos que tuvieran en su estructura las características propuestas en el farmacóforo. Como resultado del cribado se obtuvieron las estructuras de 101 compuestos con dichas características, 61 compuestos se obtuvieron de la quimioteca de ZINC, mientras que 40 compuestos se obtuvieron de la quimioteca ChEMBL30. Esta nueva colección de compuestos se llevó a ensayos de acoplamiento molecular con el protocolo establecido. La estructura química y las poses de los compuestos con mejor acoplamiento se muestran en las figuras 50, 51 y 52. Además, se muestra la orientación y conformación de los compuestos en el sitio activo de la enzima, mediante una representación de superficie de la enzima (figura 3 del material suplementario).



COMP90 (CHEMBL92481)
Puntaje de acoplamiento: -12.228 kJ
Eficiencia del ligando: -0.208

Figura 50. Acoplamiento del compuesto COMP90 (CHEMBL92481) en el sitio activo de Mpro. En la parte izquierda se muestra la interacción tridimensional del compuesto con los residuos de aminoácidos en el sitio activo de Mpro. Las líneas amarillas indican interacciones por puentes de hidrógeno, las azules indican interacciones de apilamiento π - π , y las rosas indican puentes salinos. En la parte derecha se muestran las interacciones más importantes con los residuos del sitio activo.



COMP83 (CHEMBL199740)
Puntaje de acoplamiento: -10.697 kJ
Eficiencia del ligando: -0.170

Figura 51. Acoplamiento del compuesto COMP83 (CHEMBL199740) en el sitio activo de Mpro. En la parte izquierda se muestra la interacción tridimensional del

compuesto con los residuos de aminoácidos en el sitio activo de Mpro. Las líneas amarillas indican interacciones por puentes de hidrógeno, las azules indican interacciones de apilamiento π - π , y las rosas indican puentes salinos. En la parte derecha se muestran las interacciones más importantes con los residuos del sitio activo.

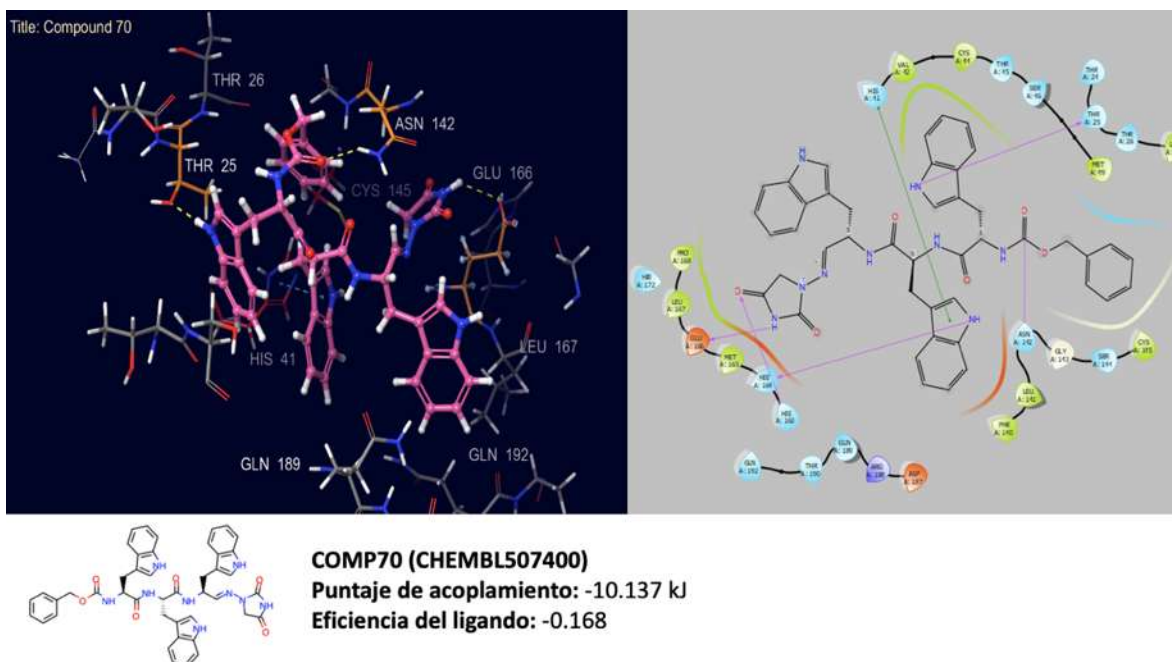


Figura 52. Acoplamiento del compuesto COMP70 (CHEMBL507400) en el sitio activo de Mpro. En la parte izquierda se muestra la interacción tridimensional del compuesto con los residuos de aminoácidos en el sitio activo de Mpro. Las líneas amarillas indican interacciones por puentes de hidrógeno, las azules indican interacciones de apilamiento π - π , y las rosas indican puentes salinos. En la parte derecha se muestran las interacciones más importantes con los residuos del sitio activo.

Los compuestos con mejor acoplamiento resultaron provenientes de la base de datos química ChEMBL, siendo el compuesto con identificador COMP90 (CHEMBL92481) el que resultó con el mejor puntaje de acoplamiento de -12.288 kJ. Este compuesto tiene una orientación y conformación en el sitio activo, de tal forma que ocupa el “bolsillo” casi en su totalidad (Fig 3, MS) y formando interacciones de tipo puente de hidrógeno con los residuos Glu166 y Gln192, además de una interacción de apilamiento π - π con la His41 del sitio catalítico. Por otro lado, el compuesto con el identificador COMP83 (CHEMBL199740) obtuvo el segundo mejor puntaje de acoplamiento de -10.697 kJ. Su orientación y conformación en la cavidad del sitio activo es apropiada y forma interacciones de tipo puente de hidrógeno con los residuos Asn142, Gln189, Glu166, Gln192, una

interacción de puente salino con Glu166 y una interacción π - π con His41. Por último, el compuesto COMP70 (ChEMBL507400), obtuvo el tercer mejor puntaje de acoplamiento de -10.137 kJ, formando interacciones por puente de hidrógeno con Asn142, Thr25, Glu166, His163, His164 y una interacción de apilamiento π - π con His41. Estos tres compuestos, presentan características estéricas y electrónicas favorables para interactuar con los residuos del sitio activo de Mpro, tienen una orientación y conformación aceptable en la cavidad de la enzima (Fig 3, MS), y forman interacciones importantes con residuos clave, siendo moléculas con el potencial de ocupar el sitio activo e inhibir la actividad de proteasa de Mpro. El acoplamiento de los compuestos permite obtener una idea de la posible interacción entre los diferentes átomos de la molécula y los residuos del sitio activo de la enzima. El paso siguiente fue determinar la estabilidad de estas interacciones, mediante simulaciones atomísticas de dinámica molecular en al menos uno de los compuestos con respecto al tiempo utilizando un campo de fuerza y observando si las interacciones predichas se mantienen y si el compuesto o los compuestos podrían ser potenciales inhibidores de la enzima.

10.18 SIMULACIÓN ATOMÍSTICA POR DINÁMICA MOLECULAR DE COMPUESTO COMP83 CON EL SITIO ACTIVO DE Mpro

Para evaluar la estabilidad de los compuestos en el sitio activo de Mpro en un ambiente fisiológico se construyó el sistema COMP83/Mpro como primer complejo entre uno de los compuestos resultantes del análisis y el sitio activo de la enzima. Este complejo fue simulado mediante la técnica de simulación atomística por dinámica molecular durante una escala de tiempo de 150 ns y subsecuentemente se analizaron los cambios conformacionales en el entorno del ligando. El complejo de proteasa y COMP83 se observa en la figura 53, donde se presenta la estructura de Mpro en su forma dimérica. El monómero A se observa en color rosa mientras que el monómero B se ilustra en color azul. Se observan además las dos moléculas de COMP83 en color amarillo la enlazada al monómero A y en color anaranjado la enlazada al monómero B. Además, se ilustran los residuos His41 y Cys145 del sitio activo de Mpro en la representación de esferas.

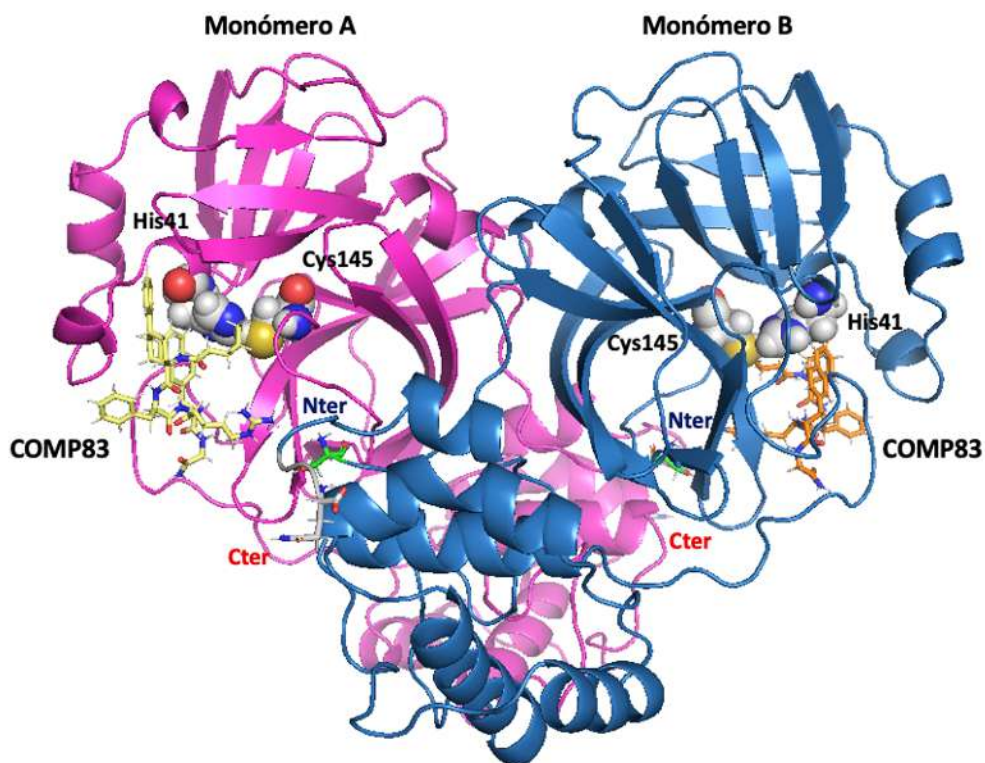
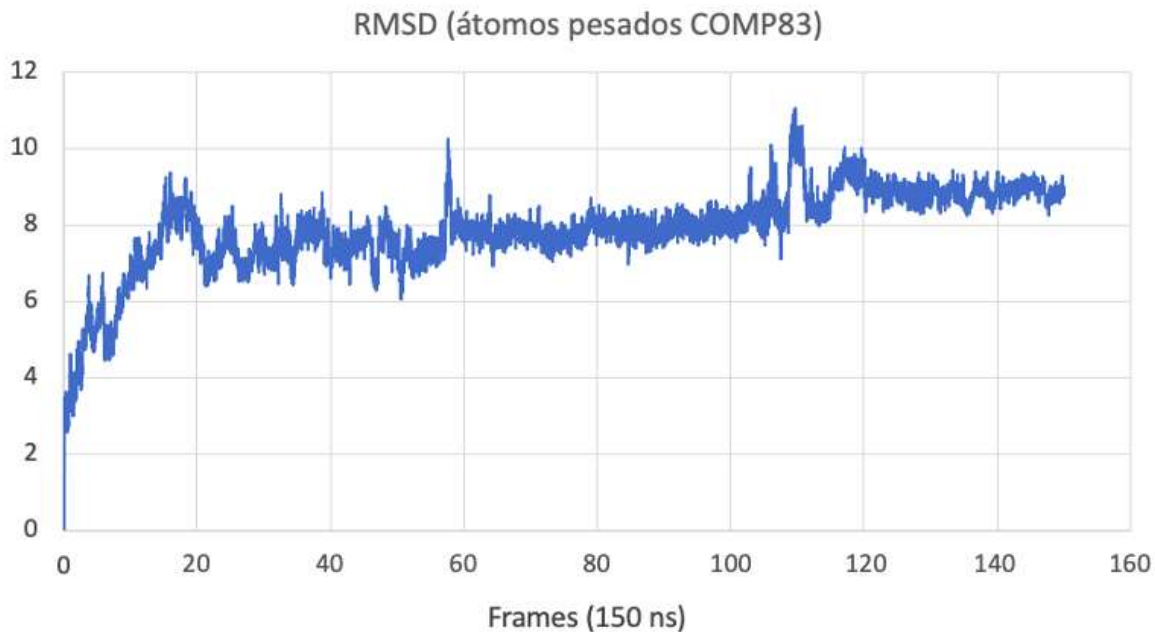


Figura 53. Sistema COMP83/Mpro en su forma dimérica. La estructura terciaria de Mpro se tomó de la estructura dilucidada por difracción de rayos X (PDB: 7JU7). Se observan de color rosa el monómero A, de color azul el monómero B, de color amarillo y en varillas el compuesto COMP83 orientado hacia el sitio activo del monómero A, de color naranja y en varillas el compuesto COMP83 orientado hacia el monómero B, a manera de esferas los residuos His41 y Cys145, de color blanco y en varillas la región C-terminal (Cter) y de color verde y en varillas la región N-terminal Nter de cada monómero.

Durante los 150 ns de simulación, el COMP83 logró permanecer de manera estable en el sitio activo de la enzima. Para el caso del monómero A, en el comienzo de la simulación, COMP83 se movió y fluctuó brevemente en la cavidad del sitio activo y poco después se estabilizó, fluctuando en un valor de 8 Å relativo a la pose molecular inicial, de acuerdo con el cálculo general de RMSD que se muestra en el gráfico 6. Una vez que el compuesto se estabilizó formó interacciones importantes con los residuos Glu166, Gln189 y Met49, ocupando por completo la cavidad en el sitio activo de la enzima. Estas interacciones pueden observarse en la figura 54, donde se observa el sistema de COMP83/Mpro y un acercamiento a lo que sucede en el monómero A, durante un paso final en la simulación.



Grafica 6. Cálculo de RMSD general de los átomos pesados de COMP83 durante 150 ns de simulación.

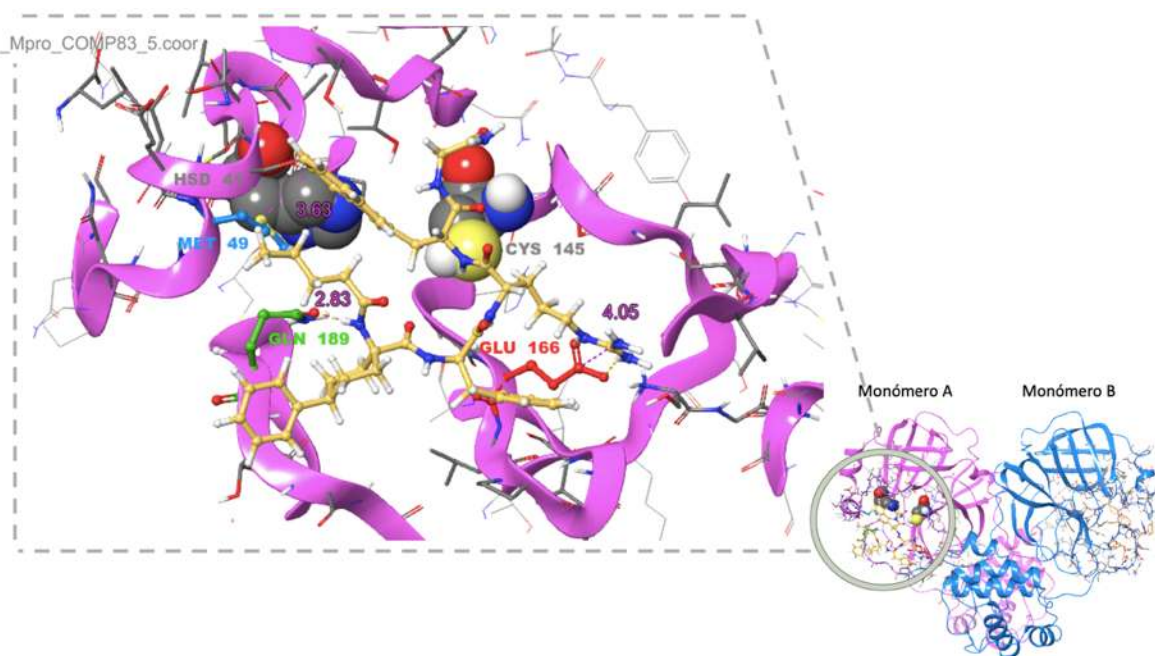


Figura 54. Acercamiento que ilustra las interacciones entre COMP83 y los residuos de aminoácidos del monómero A en el sitio catalítico de Mpro. Los residuos y las interacciones importantes, así como su distancia con átomos de la enzima se muestran con diferentes colores. De color rojo CD_Glu166/C48_COMP83, de color verde OE1_Q189/N6_COMP83 y de color azul SD_Met49/Ring1_COMP83. Las diferentes distancias se muestran en color rosa. A manera de esferas se muestran los residuos de His41 y Cys145 del sitio activo.

COMP83 es un péptido modificado con residuos de Phe, Arg y Gly (Figura 55). La interacción de mayor jerarquía en el complejo Mpro/COMP83 se establece entre Glu166 y un grupo R de residuo de Arginina en la estructura del compuesto. Esta interacción carga-carga es de gran estabilidad de acuerdo a la medición de la distancia entre el carbono Delta de Glu166 y el C48 (Figura 54 y 55) del compuesto durante los 150 ns de simulación representados en la gráfica 7, donde se puede observar la fluctuación en las interacciones más importantes entre el compuesto y los residuos de la proteasa, siendo la interacción entre Glu166 (en color rojo en la figura 54) y el C48 la que se muestra de color naranja y que se conserva a lo largo de toda la trayectoria. Este puente salino, al tener una energía de interacción alta, permite que el compuesto permanezca “anclado” adentro del sitio catalítico de la enzima, formando otras interacciones importantes como Gln189 (en color verde en la figura 54) y el N6 del compuesto mediante la formación de un puente de hidrógeno y el residuo de Met49 (en color azul en la figura 54) interaccionando entre el átomo de azufre con el par de anillos conjuntos del compuesto (figura 54 y 55). La estabilidad de estas interacciones puede observarse mediante la medición de las distancias que se muestra en la gráfica 7, donde se observa que la interacción más estable entre las últimas dos descritas es la del Oxígeno épsilon de Gln89 y el N6 del compuesto, seguida de la interacción entre el azufre delta de Met49 y los anillos conjugados del compuesto.

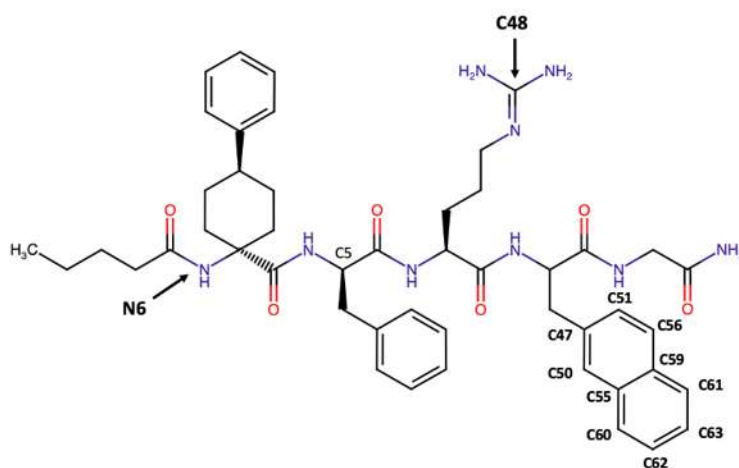


Figura 55. Representación 2D de COMP83. En negritas se indican los átomos con los que COMP83 forma interacciones con los residuos Glu166, Gln189 y Met49 del sitio activo de Mpro.

Estas interacciones sugieren que la pose predicha por el cálculo de acoplamiento es estable, tanto para el ligando como para la proteína en el monómero A del sistema, cuyo RMSD general se muestra en la gráfica 1 del material suplementario. Por otra parte, para el monómero B, el compuesto se mueve y estabiliza brevemente en la simulación, aunque prontamente pierde la interacción entre la Glu155 y el residuo modificado de Arginina del compuesto, lo que parece desanclar a COMP83 del bolsillo activo. Este lleva a cabo un re-arreglo fuera de la enzima, pero para los 150 ns de simulación aun no logra encontrar un punto de estabilización. En lo sucedido en el monómero A, se observa además que, aunque COMP83 no forma interacciones aparentes con los residuos His41 y Cys145 del sitio activo, debido su tamaño, el COMP83 logra, mediante las interacciones ya descritas con otros residuos de aminoácidos, bloquear en su totalidad el bolsillo catalítico durante los 150 ns de la simulación, funcionando como un potencial inhibidor de Mpro.



Gráfica 7. Cálculo de distancias en Amstrongs de las interacciones entre átomos pesados de COMP83 y los residuos Glu166, Gln189 y Met49 del sitio activo de Mpro durante 150 ns de simulación. De color naranja se grafican las distancias entre CD_Glu166/C48_COMP83, de color amarillo la distancia entre OE1_Q189/N6_COMP83 y de color verde SD_Met49/Ring1_COMP83.

10.19 VISUALIZACIÓN DEL ESPACIO QUÍMICO BASADO EN LAS SEIS PROPIEDADES FISICOQUÍMICAS MW, HBAcc, HBDon, nRotB, ALogP y TPSA DE COMPUESTOS CON AFINIDAD POR LA PROTEASA DE TIPO PAPAÍNA PLPRO

Para la proteasa PLpro, se graficó de manera similar al caso de Mpro, el espacio químico correspondiente a compuestos con afinidad por esta enzima utilizando la plataforma PUMA (Fig. 57 y 58). En este caso los componentes PCa y PCb recuperaron el 81% de la covarianza de los datos y el compuesto de referencia es GRL0617, un compuesto propuesto por Fu *et al.*, 2021 como inhibidor de PLpro resultante del cribado de una librería de fármacos aprobados y algunas otras moléculas propuestas como inhibidores de PLpro en el artículo "*The complex structure of GRL0617 and SARS-CoV-2 PLpro reveals a hot spot for antiviral drug discovery*". GRL0617 tiene un IC50 reportado de 2.1 μM y su estructura y propiedades fisicoquímicas se muestra en la figura 56. De acuerdo a los valores de la tabla 18, las propiedades fisicoquímicas TPSA, nHBAcc y nHBDon tuvieron un mayor peso en la contribución de la formación del PCa y una relación inversa con la propiedad ALogP, por lo que compuestos posicionados hacia valores negativos en el PCA tendrán valores de TPSA, nHBAcc y nHBDon mayores, sinónimo de hidrofiliidad y polaridad, mientras que para la formación del PCb las propiedades fisicoquímicas que contribuyeron en mayor medida fueron nRotB, MW y ALogP, lo que significa que compuestos posicionados hacia valores positivos en el PCb tendrán valores mayores de estas propiedades que son indicativas de flexibilidad, tamaño, y solubilidad.

Por su parte, el compuesto de referencia GRL0617 fué reportado por Ratia *et al.*, en 2008 como un inhibidor que interactúa de forma no covalente resultante del cribado de alto rendimiento (High Throughput screening) de una colección de 50,080 compuestos. GRL0617 se une al dominio catalítico de PLpro de SARS-CoV-2 y compete en la unión con otras moléculas como ISG15 y Ubiquitina, moléculas en las que PLpro tiene actividad, inhibiendo la actividad de PLpro con un IC50 de 2.1 μM . De manera que así como para el caso de la proteasa Mpro, se coleccionaron las moléculas cercanas al compuesto de referencia bajo la premisa de que compuestos

cercanos en el espacio químico tienen propiedades fisicoquímicas similares. En la parte izquierda de la Figura 57 se puede observar el espacio químico para compuestos con afinidad por PLpro realizado mediante PCA en la plataforma PUMA y en la parte derecha la selección de compuestos cercanos provenientes de los grupos con afinidad reportada “PL” y fármacos reposicionados “RPL”, mientras que la figura 58 muestra un acercamiento en el espacio químico y la selección de los compuestos cercanos a GRL0617 provenientes del grupo con actividad biológica reportada “ACT”.

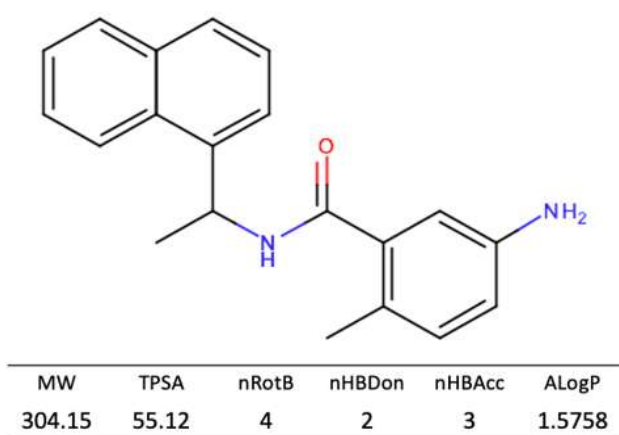


Figura 56. Estructura química de GRL0617. Los valores de sus propiedades fisicoquímicas se indican debajo del nombre del compuesto.

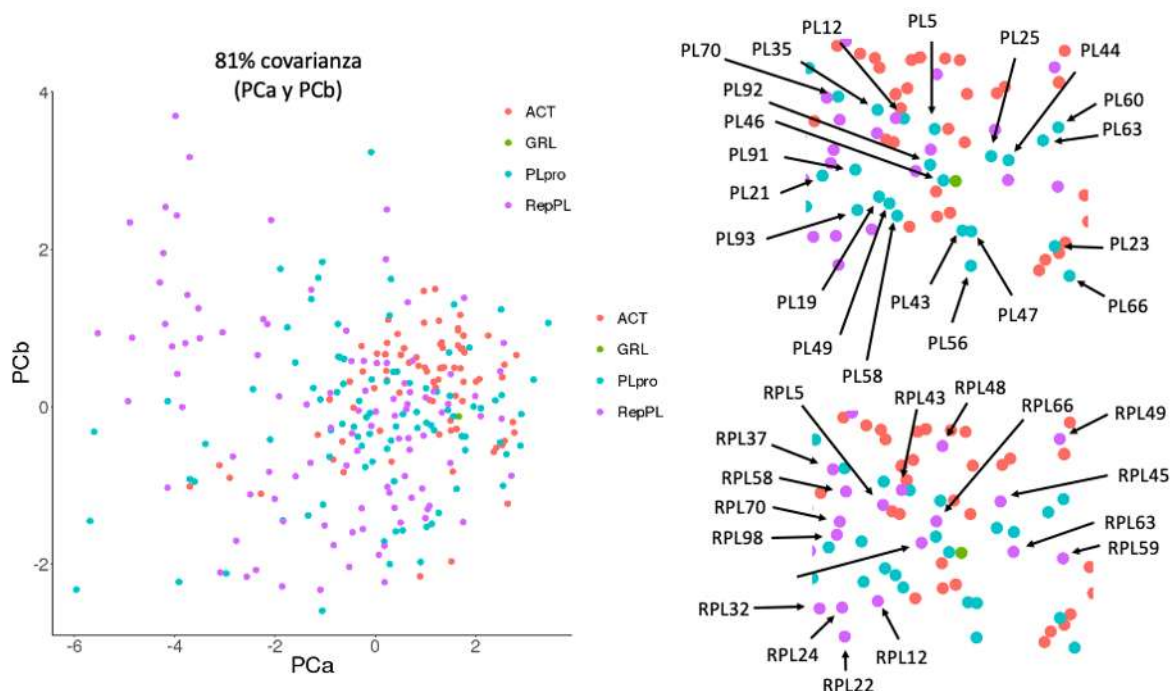


Figura 57. Representación visual del espacio químico para la base de datos de compuestos afines a PLpro mediante análisis de PCA. Los componentes principales PCa y PCb recuperaron el 81% de la covarianza de los datos. Los compuestos están divididos en compuestos reportados con afinidad por la proteasa (esferas azules), compuestos con actividad biológica reportada (esferas naranja) y fármacos reposicionados (esferas moradas). Se incluyen el compuesto de referencia GRL0617 (esfera verde).

Tabla 18. Contribución de las seis propiedades fisicoquímicas a cada PC en el espacio químico de PLpro.

	PCa	PCb	PCc	PCd	PCe	PCf
MW	-0.379	0.527	0.475	-0.055	-0.550	0.218
TPSA	-0.483	-0.158	0.158	0.138	0.552	0.627
nRotB	-0.308	0.606	-0.713	-0.033	0.164	-0.041
nHBDon	-0.400	-0.327	-0.212	0.725	-0.363	-0.177
nHBAcc	-0.483	-0.012	0.326	-0.183	0.339	-0.715
ALogP	0.367	0.472	0.300	0.647	0.346	-0.120

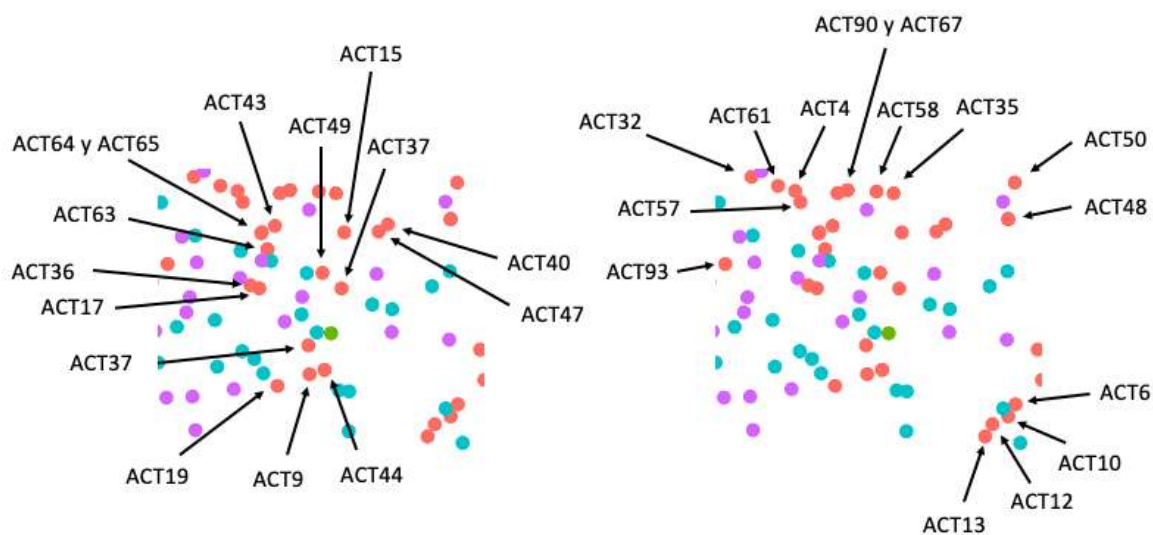


Figura 58. Acercamiento en el espacio químico de compuestos con afinidad por PLpro, en esta imagen se muestran los compuestos con actividad biológica que se recopilaron para probar por acoplamiento molecular.

Además del espacio químico en dos dimensiones generado en PUMA, se generó el espacio químico tridimensional en Data Warrior, lo que permite visualizar con mejor precisión los compuestos de los diferentes grupos que se encuentran cercanos a GRL0617. Las propiedades utilizadas para graficar el espacio químico en DataWarrior fueron las mismas que para graficarlo en la plataforma PUMA (MW, HBAcc, HBDOn, nRotB, ALogP y TPSA). En la figura 59 se observa el espacio químico tanto bidimensional como tridimensional generado en DataWarrior, donde se observan con diferentes colores y figuras los diferentes grupos de compuestos. En la figura 60 se observa en la parte izquierda la selección de compuestos cercanos a GRL0617 provenientes del grupo de compuestos con afinidad por esta enzima “PL”, mientras que de lado derecho se observa la selección de compuestos provenientes del grupo de fármacos reposicionados “RPL”, por otro lado, en la figura 61 se observa la selección de compuestos cercanos a GRL0617 provenientes del grupo de compuestos con actividad biológica conocida para esta enzima “ACT”.

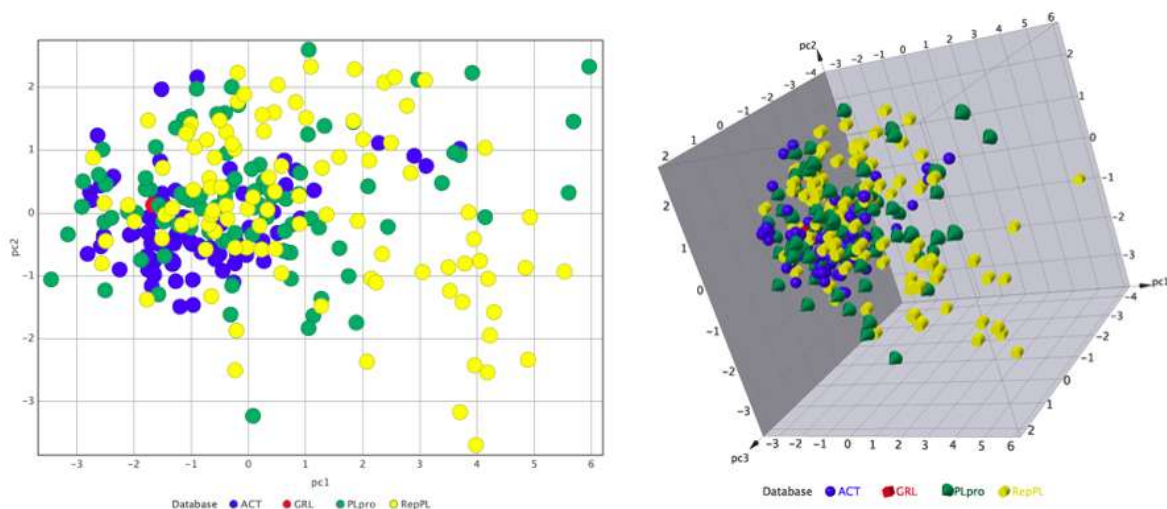


Figura 59. Representación visual 2D y 3D del espacio químico para la base de datos de compuestos afines a PLpro mediante análisis de PCA en Datawarrior. Los componentes principales PC1, PC2 y PC3 recuperaron el 88% de la covarianza de los datos. El compueseto de referencia GRL0617 se muestra en rojo.

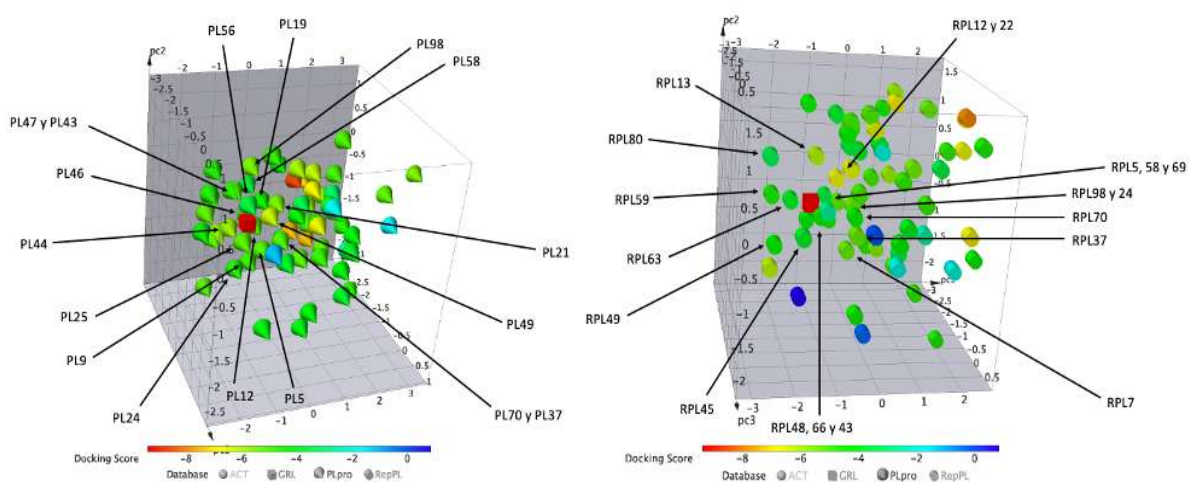


Figura 60. Representación 3D del espacio químico para el grupo de compuestos con afinidad por la proteasa PLpro mediante PCA en Datawarrior. La escala de color es representativa del score de acoplamiento obtenido en ensayo de acoplamiento molecular. El primer gráfico muestra los compuestos con afinidad por la proteasa cercanos a GRL0617, el gráfico de la derecha muestra los fármacos reposicionados cercanos a GRL0617.

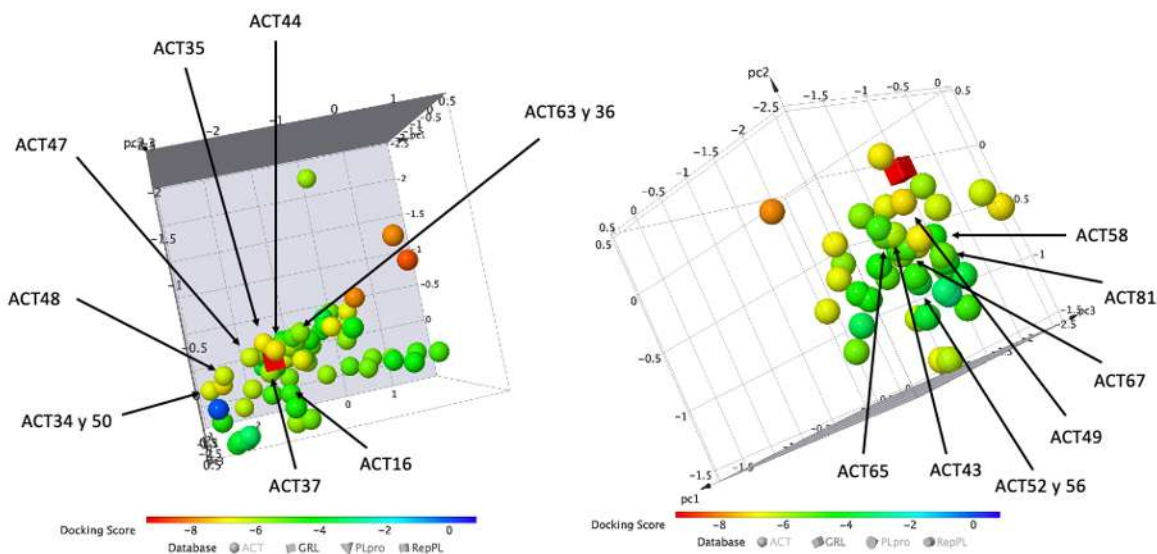
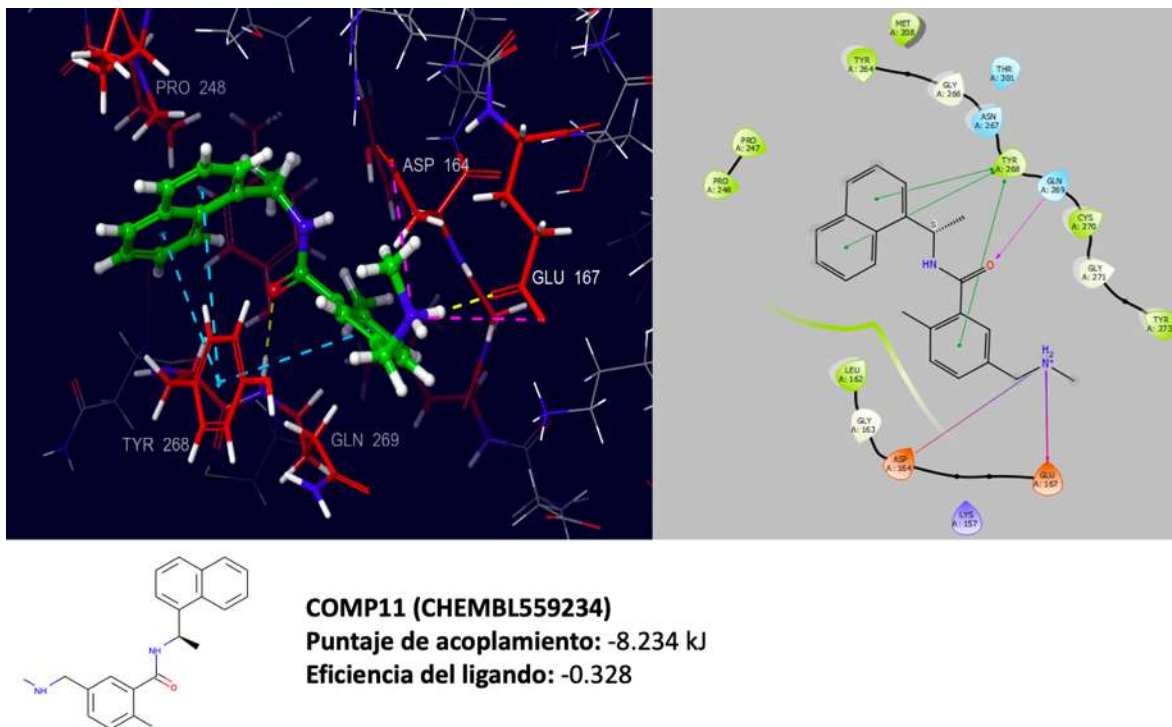
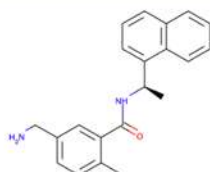
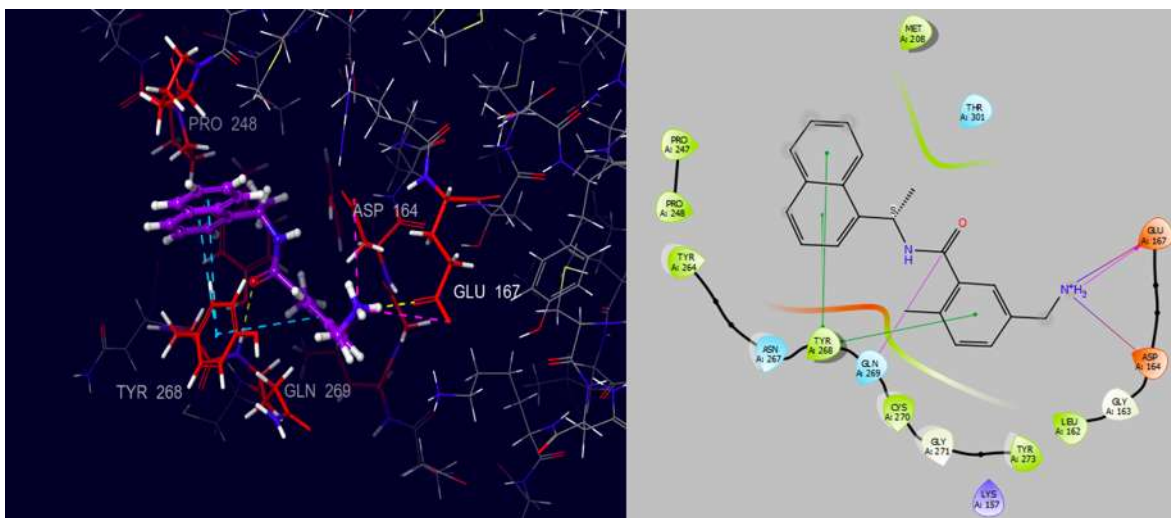


Figura 61. Representación 3D del espacio químico para el grupo de compuestos con actividad biológica reportada contra la proteasa PLpro mediante PCA en Datawarrior. La escala de color es representativa del score de acoplamiento obtenido en ensayo de acoplamiento molecular.

Después del PCA tanto en PUMA como en DataWarrior se identificaron 81 compuestos posicionados con cercanía a GRL0617 en el espacio químico. Esta nueva colección de compuestos se sometió a simulaciones de acoplamiento molecular para analizar sus interacciones y posibles modos de unión con el sitio activo de la enzima, siendo los compuestos COMP11 y COMP2 aquellos que resultaron con los mejores puntajes de acoplamiento y acomodo en la cavidad de la enzima, el acomodo de los compuestos en dicha cavidad puede observarse en la figura 4 de material suplementario donde la cavidad se presenta como una superficie. COMP11 y COMP2 fueron propuestos por Ghosh *et al.* En 2009 como inhibidores de PLpro de SARS-CoV-1, resultado del cribado de alto rendimiento de una colección de 50,080 compuestos con diversas estructuras seguido de un análisis de estructura actividad y proceso de optimización para mejorar la potencia de los compuestos en células Vero E6 infectadas con SARS-CoV. Ambos compuestos al tener estructuras altamente similares con tan solo un átomo de Carbono de diferencia forman el mismo tipo de interacciones con el sitio activo de la enzima, por su parte COMP11 obtuvo un puntaje de acoplamiento de -8.234 kJ, mientras que COMP2 obtuvo un puntaje ligeramente menor de -8.154 kJ. Ambos forman interacciones por puente de hidrógeno con Glu167 y Gln269, hasta tres

interacciones de apilamiento π - π entre los anillos del compuesto y el residuo Tyr268, y dos interacciones por puente salino con los residuos Glu167 y Asp164, todo esto mostrando un excelente acomodo en el sitio activo de la enzima (Fig. 62 y 63).





COMP2 (CHEMBL561061)
Puntaje de acoplamiento: -8.219 kJ
Eficiencia del ligando: -0.340

Figura 63. Acoplamiento del compuesto COMP2 (CHEMBL561061) en el sitio activo de PLpro. En la parte izquierda se muestra la interacción tridimensional del compuesto con los residuos de aminoácidos en el sitio activo. Las líneas amarillas indican interacciones por puentes de hidrógeno, las azules indican interacciones de apilamiento π - π , y las rosas indican puentes salinos. En la parte derecha se muestran las interacciones más importantes con los residuos del sitio activo.

11.- DISCUSIÓN

COVID-19 es una enfermedad respiratoria infecciosa potencialmente mortal ocasionada por el coronavirus SARS-CoV-2 (Síndrome respiratorio agudo severo por coronavirus-2), identificado por primera vez en la ciudad china de Wuhan en noviembre de 2019 (Lippi *et al.*, 2023). Pronto después de su aparición el virus se propagó rápidamente, siendo declarado una emergencia sanitaria mundial el 30 de enero del 2020 por la Organización Mundial de la Salud (OMS). Misma institución que poco después, el 11 de marzo del mismo año declararía a la enfermedad COVID-19 como pandemia (Cucinotta & Vanelli, 2020). Los síntomas más comunes de COVID-19 reportados en la literatura incluyen fiebre, tos, fatiga y dificultad para respirar, que son similares a otras infecciones virales como la gripe estacional (Chams *et al.*, 2020). No obstante, la facilidad de contagio junto con la severidad de la enfermedad, han permitido que el número de muertes por COVID continúe aumentando, dando como resultado complicaciones en los sectores salud, económico, educativo, político, entre otros.

El descubrimiento y desarrollo de fármacos es un proceso largo que puede durar en promedio entre 10 y 15 años de manera convencional (Gurung *et al.*, 2021). La etapa de descubrimiento y optimización es por lo general la más tardada y el proceso en su totalidad es bastante costoso, dependiendo de varios factores (entre 1,395 - 2,558 millones de dólares, DiMasi *et al.*, 2016). En los últimos años, el diseño de fármacos asistido por computadora (DIFAC) ha contribuido de manera importante en la búsqueda, diseño y desarrollo de fármacos para distintas enfermedades, minimizando el tiempo y el costo del proceso de manera significativa (Gurung *et al.*, 2021).

En el contexto de COVID-19, la importancia del DIFAC radica en la confianza que los investigadores han depositado en estas técnicas computacionales para la rápida identificación de moléculas prometedoras candidatas a fármacos contra diversas dianas farmacológicas implicadas en la patogénesis de SARS-CoV-2 (Gurung *et al.*,

2021). De manera general el proceso para el descubrimiento de un nuevo fármaco comienza con la identificación de la diana terapéutica adecuada, la validación de dicha diana, el descubrimiento de compuestos llamados comúnmente “*hits*”, la optimización de los compuestos para hacerlos “*líderes*” y los estudios preclínicos y clínicos (Vohora & Singh., 2018).

SARS-CoV-2 es un coronavirus de ARN monocatenario en sentido positivo de aproximadamente 30 kb, que utiliza la maquinaria celular del huésped para ejecutar diversos procesos patogénicos como la entrada viral, la replicación genómica y la síntesis proteica para el ensamblaje de nuevos viriones (Amin & Jha., 2020). Al igual que otros coronavirus que surgieron con anterioridad como el SARS-CoV y el MERS-CoV, el genoma del SARS-CoV-2 codifica para cuatro proteínas estructurales, dieciséis proteínas no estructurales (nsps) y algunas otras proteínas accesorias (Goyal & Goyal., 2020). Dentro de las nsps se encuentran la proteasa principal “Mpro” y la proteasa tipo papaína “PLpro”, las cuales desempeñan un papel fundamental en el ciclo infeccioso del virus, estando involucradas en la producción de enzimas que forman el complejo de replicación-transcripción del virus (CRT) que a su vez está encargado de la producción de ARN subgenómicos (Dai *et al.*, 2020). Lo anterior las convierte por tanto en blancos terapéuticos prometedores para el diseño y desarrollo de agentes antivirales contra COVID-19 (Goyal & Goyal., 2020).

La pronta dilucidación de la estructura tridimensional de las proteínas que funcionarían como dianas terapéuticas, como Mpro y PLpro, así como la búsqueda, diseño y optimización de miles de compuestos candidatos a ensayos preclínicos contra COVID-19 son el resultado de la combinación de métodos de DIFAC, tanto de aquellos basados en la estructura como en el ligando (Huang *et al.*, 2010). Mientras que el diseño de fármacos basado en la estructura parte de conocer la estructura tridimensional del receptor diana y sus sitios activos para comprender la interacción molecular entre el receptor y el ligando, el diseño de fármacos basado en el ligando depende del conocimiento de moléculas que interactúan con el receptor determinado (Huang *et al.*, 2010). Técnicas de DIFAC como el modelado de proteínas, el cribado virtual, el acoplamiento molecular, el diseño de modelos de

farmacóforo, algunos métodos de aprendizaje estadístico, las simulaciones atomísticas de dinámica molecular, entre otras., han sido empleados para la identificación de miles de moléculas con afinidad y potencial actividad farmacológica contra COVID-19 a partir de bases de datos químicas también llamadas quimiotecas (Gurung *et al.*, 2021).

En este sentido, una gran cantidad de información concerniente a la estructura y propiedades químicas de moléculas con afinidad predicha por métodos computacionales por las proteasas de SARS-CoV-2, quedó disponible y dispersa en los diferentes gestores de información científica. El uso de la tecnología para la gestión de información es en la actualidad una parte fundamental del proceso de descubrimiento de fármacos, y la quimioinformática es la disciplina que combina esos recursos de información para transformar datos en información e información en conocimiento con el propósito de tomar mejores decisiones, de manera más rápida y asertiva en la identificación y organización de moléculas líderes (Brow & Bristol, 1998). Quimioinformática es un término genérico que abarca el diseño, la creación, la organización, la gestión, la recuperación, el análisis, la difusión, la visualización y el uso de información química (Gasteiger & Funatsu., 2006).

Un concepto clave en quimioinformática es el espacio químico, que puede definirse como un espacio matemático M-dimensional ocupado por moléculas y donde la posición de cada molécula en el espacio estará definida por sus propiedades (Arús-Pous *et al.*, 2019). En el DIFAC, estas propiedades están por lo general asociadas a que una molécula sea “*de tipo fármaco*” incluyendo propiedades fisicoquímicas como MW, clogP, nHBDon, nHBAcc, nRotB, TPSA, entre otras posibles, que favorezcan las características ADME (absorción, distribución, metabolismo y excreción) de los compuestos (Medina-Franco *et al.*, 2021). En este contexto, el objetivo de la quimioinformática es encontrar una forma racional de representar el espacio químico y navegar por él, convirtiendo datos en información e información en conocimiento relevante para la búsqueda y desarrollo de nuevos compuestos biológicamente activos (Varnek & Baskin., 2011).

El presente trabajo adopta un enfoque quimioinformático, partiendo de datos e información publicada de moléculas con afinidad predicha por métodos computacionales contra las proteasas de SARS-CoV-2, generando visualizaciones del espacio químico definido por descriptores moleculares de importancia farmacéutica y culminando con la combinación de una serie de metodologías de DIFAC como son el modelado y acoplamiento molecular, el diseño de modelos de farmacóforo y las simulaciones atomísticas de dinámica molecular para proponer moléculas nuevas candidatas a ser probadas en el laboratorio para inhibir a sus enzimas correspondientes.

A pocas semanas después del primer brote de COVID-19 en diciembre del 2019, grupos científicos trabajaron arduamente para dilucidar las estructuras cristalográficas de proteínas clave en la replicación y ciclo infeccioso de SARS-CoV-2, entre ellas las proteasas Mpro y PLpro. A partir de dichas estructuras y con información sobre la investigación de las proteasas de los coronavirus pasados comenzó la búsqueda de agentes antivirales capaces de inhibir a las proteasas y así frenar el impacto de la enfermedad en las personas. Métodos computacionales para el cribado de moléculas por acoplamiento molecular fueron utilizados para proponer compuestos que se unieran con afinidad a los sitios activos de las proteasas, por lo que se recopilaron 97 artículos científicos con información concerniente a dichos compuestos, incluyendo además artículos que propusieran compuestos que se unieran a las proteasas Mpro y/o PLpro de SARS-CoV debido a su alto porcentaje de homología (96% para Mpro y 83% para PLpro) y fármacos aprobados, en un intento por ser reposicionados contra COVID-19.

El resultado de la búsqueda fue una quimioteca de 866 moléculas, cuya información provino de una variedad de grupos geográficamente dispersos, y con ello el origen de los compuestos, tanto natural como sintético, resultó diverso también. Das *et al.*, 2020 utilizaron por ejemplo un enfoque de acoplamiento molecular para identificar potenciales inhibidores de Mpro, mediante el cribado de compuestos provenientes de productos naturales, antivirales, antifúngicos, antinematodos y antiprotozoarios. Los compuestos identificados incluyen la rutina (compuesto natural), emetina

(antiprotozoario), hesperidina (compuesto natural) y el indinavir (antiviral) (Das *et al.*, 2020). Umesh *et al.*, 2021, identificaron el carnosol, arjunglucósido-1 y el rosmanol como inhibidores de Mpro a partir de examinar moléculas provenientes de especias de la India utilizando acoplamiento y simulaciones de dinámica molecular (Umesh *et al.*, 2021). Gentile *et al.*, 2020, examinaron una biblioteca de productos naturales marinos (MNP) con la ayuda de un modelo de farmacóforo, acoplamiento y dinámica molecular e identificaron 17 compuestos provenientes del alga parda *Sargassum spinuligerum* como potenciales inhibidores de Mpro (Gentile *et al.*, 2020). Delre *et al.*, 2020 utilizaron acoplamiento molecular para acoplar 688 fármacos en fase III y 1,702 fármacos en fase IV de ensayos clínicos, a la proteasa PLpro de SARS-CoV-2, encontrando 24 fármacos como potenciales inhibidores dentro de los cuales se encuentran vilanterol, arformeterol, atenolol (moduladores de receptores adrenérgicos), edoxaban y rivaroxaban (anticoagulantes), anagliptina, canaglifozina (antidiabético) entre otros., para considerarse candidatos a probarse en laboratorio contra PLpro (Delre *et al.*, 2020). Rao *et al.*, 2022 identificaron seis compuestos (Fonsecina, Piranonigrina-B, Nigerloxina, Flaviolina, Tensidiol A y Ocratoxina Beta) provenientes de una biblioteca de metabolitos secundarios de hongos con naturaleza aromática mediante acoplamiento molecular, sugiriendo que Fonsecina es un compuesto líder candidato a probarse experimentalmente para inhibir a PLpro de SARS-CoV-2 (Rao *et al.*, 2022). Además, algunos grupos utilizaron el enfoque de evaluar moléculas que pudieran funcionar como inhibidores duales contra ambas proteasas de SARS-CoV-2. Por ejemplo Mitra *et al.*, 2020 utilizó un modelo de farmacóforo y simulaciones de acoplamiento y dinámica molecular en la identificación de tres compuestos (nelfinavir, tipranavir y licochalcona-D) capaces de unirse con buena afinidad a los sitios activos de ambas enzimas (Mitra *et al.*, 2020). Rajpot *et al.*, 2020 por otra parte encontraron nueve compuestos provenientes de la base de datos ZINC (ZINC389747, ZINC1530637, ZINC895199, ZINC49153, ZINC113355, ZINC39089, ZINC20259, ZINC56568, ZINC53) también capaces de interactuar con los sitios activos de Mpro y PLpro de SARS-CoV-2 (Rajpot *et al.*, 2020).

El primer análisis de diversidad del presente trabajo comparó los compuestos reportados con afinidad por Mpro contra aquellos con afinidad por PLpro, encontrando una diferencia significativa en el tamaño de las moléculas, siendo aquellas con afinidad por Mpro las más grandes, con valores de MW, nHBacc, nHBDon y TPSA mayores. Esto permite que las moléculas con afinidad por Mpro formen más interacciones electrostáticas con su entorno lo que puede ser indicativo de mayor hidrofiliidad. Además, al graficar mediante un análisis de componentes principales (PCA) el espacio químico para ambos conjuntos de moléculas y aplicar las métricas de distancia euclídeana e índice de Tanimoto para analizar diversidad molecular basada en comparaciones sistemáticas por pares, se observó que los compuestos con afinidad por Mpro se encuentran más dispersos en el espacio químico comparados con los compuestos con afinidad por PLpro. Esto puede deberse a que existe una mayor cantidad de publicaciones en las que el blanco biológico a inhibir es la proteasa principal Mpro y por tanto el número de moléculas reportadas con afinidad por Mpro es mayor y más diverso que aquellas reportadas con afinidad por PLpro. Al teclear en el buscador Pubmed las palabras clave “SARS-CoV-2 Mpro inhibitors” la página arroja 1,170 publicaciones desde el año 2020 a la fecha, mientras que al teclear las palabras “SARS-CoV-2 PLpro inhibitors” para el mismo periodo el buscador arroja tan solo 323 publicaciones. La acción de la actividad catalítica de Mpro sobre las poliproteínas virales da lugar a la liberación y correcto funcionamiento de 12 proteínas efectoras no estructurales con un papel fundamental en la replicación y transcripción viral al escindir al menos 11 sitios alrededor de las regiones C-terminal y central de las poliproteínas virales con secuencia consenso Leu-Gln/Ser-Ala-Gly, y como se mencionó anteriormente tiene una identidad de secuencia del 96% con la Mpro de SARS-CoV-1 (Zhang *et al.*, 2020). Por otro lado, la actividad catalítica de PLpro permite tan solo la generación de cuatro proteínas efectoras no estructurales, tiene un 83% de identidad de secuencia con PLpro de SARS-CoV y su secuencia de reconocimiento para la escisión de las poliproteínas es Leu-X-Gly-Gly/X, una secuencia que se encuentra presente también en la ubiquitina (Ub) y en el producto génico estimulado por interferón 15 (ISG15) por lo que tiene actividad en ambas moléculas (Arya *et al.*,

2021). Estas son algunas razones por las que la comunidad científica ha elegido a Mpro como el blanco terapéutico idóneo para inhibir la replicación de SARS-CoV-2.

El acoplamiento molecular permite simular la interacción molecular y predecir el modo de unión y la afinidad entre receptores y ligandos. Esta información es relevante ya que pueden proporcionar información estructural clave para la búsqueda de nuevas moléculas que tengan mejor afinidad por la enzima. Como referencia para llevar a cabo los protocolos de acoplamiento se tomaron en cuenta dos inhibidores conocidos para su respectiva enzima. Para el caso de Mpro el compuesto referente fue Masitinib, un fármaco que se probó tanto computacional como experimentalmente por Drayman *et al.*, 2021. Este compuesto logró reducir >200 veces los títulos virales en los pulmones y nariz de ratones infectados con SARS-CoV-2. Para PLpro se utilizó el compuesto GRL-0617, el cual actúa mediante un mecanismo de inhibición competitiva con la proteína ISG15 inhibiendo la actividad de proteasa y por tanto la replicación de SARS-CoV y SARS-CoV-2 (Fu *et al.*, 2021). Además, con la finalidad de obtener información más precisa con respecto al modo de unión y la estructura química de los compuestos que se reportaron no solo con afinidad por su respectiva enzima si no que también reportaron inhibición en experimentos en laboratorio, se decidió separar en diferentes grupos a las moléculas, teniendo las categorías ACT (actividad), Mpro/PLpro (compuestos con afinidad por su enzima correspondiente) y RepMP/PL (fármacos reposicionados para su enzima correspondiente). El análisis respectivo a los compuestos con afinidad por Mpro demostró que aquellas moléculas con actividad biológica reportada son más pequeñas y tienen valores de AlogP mayores que pueden ser indicativos de mejor solubilidad, de acuerdo con las métricas de diversidad son además más parecidos entre ellos comparados con los otros grupos de compuestos, teniendo un mayor número de estructuras núcleo en común. Para el caso de PLpro, aunque los compuestos con actividad biológica no presentaron diferencia en tamaño, si tienen mayor AlogP y también son más parecidos en sus propiedades fisicoquímicas y en sus estructuras químicas entre sí. Estos resultados sugieren la posibilidad de encontrar características químicas estructurales dentro

del grupo de compuestos con actividad para ambas enzimas que puedan estar relacionadas con su actividad biológica. Complementario a este análisis se graficó y analizó el espacio químico para ambos sets de moléculas por separado (aquellos con afinidad por Mpro, PLpro y sus subgrupos correspondientes, incluyendo los compuestos Masitinib y GRL-0617 de referencia. En el espacio químico las moléculas ocupan su lugar dependiendo de los valores de sus propiedades fisicoquímicas, y sobre todo, los compuestos biológicamente activos no se distribuyen de manera uniforme en el espacio sino que forman regiones compactas en el (Varnek & Baskin., 2011) Por tanto se coleccionaron moléculas cercanas a los compuestos de referencia, bajo el supuesto de que las propiedades fisicoquímicas responsables de su posicionamiento son lo más parecidas a las propiedades de compuestos bien establecidos como afines y activos. Estas colecciones se acoplaron a su enzima correspondiente, encontrando características en sus estructuras responsables de la interacción entre los compuestos y el sitio activo de las enzimas. Estas características funcionaron como punto de partida para la búsqueda de nuevas moléculas mediante el empleo, en el caso de Mpro, de un modelo de farmacóforo y el cribado de las bases de datos ChEMBL y ZINC. Se encontraron en el presente trabajo a los compuestos COMP83 (ChEMBL199740), COMP90 (ChEMBL92481) y COMP70 (ChEMBL507400), los cuales presentan afinidad por los residuos de aminoácido de la cavidad del sitio activo de Mpro a juzgar por sus puntajes de acomodamiento, así como buen acomodo en dicha cavidad. De los tres compuestos encontrados, se evaluó la estabilidad con respecto al tiempo de las interacciones establecidas entre el compuesto COMP83 y los residuos de aminoácido que conforman el sitio activo de Mpro mediante simulaciones atómicas de dinámica molecular. Los resultados de la dinámica indican que COMP83 es capaz de permanecer interaccionando con los residuos Glu166, Gln189 y Met49 del sitio activo de Mpro durante 150 ns de simulación bajo condiciones fisiológicas de presión y temperatura, abarcando además durante el tiempo de simulación la cavidad del sitio activo, por lo que se sugiere que el compuesto impediría la actividad catalítica de Mpro sobre las poliproteínas virales,

evitando la formación de las proteínas virales requeridas para la replicación del virus.

Por otro lado para el caso de PLpro los compuestos con mejor acoplamiento a partir del análisis del espacio químico resultaron análogos al compuesto de referencia GRL-0617. Estos compuestos son COMP11 (ChEMBL559234) y COMP2 (ChEMBL561061), y tienen en su estructura un andamio molecular con el tamaño y propiedades idóneas para en un proyecto continuativo optimizarlo y mejorar su afinidad y aumentar la probabilidad de tener una actividad biológica mayor a la del compuesto de referencia.

Como perspectivas del presente trabajo se encuentra realizar las simulaciones de dinámica molecular para los compuestos COMP90 y COMP70 con el sitio catalítico de Mpro para evaluar la estabilidad de sus interacciones, además de probar experimentalmente mediante ensayos enzimáticos la actividad inhibitoria de al menos el COMP83 contra Mpro. De manera análoga, se sugiere evaluar la estabilidad de COMP11 y COMP2 en el sitio activo de PLpro mediante simulaciones de dinámica molecular. Los resultados de la dinámica permitirán tomar decisiones sobre la optimización del compuesto y si es apropiado probarlos experimentalmente.

La caracterización precisa de la pandemia por COVID-19 es un reto por muchas razones. Comenzando con la infravalorización del número de casos oficiales de contagios debido principalmente a la falta de pruebas y de notificación oportuna a las instancias de salubridad en el mundo. Aunque la gravedad clínica de la enfermedad ha disminuido considerablemente con el tiempo debido a la inmunidad adquirida ya sea natural o inducida por la vacunación o ambas, el número global de muertes por COVID está también infravalorado, pudiendo llegar a ser hasta tres veces superior al valor notificado oficialmente según datos epidemiológicos de la OMS (Lippi *et al.*, 2023). Además, las mutaciones son los componentes básicos de la evolución, ya que permiten la selección natural de rasgos beneficiosos para el virus, como una mayor virulencia, adaptabilidad y capacidad de evolución (Duffy., 2018). De acuerdo con Cao *et al.*, 2023, existen por lo menos 3000 descendientes

originados a partir del primer coronavirus 19-A identificado, siendo para finales del 2022 la variante Omicron la más reciente y que ha sufrido ya un proceso de intensa mutación y recombinación, generando varias docenas de sublinajes en la actualidad (Cao *et al.*, 2023). Por tanto, los esfuerzos por encontrar fármacos antivirales no deben parar y en este contexto el DIFAC permite emplear los recursos académicos y tecnológicos existentes en biología molecular, química teórica, quimioinformática, ciencias de la computación, entre otras, para mantener el diseño y la búsqueda de nuevos compuestos efectiva no solo contra COVID-19 sino para cualquier otra enfermedad. En la actualidad, la elucidación estructural de dianas farmacológicas permite a los investigadores la identificación virtual de inhibidores basada en estructuras y por otro lado, el descubrimiento de unas pocas moléculas líderes puede conducir al diseño y búsqueda de nuevos compuestos que pueden optimizarse mediante el diseño racional de fármacos basado en ligandos (Gurung *et al.*, 2021). El presente trabajo propone nuevas moléculas candidatas a probarse experimentalmente como inhibidores de las proteasas de SARS-CoV-2, pero también propone una metodología basada en la recopilación, manejo y uso de la información para la búsqueda de nuevas moléculas biológicamente activas, remarcando la importancia de la quimioinformática en la investigación farmacéutica actual.

12.- CONCLUSIÓN

El análisis del espacio químico de moléculas reportadas con afinidad y actividad inhibitoria contra las proteasas Mpro y PLpro de SARS-CoV-2 permitió identificar moléculas con patrones de estructura y propiedades fisicoquímicas clave para la búsqueda e identificación de nuevas moléculas candidatas a ser probadas experimentalmente para inhibir a su enzima correspondiente y entrar en una fase preclínica como potenciales fármacos contra COVID-19.

13.- BIBLIOGRAFÍA

Agatonovic-Kustrin, S., Ling, L. H., Tham, S. Y., y Alany, R. G. 2002. In Silico prediction of Oral bioavailability. *J. Pharm. Biomed. Anal.* 29: 103–119.

Agoni, C., Olotu, F. A., Ramharack, P., y Soliman, M.E. 2020. Druggability and drug-likeness concepts in drug design: are biomodelling and predictive tools having their say? *J. Mol. Model.*, 26: 6. <https://doi.org/10.1007/s00894-020-04385-6>

Agrafiotis, D. K., Shemanarev, K., Connolly, P. J., Farnum, M. y Lobanov, V. S. 2007. SAR Maps: A New SAR Visualization Technique for Medicinal Chemists. *J. Med. Chem.*, 50: 5926–5937.

Alanagreh, L., Alzoughool, F. y Atoum, M. 2020. The human coronavirus disease COVID-19: its origin, characteristics, and insights into potential drugs and its mechanisms. *Pathogens*, 9: 331. <https://doi.org/10.3390/pathogens9050331>

Amin, S. A. y Jha, T. 2020. Fight against novel coronavirus: a perspective of medicinal chemists. *Euro. J. Med. Chem.* 201: 112559.

Arús-Pous, J., Awale, M., Probst, D. y Reymond J.L. 2019. Exploring chemical space with machine learning. *Chimia (Aarau)* 73: 1018 – 1023. <https://doi.org/10.2533/chimia.2019.1018>

Arya, R., Kumari, S., Pandey, B., Mistry, H., Bihani, S.C., Das, A., Prashar, V., Gupta, G.D., Panicker, L. y Kumar, M. 2021. Structural insights into SARS-CoV-2 proteins. *J. Mol. Biol.* 433: 166725 <https://doi.org/10.1016/j.jmb.2020.11.024>

Báez-Santos, Y. M., John, S. E. y Mesecar, A. D. 2015. The SARS-coronavirus papain-like protease: structure, function, and inhibition by designed antiviral compounds. *Antivir. Res.* 115: 21–38.

Báez-Santos, Y.M., Barraza, S.J., Wilson, M.W., Agius, M.P., Mielech, A.M., Davis, N.M., Baker, S.C., Larsen, S. D., et al. 2014. X-ray structural and biological evaluation of a series of potent and highly selective inhibitors of human coronavirus papain-like proteases. *J. Med. Chem.* 57: 2393 – 2412. <https://doi.org/10.1021/jm401712t>.

Bajorath, J., Peltason, L., Wawer, M., Guha, R., Lajiness, M.S. y Van Drie J.H. 2009. Navigating structure-activity landscapes. *Drug Discov. Today*, 14, 13-14: 698-705. <https://doi.org/10.1016/j.drudis.2009.04.003>

Barretto, N., Jukneliene, D., Ratia, K., Chen, Z., Mesecar, A.D. y Baker S.C. 2005. The papain-like protease of severe acute respiratory syndrome coronavirus has deubiquitinating activity. *J. Virol.*, 79: 15189 – 15198. <https://doi.org/10.1128/JVI.79.24.15189-15198.2005>.

Bickerton, G. R., Paolini, G. V., Besnard, J., Muresan, S. y Hopkins A. L. 2012. Quantifying the Chemical Beauty of Drugs. *Nat. Chem.* 4: 90–98.

Bickerton, G.R., Paolini, G. V., Besnard, J., Muresan, S., y Hopkins, A. L. 2012. Quantifying the chemical beauty of drugs. *Nat. Chem.*, 4, 2: 90 - 98. <https://doi.org/10.1038/nchem.1243>

Brown, F. K. y Bristol J. A. 1998. Chapter 35 - Chemoinformatics: What Is It and How Does It Impact Drug Discovery. *Annu. Rep. Med. Chem.* 33: 375–384.

Cambridge Dictionary; Cambridge University Press, 2020. <https://dictionary.cambridge.org/> (consultado en Diciembre 2021).

Cao, Y., Jian, F., Wang, J. et al. 2023. Imprinted SARS-CoV-2 humoral immunity induces convergent Omicron RBD evolution. *Nature* 614: 521 – 529. <https://doi.org/10.1038/s41586-022-05644-7>

Chams, N., Chams, S., Badran, R., Shams, A., Araji, A., Raad, M., Mukhopadhyay, S., Stroberg, E., Duval, E. J., Barton, L. M. y Hajj Hussein, I. 2020. COVID-19: A Multidisciplinary Review. *Front. Public Health* 8:383. <https://doi.org/10.3389/fpubh.2020.00383>

Chen, Y., Liu, Q. y Guo, D. 2020. Emerging coronaviruses: genome structure, replication, and pathogenesis. *J. Med. Virol.* 92: 418 - 423. <https://doi.org/10.1002/jmv.25681>.

Clark, D.E. 1999. Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. 1. Prediction of intestinal absorption. *J. Pharm. Sci.* 88: 807-14.

Cobb, R.R. y Molony L.J. 2007. Drug discovery: design and serendipity. In *Principles and Practice of Pharmaceutical Medicine* (Edwards, L.D., Fletcher, A.J., Fox, A.W., Stonier, P.D., eds), pp. 43–50

Coelho, C., Gallo, G., Campos, C.B., Hardy, L. y Würtele, M. 2020. Biochemical screening for SARS-CoV-2 main protease inhibitors. *PLoS ONE* 15, 10: e0240079. <https://doi.org/10.1371/journal.pone.0240079>

Cosconati, S., Forli, S., Perryman, A.L., Harris, R., Goodsell, D.S. y Olson A.J. 2010. Virtual screening with AutoDock: Theory and Practice. *Expert Opin. Drug Discov.* 6: 597 - 607.

Cucinotta, D. y Vanelli, M. 2020. WHO declares COVID-19 a pandemic. *Acta Biomed.* 91: 157-160.

Dafydd, R. Owen, et al. 2021. An oral SARS-CoV-2 Mpro inhibitor clinical candidate for the treatment of COVID-19. *Science* 485, 1586-1593. <https://doi.org/10.1126/science.abl4784>

Dai, W., Zhang B., Xian-Ming Jiang et al. 2020. Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science* 368: 1331-1335 <https://doi.org/10.1126/science.abb4489>

Das, S., Sarmah, S., Lyndem, S., y Singha-Roy, A. 2020. An investigation into the identification of potential inhibitors of SARS-CoV-2 main protease using molecular docking study. *J. of Biomol. Struct.*, pp. 1–11

Dávila, E.M., Patricio, F., Rebolledo-Bustillo, M., García-Gomez, D., Hernandez, J.C.G., Sanchez-Gaytan, B.L., Limón, I.D. y Perez-Aguilar, J.M. 2022. Interacting binding insights and conformational consequences of the differential activity of cannabidiol with two endocannabinoid-activated G-protein-coupled receptors. *Front. Pharmacol.* 13:945935. <https://doi.org/10.3389/fphar.2022.945935>

Delre, P., Caporuscio, F., Saviano, M., y Mangiatordi, G. F. 2020. Repurposing Known Drugs as Covalent and Non-covalent Inhibitors of the SARS-CoV-2 Papain-Like Protease. *Front. Chem.*, 8. <https://doi:10.3389/fchem.2020.594009>

DiMasi, J. A., Grabowski, H. G., y Hansen, R. W. 2016. Innovation in the pharmaceutical industry: New estimates of R&D costs. *J. Health Econ.*, 47: 20–33. <https://doi.org/10.1016/j.jhealeco.2016.01>

DiMasi, J.A., Feldman, L., Seckler, A. y A. Wilson. 2010. Trends in risks associated with new drug development : success rates for investigational drugs. *Clin. Pharmacol. Ther.* 87: 272–277. <https://doi.org/10.1038/clpt.2009.295>

Dobson, C.M. 2004. Chemical space and biology. *Nature*, 432, 7019: 824 - 828. <https://doi.org/10.1038/nature03192>

Drayman, N., DeMarco, J.K., Jones, K. A., Azizi, S. A., Froggatt, H. M., Tan, K., Maltseva, N. I., Chen, S., Nicolaescu, V., Dvorkin, S., Furlong, K., Kathayat, R. S., Firpo, M. R., Mastrodomenico, V., Bruce, E. A., Schmidt, M. M., Jedrzejczak, R., Muñoz-Alía, M., Schuster, B., ... Tay, S. 2021. Masitinib is a broad coronavirus 3CL inhibitor that blocks replication of SARS-CoV-2. *Science*, 373, 6557: 931 - 936. <https://doi.org/10.1126/science.abg5827>

Duffy, S. 2018. Why are RNA virus mutation rates so damn high? *PLoS Biol.* 16: 8.

Saldívar-Gonzalez, F.I., Valli M., Andricopulo, A.D., da Silva-Bolzani, V. y Medina-Franco, J.L. 2019. Chemical space and diversity of the NuBBE database: a chemoinformatic characterization, *J. Chem. Inf. Model.* 59: 74 - 85.

Fan, J., Fu, A. y Zhang, L. 2019. Progress in molecular docking. *Quant. Biol.*, 7, 2: 83-89. <https://doi.org/10.1007/s40484-019-0172-y>

Fehr, A.R. y Perlman, S. 2015. Coronaviruses: An Overview of their Replication and Pathogenesis, *Coronaviruses*, Springer, pp. 1 – 23.

Fernández-De Gortari, E., García-Jacas, C. R., Martínez-Mayorga, K., y Medina-Franco, J.L. 2017. Database fingerprint (DFP): an approach to represent molecular databases. *J. Cheminformatics*, 9, 1: 1-9. <https://doi.org/10.1186/s13321-017-0195-1>

Fourches, D., Muratov, E. y A. Tropsha. 2010. Trust but verify on the importance of chemical structure curation in cheminformatics and QSAR modeling research. *J Chem. Inf. Model.*, 50: 1189–204.

Friesner, R.A., Banks, J.A. y Murphy, R.B. et al. 2004. Glide: A new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy. *J. Med. Chem.* 47: 1739 - 1749.

Fu, Z., Huang, B., Tang, J., Liu, S., Liu, M., Ye, Y., Huang, H. et al. 2021. The complex structure of GRL0617 and SARS-CoV-2 PLpro reveals a hot spot for antiviral drug discovery. *Nat. Commun.*, 12: 1. <https://doi.org/doi:10.1038/s41467-020-20718-8>

Fung, T.S. y Liu, D.X. 2014. Coronavirus infection, ER stress, apoptosis and innate immunity. *Front. Microbiol.*, 5: 296. <https://doi.org/10.3389/fmicb.2014.00296>

Gasteiger, J. y Funatsu, K. 2006. Chemoinformatics – An important scientific discipline. *J. Comput. Chem.* 5, 2: 53 – 8. <https://doi.org/10.2477/jccj.5.53>

Gasteiger, J. 2020. Chemistry in Times of Artificial Intelligence. *Chem. Phys. Chem.* 21: 2233 – 2242.

Gentile, D., Patamia, V., Scala, A., Sciortino, M. T., Piperno, A., y Rescifina, A. 2020. Putative Inhibitors of SARS-CoV-2 Main Protease from A Library of Marine Natural Products: A Virtual Screening and Molecular Modeling Study. *Mar. Drugs*, 18, 4: 225 <https://doi.org/10.3390/md18040225>

Ghosh, A. K., Takayama, J., Aubin, Y., Ratia, K., Chaudhuri, R., Baez, Y., Mesecar, A. D. et al. 2009. Structure-based design, synthesis, and biological evaluation of a series of novel and reversible inhibitors for the severe acute respiratory syndrome–coronavirus papain-like protease. *J. Med. Chem.*, 52, 16: 5228 – 5240. <https://doi.org/10.1021/jm900611t>

Giordano, D., Biancaniello, C., Argenio, M.A. y Facchiano, A. 2022. Drug Design by Pharmacophore and Virtual Screening Approach. *Pharmaceuticals*. 15: 646. <https://doi.org/10.3390/ph15050646>

González-Medina, M. y Medina-Franco, J.L. 2019. Chemical Diversity of Cyanobacterial Compounds: A Chemoinformatics Analysis. ACS Omega, 4, 4: 6229 – 6237. <https://doi.org/10.1021/acsomega.9b00532>

González-Medina, M. y Medina-Franco, J.L. 2017. Platform for Unified Molecular Analysis: PUMA. J. Chem. Inf. Model., 57, 8: 1735 – 1740. <https://doi.org/10.1021/acs.jcim.7b00253>

González-Medina, M., Méndez-Lucio, O. y Medina-Franco J.L. 2017. Activity landscape plotter: a web-based application for the analysis of structure-activity relationships. J. Chem. Inf. Model., 57: 397– 402. <https://doi.org/10.1021/acs.jcim.6b00776>.

González-Medina, M., Owen, J. R., El-Elimat, T., Pearce, C. J., Oberlies, N. H., Figueroa, M., y Medina-Franco, J. L. 2017. Scaffold Diversity of Fungal Metabolites. Front. Pharmacol., 8. <https://doi.org/10.3389/fphar.2017.00180>

Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., O'Meara, M. J., Rezelj, V. V., Guo, J. Z., Swaney, D. L., et al. 2020. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature 583: 459–468.

Goyal, B. y Goyal, D. 2020. Targeting the dimerization of the main protease of coronaviruses: A potential broad-spectrum therapeutic strategy. ACS Comb. Sci., 22, 6: 297–305

Guedes, I. A., Costa, L. S. C., dos Santos, K. B., Karl, A. L. M., Rocha, G. K., Teixeira, I. M., Dardenne, L. E. et al. 2021. Drug design and repurposing with DockThor-VS web server focusing on SARS-CoV-2 therapeutic targets and their non-synonym variants. Scientific Reports, 11: 1. <https://doi.org/10.1038/s41598-021-84700-0>

Guha, R. 2012. Exploring structure-activity data using the landscape paradigm. Wiley Interdiscip. Rev. Comput. Mol. Sci. <https://doi.org/10.1002/wcms.1087>

Guha, R. y Van Drie, J.H. 2008. Structure–activity landscape index: identifying and quantifying activity cliffs. J. Chem. Inf. Model 48: 646–658

Guo, Y., Cao, Q., Hong, Z. et al. 2020 The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak - an update on the status. Mil. Med. Res. 7: 11.

Gurung, A. B., Ali, M. A., Lee, J., Farah, M. A. y Al-Anazi, K. M. 2021. An Updated Review of Computer-Aided Drug Design and Its Application to COVID-19. Biomed. Res. Int. <https://doi.org/10.1155/2021/8853056>

Harcourt, B.H., Jukneliene, D., Kanjanahaluethai, A., Bechill, J., Severson, K.M., Smith, C.M., Rota, P.A. y Baker S.C. 2004. Identification of severe acute respiratory syndrome coronavirus replicase products and characterization of papain-like protease activity. *J. Virol.*, 78: 13600 – 13612. <https://doi.org/10.1128/JVI.78.24.13600-13612.2004>

Harrison, A. G., Lin, T., y Wang P. 2020. Mechanisms of SARS-CoV-2 Transmission and Pathogenesis. *Trends Immunol.* <https://doi.org/10.1016/j.it.2020.10.004>

Harrold, M.W y Zavod, R.M. 2018. Salts and solubility (Solubility and partition coefficients) en Basic concepts in Medicinal Chemistry. 2nd Ed. Pp. 30, 175. Ashp publications.

Helmy, Y.A., Fawzy, M., Elasad, A., Sobieh, A., Kenney, S.P. y Shehata, A.A. 2020. The COVID-19 pandemic: a comprehensive review of taxonomy, genetics, epidemiology, diagnosis, treatment, and control. *J. Clin. Med.*, 9: 1225. <https://doi.org/10.3390/jcm9041225>.

Hermann, M. y Bogunovic, D. 2017. ISG15: in sickness and in health. *Trends Immunol.* 38: 79–93.

Hersey, A., Chambers, J., Bellis, L., Patrícia Bento, A., Gaulton, A., y Overington J.P. 2015. Chemical databases: Curation or integration by user-defined equivalence? *Drug Discov. Today: Technologies.* 14: 17 – 24. <https://doi.org/10.1016/j.ddtec.2015.01.005>

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., et al. 2020. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*, 181: 271–280. <https://doi.org/10.1016/j.cell.2020.02.052>

Hollingsworth, S. A. y Dror, R.O. 2018. Molecular Dynamics Simulation for All. *Neuron*, 99, 6: 1129–1143. <https://doi.org/10.1016/j.neuron.2018.08.011>

Hu, Y., Stumpfe, D. y Bajorath, J. 2016. Computational Exploration of Molecular Scaffolds in Medicinal Chemistry. *J. Med. Chem.* 59: 4062 – 4076.

Huang, H.-J., Yu, H. W., Chen, C.-Y., et al. 2010. Current developments of computer-aided drug design. *J. Taiwan Inst. Chem.*, 41, 6: 623 – 635.

Huynh, T., Cornell, W., y Luan, B. 2021. In silico Exploration of Inhibitors for SARS-CoV-2's Papain-Like Protease. *Front. Chem.*, 8: 1 – 9. <https://doi.org/10.3389/fchem.2020.624163>

Ibrahim, O.O. 2020. Coronavirus SARS-CoV-2 is the newly emerged zoonotic virus causing pandemic death and economic loss, *EC Pulmonol. Respir. Med.* 9: 65–75.

Karplus, M. y McCammon, J.A. 2002. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* 9: 646–652.

Kim, Y., Liu, H., Galasiti-Kankanamalage, A.C. et al. 2016. Reversal of the progression of fatal coronavirus infection in cats by a broad-spectrum coronavirus protease inhibitor. *PLoS Pathogens* 12: 1005531. <https://doi.org/10.1371/journal.ppat.1005531>

Kirchmair, J. 2020. Molecular informatics in natural products research. *Mol Inf* 39, 11: 2000206. <https://doi.org/10.1002/minf.202000206>

Kola, I. y Landis, J. 2004. *Nat. Rev. Drug Discov.*, 3: 711–715.

Kong, R., Yang, G., Xue, R., Liu, M., Wang, F., Hu, J., Guo, X. y S. Chang. 2020. COVID-19 Docking Server: An Interactive Server for Docking Small Molecules, Peptides and Antibodies Against Potential Targets of COVID-19. *Bioinformatics.* 36, 20: 5109 - 5111 <https://doi.org/10.1093/bioinformatics/btaa645>

Lander, E. S. et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409: 806–921.

Lang, Y., Li, W., Li, Z., Koerhuis, D., van den Burg, A.C., Rozemuller, E., Bosch, B.J., van Kuppeveld, F.J., Boons, G.J.P. y Huizinga, E.G. 2020. Coronavirus hemagglutinin-esterase and spike proteins co-evolve for functional balance and optimal virion avidity, *bioRxiv.* <https://doi.org/10.1101/2020.04.03.003699>.

Langdon, S. R., Brown, N., y Blagg, J. 2011. Scaffold Diversity of Exemplified Medicinal Chemistry Space. *J. Chem. Inf. Model.*, 51(9), 2174 – 2185. <https://doi.org/10.1021/ci2001428>

Li, D., Kerns, E. y Carter, G. 2009. Drug-like property concepts in pharmaceutical design. *Curr. Pharm. Des.*, 15: 2184 – 2194

Li, S.F., He, H.D., Parthiban, L.J., Yin, H.Q. y Serajuddin, A.T.M. 2005. IV-IVC considerations in the development of immediate-release oral dosage form. *J. Pharm. Sci.* 94: 1396-417.

Lindner, H.A., Lytvyn, V., Qi, H., Lachance, P., Ziomek, E. y R. Menard. 2007. Selectivity in ISG15 and ubiquitin recognition by the SARS coronavirus papain-like protease. *Arch. Biochem. Biophys.*, 466: 8 – 14. <https://doi.org/10.1016/j.abb.2007.07.006.36>

Lipinski, C. A., Lombardo, F., Dominy, B. W. y Feeney, P.J. 2001. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Deliv. Rev.* 46, 1–3: 3–26.

Lippi, G., Sanchis-Gomar, F. y Henry, B. M. 2023. COVID-19 and its long-term sequelae: what do we know in 2023? *Pol Arch Intern Med.* 133: 16402. <https://doi.org/10.20452/pamw.16402>

Liu, J., Sun, Y., Qi, J., Chu, F., Wu, H., Gao, F, Li, T., Yan, J. y Gao, G.F. 2010. The membrane protein of severe acute respiratory syndrome coronavirus acts as a dominant immunogen revealed by a clustering region of novel functionally and structurally defined cytotoxic T-lymphocyte epitopes. *J. Infect. Dis.* 202, 8: 1171–1180.

Liu, X., y Wang, X.-J. 2020. Potential inhibitors against 2019-nCoV coronavirus M protease from clinically approved medicines. *J. Genet. Genomics.* <https://doi.org/10.1016/j.jgg.2020.02.001>

López-López, E., Bajorath, J. y Medina-Franco, J.L. 2021. Informatics for chemistry, biology, and biomedical sciences. *J. Chem. Inf. Model.* 61: 26–35. <https://doi.org/10.1021/acs.jcim.0c01301>

López-López, E., Naveja, J.J. y Medina-Franco, J.L. 2019. DataWarrior: an evaluation of the open-source drug discovery tool. *Expert Opin. Drug Discov.* 14: 335–341. <https://doi.org/10.1080/17460441.2019.1581170>

López-López, E., Rabal, O., Oyarzabal, J. y Medina-Franco, J.L. 2020. Towards the understanding of the activity of G9a inhibitors: an activity landscape and molecular modeling approach. *J. Comput. Aided Mol. Des.*, 34, 6: 659 - 669. <https://doi.org/10.1007/s10822-020-00298-x>

Medina-Franco, J. L. 2012. Interrogating Novel Areas of Chemical Space for Drug Discovery using Chemoinformatics. *Drug Dev. Res.* 73, 7: 430–438.

Medina-Franco, J. L., y Saldívar-González, F. I. 2020. Cheminformatics to Characterize Pharmacologically Active Natural Products. *Biomolecules*, 10, 11: 1566. <https://doi.org/10.3390/biom10111566>

Medina-Franco, J. L., Norberto, F., Cruz, S., López, E. 2021. Progress on open chemoinformatic tools for expanding and exploring the chemical space. *J. Comput. Aided Mol. Des.*, 0123456789. <https://doi.org/10.1007/s10822-021-00399-1>

Medina-Franco, J.L. 2012. Scanning structure-activity relationships with structure-activity similarity and related maps: from consensus activity cliffs to selectivity switches. *J. Chem. Inf. Model.*, 52: 2485 – 2493. <https://doi.org/10.1021/ci300362x>

Medina-Franco, J.L., Martínez-Mayorga, K., Bender, A. y Scior, T. 2009. Scaffold diversity analysis of compound data sets using an entropy-based measure. *QSAR and Combinatorial Science*, 28, 11 - 12: 1551 - 1560. <https://doi.org/10.1002/qsar.200960069>

Medina-Franco, J.L., Martínez-Mayorga, K., Giulianotti, M.A., Houghten, R.A. y Pinilla, C. 2008. Visualization of the chemical space in drug discovery. *Curr. Comput. Aided Drug Des.* 4: 322 - 333. <https://doi.org/10.2174/157340908786786010>

Medina-Franco, J.L., Naveja J.J. y López-López, E. 2019. Reaching for the bright StARs in chemical space. *Drug Discov Today*. <https://doi.org/10.1016/j.drudis.2019.09.013>

Mengist, H. M., Dilnessa, T., y Jin, T. 2021. Structural Basis of Potential Inhibitors Targeting SARS-CoV-2 Main Protease. *Front. Chem.*, 9, 1 - 19. <https://doi.org/10.3389/fchem.2021.622898>

Milardi, D. y M. Pappalardo. 2015. Molecular dynamics: New advances in drug discovery. *Euro. J. Medicinal Chem.*, 91: 1-3. <https://doi.org/10.1016/j.ejmech.2014.10.078>

Miranda-salas, J., Peña-varas, C., Valenzuela, I., Olmedo, D.A., Zamora, W.J., Chávez-fumagalli, M. A., Azevedo, D., Q., Oliveira, R., Maltarollo, V.G., Ramírez, D., y Medina-franco, J.L. 2023. Artificial Intelligence in the Life Sciences Trends and challenges in chemoinformatics research in Latin America. <https://doi.org/10.1016/j.ailsi.2023.100077>

Mitra, K., Ghanta, P., Acharya, S., Chakrapani, G., Ramaiah, B., y Doble, M. 2020. Dual inhibitors of SARS-CoV-2 proteases: pharmacophore and molecular dynamics based drug repositioning and phytochemical leads. *J. Biomol. Struct.*, 1-14. <https://doi.org/10.1080/07391102.2020.1796802>

Morris, G. M. y Lim-Wilby, M. 2008 Molecular docking. *Methods Mol. Biol.*, 443: 365 - 382.

Mu, J., Xu, J., Zhang, L., Shu, T., Wu, D., Huang, M., Ren, Y., Li, X., Geng, Q. y Xu, Y. 2020. SARS-CoV-2-encoded nucleocapsid protein acts as a viral suppressor of RNA interference in cells, *Sci. China Life Sci.* 1 - 4.

Muratov, E.N., Amaro, R., Andrade, C.H., Brown, N., Ekins, S., Fourches, D., Isayev, O., Kozakov, D., Medina-Franco, J.L., Merz, K.M., Oprea, T.I., Poroikov, V., Schneider, G., Todd, M.H., Varnek, A., Winkler, D.A., Zakharov, A.V., Cherkason, A. y Tropsha, A. 2021. A critical overview of computational approaches

employed for COVID-19 drug discovery. Chem. Soc. Rev..
<http://doi.org/10.1039/d0cs01065k>

Naveja, J.J., Vogt, M., Stumpfe, D., Medina-Franco, J.L. y Bajorath, J. 2019. Systematic Extraction of Analogue Series from Large Compound Collections Using a New Computational Compound-Core Relationship Method. ACS Omega, 4, 1: 1027 - 1032. <https://doi.org/10.1021/acsomega.8b03390>

Nikolova, N., y Jaworska, J. 2003. Approaches to Measure Chemical Similarity– a Review. QSAR Comb. Sci., 22, 910: 1006 – 1026. <https://doi.org/10.1002/qsar.200330831>

Perez, J. J. 2005. Managing molecular diversity. Chem. Soc. Rev. 34, 2: 143. <https://doi.org/10.1039/b209064n>

Perng, Y. C. y Lenschow, D. J. 2018. ISG15 in antiviral immunity and beyond. Nat. Rev. Microbiol. 16: 423–439.

Poduri, R., Joshi, G., y Jagadeesh, G. 2020. Drugs targeting various stages of the SARS-CoV-2 life cycle: Exploring promising drugs for the treatment of Covid-19. Cell. Signal., 74: 109721. <https://doi.org/10.1016/j.cellsig.2020.109721>

Prasanna, S. y Doerksen, R. 2008. Topological Polar Surface Area: A Useful Descriptor in 2D-QSAR. Curr. Med. Chem., 16, 1: 21 – 41
<https://doi.org/10.2174/092986709787002817>

Rajpoot, S., Alagumuthu, M. y Baig, M. S. 2020. Dual Targeting of 3CLpro and PLpro of SARS-CoV-2: A Novel Structure-Based Design Approach to treat COVID-19. CRSB. <https://doi.org/10.1016/j.crstbi.2020.12.001>

Rangel-Galván, M., Rangel, A., Romero-Méndez, C., Dávila, E.M., Castro, M.E., Caballero, N.A. y Perez-Aguilar, J.M. 2021. Inhibitory Mechanism of the Isoflavone Derivative Genistein in the Human CaV3.3 Channel. ACS Chem. Neurosci. 12, 4: 651-659. <https://doi.org/10.1021/acchemneuro.0c00684>

Rao, P., Patel, R., Shukla, A., et al. 2022. Identifying structural-functional analogue of GRL0617, the only well-established inhibitor for papain-like protease (PLpro) of SARS-CoV-2 from the pool of fungal metabolites using docking and molecular dynamics simulation. Mol Divers., 26: 309-329. <https://doi.org/10.1007/s11030-021-10220-8>

Rao, P., Patel, R., Shukla, A., Parmar, P., Rawal, R. M., Saraf, M. y Goswami, D. 2021. Identifying structural–functional analogue of GRL0617, the only well-established inhibitor for papain-like protease (PLpro) of SARS-CoV2 from the pool of fungal metabolites using docking and molecular dynamics simulation. Mol. Divers., <https://doi.org/10.1007/s11030-021-10220-8>

Ratia, K., Kilianski, A., Baez-Santos, Y. M., Baker, S. C. y Mesecar, A. 2014. Structural basis for the ubiquitin-linkage specificity and deISGylating activity of SARS-CoV papain-like protease. *PLoS Pathog.* 10, e1004113.

Ratia, K., Pegan, S., Takayama, J., Sleeman, K., Coughlin, M., Baliji, S., Chaudhuri, R., Fu, W., Mesecar, A. D. et al. 2008. A noncovalent class of papain-like protease/ deubiquitinase inhibitors blocks SARS virus replication. *PNAS*, 105, 42: 16119 – 16124. <https://doi.org/10.1073/pnas.0805240105>

Ratia, K., Saikatendu, K.S., Santarsiero, B.D., Barretto, N., Baker, S.C., Stevens, R.C. y Mesecar, A.D. 2006. Severe acute respiratory syndrome coronavirus papain-like protease: structure of a viral deubiquitinating enzyme. *PNAS*, 103: 5717 – 5722. <https://doi.org/10.1073/pnas.0510851103>

Rebolledo-Bustillo, M., Garcia-Gomez, D., Dávila, E.M., Castro, M.E., Caballero, N.A., Melendez, F.J., Baizabal-Aguirre, V.M., Sanchez-Gaytan, B.L. y Perez-Aguilar, J.M. 2023. Structural Basis of the Binding Mode of the Antineoplastic Compound Motixafortide (BL-8040) in the CXCR4 Chemokine Receptor. *Int. J. Mol. Sci.* 24, 5: 4393. <https://doi.org/10.3390/ijms24054393>

Robson, B. 2020. COVID-19 Coronavirus spike protein analysis for synthetic vaccines, a peptidomimetic antagonist, and therapeutic drugs, and analysis of a proposed achilles' heel conserved region to minimize probability of escape mutations and drug resistance, *Comput. Biol. Med.* 103749.

Romano, M., Ruggiero, A., Squeglia, F., Maga, G., Berisio, R. 2020. A structural view of SARS-CoV-2 RNA replication machinery: RNA synthesis, proofreading and final capping. *Cells*, 9: 1267. <https://doi.org/10.3390/cells9051267>

Rut, W., Lv, Z., Zmudzinski, M., Patchett, S., Nayak, D., Snipas, S.J., Oualid, F.E., Huang, T.T., et al., 2020. Activity profiling and crystal structures of inhibitor-bound SARS-CoV-2 papain-like protease: a framework for anti- COVID-19 drug design. *Sci. Adv.*, 6, eabd4596. <https://doi.org/10.1126/sciadv.abd4596>

Sadowski, J. 2000. Optimization of the drug-likeness of chemical libraries. *Perspect. Drug Discov. Des.* 20: 17-28.

Safiabadi Tali, S.H., LeBlanc, J.J., Sadiq, Z., Oyewunmi, O.D., Camargo, C., Nikpour, B., Armanfard, N., Sagan, S.M. y Jahanshahi-Anbuhi, S. 2021. Tools and techniques for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)/COVID-19 detection. *Clin Microbiol. Rev.* 34: e00228-20. <https://doi.org/10.1128/CMR.00228-20>

Saldívar-González, F., et al. 2016. Descubrimiento y desarrollo de fármacos: un enfoque computacional. *Educación Química* <http://dx.doi.org/10.1016/j.eq.2016.06.002>

Santibáñez-Morán, M.G., López-López, E., Prieto-Martínez, F.D., Sánchez-Cruz, N. y Medina-Franco, J.L. 2020. Consensus Virtual Screening of Dark Chemical Matter and Food Chemicals Uncover Potential Inhibitors of SARS-CoV-2 Main Protease. *RSC Adv.* 10: 25089–25099.

Schuurman, N. y Leszczynski, A. 2008. Ontologies for Bioinformatics. *Bioinf. Biol. Insights* 2: 187–200.

Shanmugasundaram, V. y G.M. Maggiora. 2001. Characterizing property and activity landscapes using an information-theoretic approach. 222nd American Chemical Society National Meeting, Division of Chemical Information Abstract no. 77

Shereen, M.A., Khan, S., Kazmi, A., Bashir, N. y Siddique, R. 2020. COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses, *J. Adv. Res.* 24: 91 – 98, <https://doi.org/10.1016/j.jare.2020.03.005>

Shoichet, B.K. 2004. Virtual Screening of chemical libraries. *Nature.* 432: 862-865.

Strazielle, N. y Gherzi-Egea, J.F. 2005. Factors affecting delivery of antiviral drugs to the brain. *Rev. Med. Virol.* 15: 105-33.

Sutch, B.T., Romero R.M., Neamati N., et al. 2012. Integrated teaching of structure-based drug design and biopharmaceutics: A computer-based approach. *JCE.*, 89: 45-51.

Tanimoto, T. T. 1958. An Elementary Mathematical Theory of Classification and Prediction, IBM Internal Report.

Tian, S., Wang, J., Li, Y., Li, D., Xu, L., y Hou, T. 2015. The application of in silico drug-likeness predictions in pharmaceutical research. *Adv. Drug Deliv. Rev.*, 86: 2–10. <https://doi.org/10.1016/j.addr.2015.01.009>

Trott O. y Olson, A.J. 2010. Software News and Update, AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* 31: 455 - 461.

Tsaioun, K., Bottlaender, M. y Mabondzo, A. 2009. *BMC Neurol.* 9 (Suppl. 1), S1. <https://doi.org/10.1186/1471-2377-9-S1-S1>

Turner, J. V. 2007. In Silico Prediction of Oral Bioavailability. *Comprehensive Med. Chem.*, 29, 5: 699 – 724.

Umesh D., Kundu C., Selvaraj C., Singh S. K., y Dubey, V.K. 2021 Identification of new anti-nCoV drug chemical compounds from Indian spices exploiting SARS-CoV-2 main protease as target. *J. Biomol. Struct.*, 39, 9: 3428–3434

Van de Waterbeemd, H. 2000. Intestinal Permeability: Prediction from Theory. In: Oral Drug Absorption: Prediction and Assessment; Dressman, J.B., Lennerhäs, H. Eds.; Drugs and the Pharmaceutical Sciences 106; Marcel Dekker: New York, 2000, 31–49.

Van de Waterbeemd, H., Smith, D. A. y Jones, B.C.J. 2001. Computer Aided Drug Design: Methods and Applications. *Comput. Aided Mol. Des.* 15: 273–286.

Varnek, A. y Baskin, I. I. 2011. Chemoinformatics as a Theoretical Chemistry Discipline. *Mol. Inform.*, 30, 1: 20–32. <https://doi.org/10.1002/minf.201000100>

Veber, D. F., Johnson, S. R., Cheng, H. Y., Smith, B. R., Ward, K. W., y Kopple, K. D. 2002. Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.*, 45, 12: 2615 – 2623. <https://doi.org/10.1021/jm020017n>

Voet, A., Banwell, E.F., Sahu, K.K., Heddle, J.G. y Zhang, K.Y.J. 2013. Protein Interface Pharmacophore Mapping Tools for Small Molecule Protein: Protein Interaction Inhibitor Discovery. *Curr. Top. Med. Chem.* 3, 989 – 1001.

Vohora, D. y Singh, G. 2018. *Pharmaceutical Medicine and Translational Clinical Research*, Academic Press

Vuorinen, A. y Schuster, D. 2015. Methods for generating and applying pharmacophore models as virtual screening filters and for bioactivity profiling. *Methods* 71: 113 – 134.

Walls, A.C., Park, Y.J., Tortorici, M.A., Wall, A., McGuire, A.T. y Velesler, D. 2020. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 181, 2: 281 - 292. <https://doi.org/10.1016/j.cell.2020.02.058>

Welchman, R.L. y Gordon, C. R.J. Mayer. 2005. Ubiquitin and ubiquitin-like proteins as multifunctional signals. *Nature Rev. Mol. Cell Biol.*, 6: 599 – 609. <https://doi.org/10.1038/nrm1700>

Wermuth, C.G. 2008. *The Practice of Medicinal Chemistry*; 3rd ed. Ed.; Academic Press-Elsevier: Burlington, San Diego, London, U.K.

Wetzel, S., Schuffenhauer, A., Roggo, S., Ertl, P. y Waldmann H. 2007. Cheminformatic analysis of natural products and their chemical space. *Chimia*, 61, 6: 355 – 360. <https://doi.org/10.2533/chimia.2007.355>

Williams, A.J., Ekins, S., y Tkachenko, V. 2012. Towards a gold standard: regarding quality in public domain chemistry databases and approaches to improving the situation. *Drug Discov. Today*. 17,13 – 14: 685 – 701.

Wu, Y. Peng, B. Huang, X. Ding, X. Wang, P. Niu, J. Meng, Z. Zhu, Z. Zhang, J. Wang, J. Sheng, L. Quan, Z. Xia, W. Tan, G. Cheng y Jiang, T. 2020. Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China, *Cell Host Microbe.* 27, 3: 325 - 328. <https://doi.org/10.1016/j.chom.2020.2020.02.001>

Xu, H., Zhong, L., Deng, J., Peng, J., Dan, H., Zeng, X., Li, T. y Chen, Q. 2020. High expression of ACE2 receptor of 2019-nCoV on the epithelial cells of oral mucosa. *Int. J. Oral Sci.*, 12: 1 – 5. <https://doi.org/10.1038/s41368-020-0074-x>

Zeng, Q., Langereis, M.A., van Vliet, A.L., Huizinga, E.G. y de Groot, R.J. 2008. Structure of coronavirus hemagglutinin-esterase offers insight into corona and influenza virus evolution, *Proc. Natl. Acad. Sci.* 105, 26: 9065 – 9069.

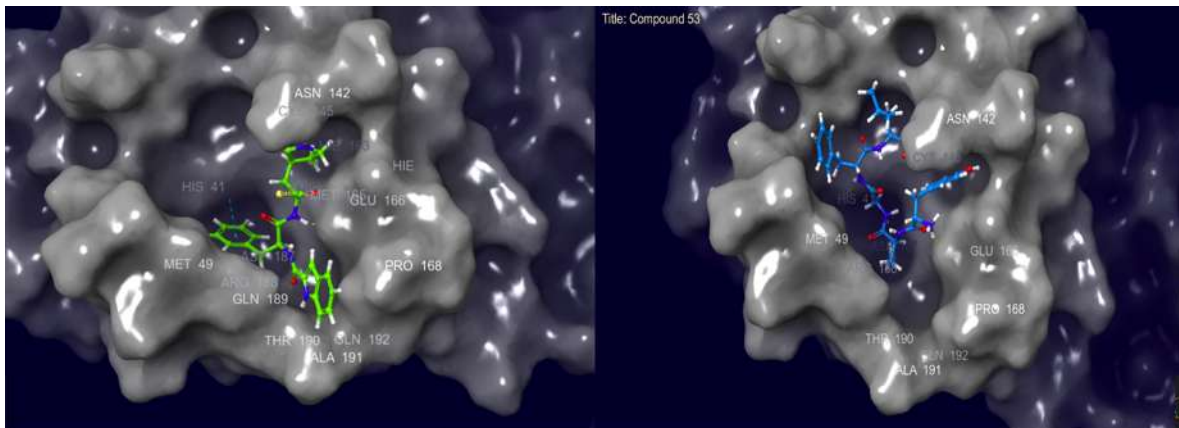
Zhang, L., Lin, D., Sun, X. y Curth, U. 2020. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α -ketoamide inhibitors. *Science* <https://doi.org/10.1126/science.abb3405>

Zhao, Y. y Truhlar, G. 2008. The Mo6 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four Mo6-class functionals and 12 other functionals. *Theor. Chem. Acc.* 20: 215 – 241.

Zhou, G., Chen, S. y Chen, Z. 2020. Advances in COVID-19: the virus, the pathogenesis, and evidence-based control and therapeutic strategies. *Front. Med.*, 14: 117 – 125. <https://doi.org/10.1007/s11684-020-0773-x>

Zhou, Q. A., Kato-Weinstein, J., Li, Y., Deng, Y., Granet, R., Garner, L., Liu, C., Polshakov, D., Gessner, C. y Watkins, S. 2020. Potential Therapeutic Agents and Associated Bioassay Data for COVID-19 and Related Human Coronavirus Infections. *ACS Pharmacol. Transl.*, 3, 5: 813 – 834. <https://doi.org/10.1021/acspsci.0c00074>

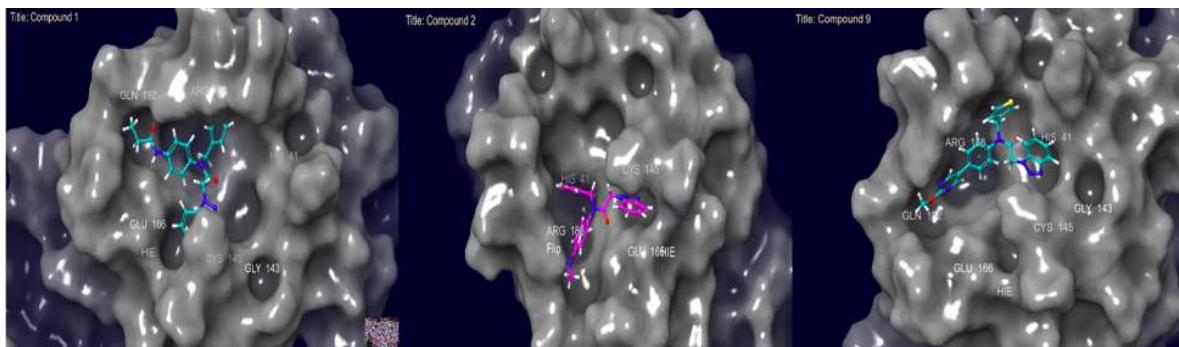
14.- MATERIAL SUPLEMENTARIO



Nombre del compuesto: MP201
Puntaje de acoplamiento: -8.780 kJ
Eficiencia del ligando: -0.251

Nombre del compuesto: RMP26
Puntaje de acoplamiento: -8.626 kJ
Eficiencia del ligando: -0.197

Figura 1 (MS). Posicionamiento de MP201 y RMP26 en el sitio activo de Mpro con una vista de superficie.



Nombre del compuesto: MP462
Puntaje de acoplamiento: -7.028 kJ
Eficiencia del ligando: -0.234

Nombre del compuesto: MP463
Puntaje de acoplamiento: -7.193 kJ
Eficiencia del ligando: -0.266

Nombre del compuesto: MP470
Puntaje de acoplamiento: -6.968 kJ
Eficiencia del ligando: -0.211

Figura 2 (MS). Posicionamiento de MP462, MP463 y MP470 en el sitio activo de Mpro con una vista de superficie.

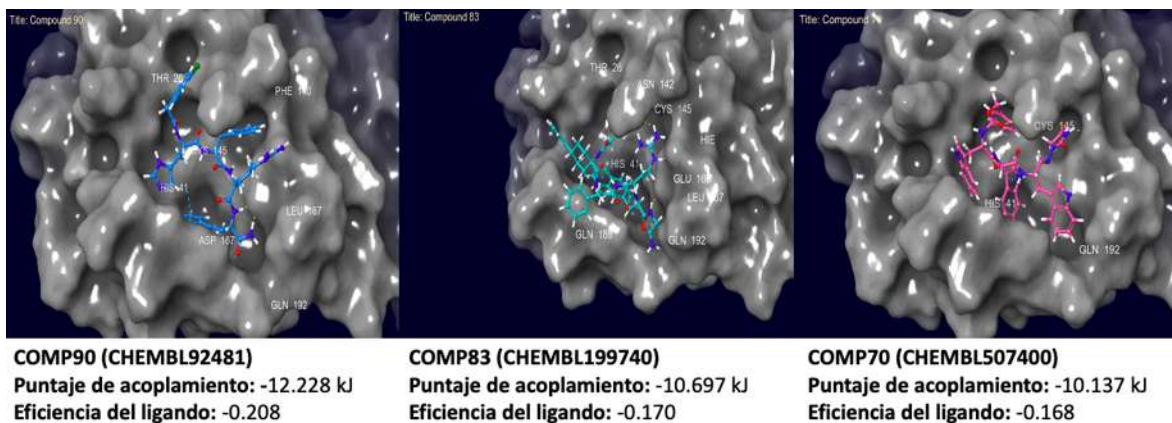


Figura 3 (MS). Posicionamiento de COMP90 (CHEMBL92481), COMP83 (CHEMBL199740) y COMP70 (CHEMBL507400) en el sitio activo de Mpro con una vista de superficie.

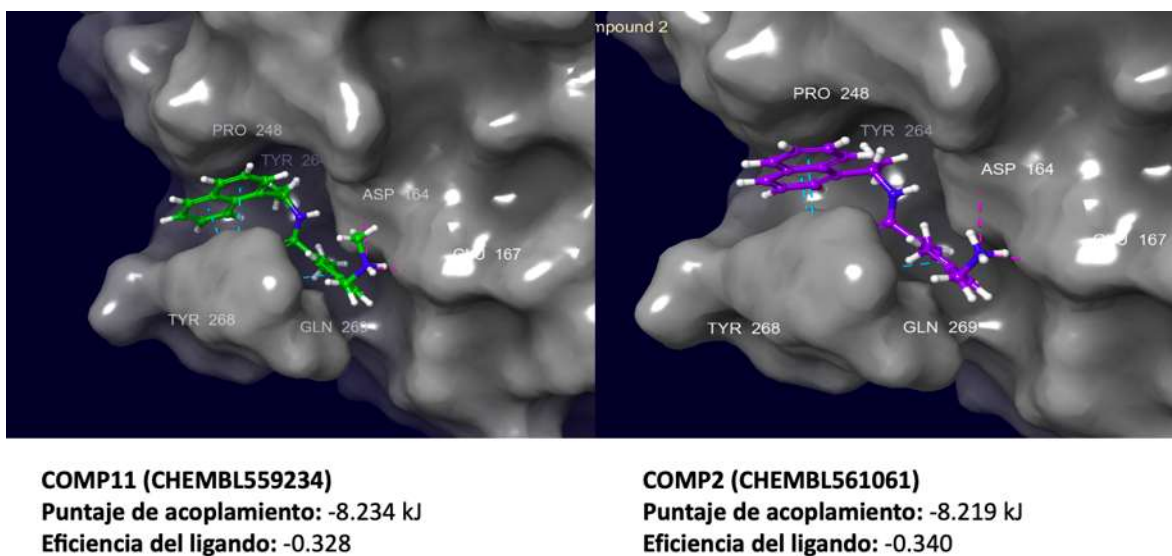
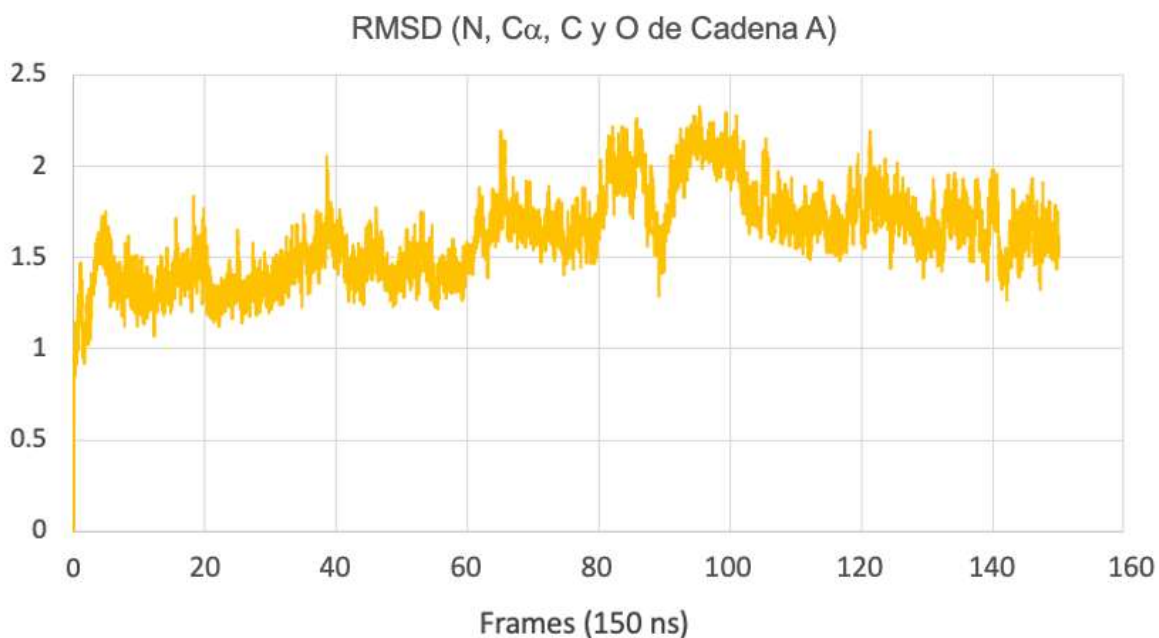


Figura 4 (MS). Posicionamiento de COMP11 (CHEMBL559234) y COMP2 (CHEMBL561061) en el sitio activo de PLpro con una vista de superficie.



Gráfica 1 (MS). Cálculo de RMSD general del “backbone” (N, C α , C y O de la cadena principal de A de Mpro) durante 150 ns de simulación atómica de dinámica molecular.

Tabla 1. Título de los 50 artículos revisados para la formación de la base de datos del presente trabajo (bibliografía en material suplementario).

No. Art.	Título del Artículo
1	Identification of natural compounds as potent inhibitors of SARS-CoV-2 main protease using combined docking and molecular dynamics simulations (Shamim-Jairapuri et al., 2021)
2	Putative Inhibitors of SARS-CoV-2 Main Protease from A Library of Marine Natural Products: A Virtual Screening and Molecular Modeling Study (Gentile et al., 2020)
3	Molecular docking studies, molecular dynamics and ADME/tox reveal therapeutic potentials of STOCK1N-69160 against papain-like protease of SARS-CoV-2 (Elekofehinti et al., 2020)
4	Dual inhibitors of SARS-CoV-2 proteases: pharmacophore and molecular dynamics based drug repositioning and phytochemical leads (Mitra et al., 2020)
5	Identification of polyphenols from <i>Broussonetia papyrifera</i> as SARS CoV-2 main protease inhibitors using in silico docking and molecular dynamics simulation approaches (Ghosh et al., 2020)
6	Identification of natural inhibitors against Mpro of SARS-CoV-2 by molecular docking, molecular dynamics simulation, and MM/PBSA methods (Sharma et al., 2020)
7	Microbial Natural Products as Potential Inhibitors of SARS-CoV-2 Main Protease (Mpro) (Sayed et al., 2020)
8	Identification of high-affinity inhibitors of SARS-CoV-2 main protease: Towards the development of effective COVID-19 therapy (Mohammad et al., 2020)
9	Cheminformatics-Based Identification of Potential Novel Anti-SARS-CoV-2 Natural Compounds of African Origin (Kwofie et al., 2021)

10	High throughput virtual screening reveals SARS-CoV-2 multi-target binding natural compounds to lead instant therapy for COVID- 19 treatment (Naik et al., 2020)
11	Computational Determination of Potential Inhibitors of SARS-CoV-2 Main Protease (Ngo et al., 2020)
12	Identification of new anti-nCoV drug chemical compounds from Indian spices exploiting SARS- CoV-2 main protease as target (Umesh et al., 2020)
13	Virtual Screening of Natural Metabolites and Antiviral Drugs with Potential Inhibitory Activity against 3CL-PRO and PL-PRO (Contreras-Puentes y Alviz-Amador, 2020)
14	In silico screening of Chinese herbal medicines with the potential to directly inhibit 2019 novel coronavirus (Zhang et al., 2020)
15	Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods (Wu et al., 2020)
16	Targeting COVID-19 (SARS-CoV-2) main protease through active phytochemicals of ayurvedic medicinal plants – <i>Withania somnifera</i> (Ashwagandha), <i>Tinospora cordifolia</i> (Giloy) and <i>Ocimum sanctum</i> (Tulsi) – a molecular docking study (Shree et al., 2020)
17	Potential inhibitors of coronavirus 3-chymotrypsin-like protease (3CLpro): an in silico screening of alkaloids and terpenoids from African medicinal plants. (Gyebi et al., 2020)
18	Screening of plant-based natural compounds as a potential COVID-19 main protease inhibitor: an in silico docking and molecular dynamics simulation approach (Majumder y Mandal, 2020)
19	Identification of bioactive molecule from <i>Withania somnifera</i> (Ashwagandha) as SARS-CoV-2 main protease inhibitor (Tripathi et al., 2020)
20	Evaluation of green tea polyphenols as novel corona virus (SARS CoV-2) main protease (Mpro) inhibitors – an in silico docking and molecular dynamics simulation study (Ghosh et al., 2020)
21	A Computational Study to Identify Potential Inhibitors of SARS-CoV-2 Main Protease (Mpro) from <i>Eucalyptus</i> Active Compounds (Muhammad et al., 2020)
22	Potential bioactive glycosylated flavonoids as SARS-CoV-2 main protease inhibitors: A molecular docking and simulation studies (Cherrak et al., 2020)
23	Computer aided identification of potential SARS CoV-2 main protease inhibitors from diterpenoids and biflavonoids of <i>Torreya nucifera</i> leaves (Ghosh et al., 2020a)
24	Depicting the inhibitory potential of polyphenols from <i>Isatis indigotica</i> root against the main protease of SARS CoV-2 using computational approaches (Ghosh et al., 2020b)
25	Structure-Based Screening to Discover New Inhibitors for Papain-like Proteinase of SARS-CoV-2: An In Silico Study (Jamalan et al., 2021)
26	Interaction of small molecules with the SARS-CoV-2 papain-like protease: In silico studies and in vitro validation of protease activity inhibition using an enzymatic inhibition assay (Pitsillou et al., 2021)
27	Lead compounds for the development of SARS-CoV-2 3CL protease inhibitors (Iketani et al., 2021)
28	Masitinib is a broad coronavirus 3CL inhibitor that blocks replication of SARS-CoV-2 (Drayman et al., 2021)
29	Potential Therapeutic Agents and Associated Bioassay Data for COVID-19 and Related Human Coronavirus Infections (Zhou et al., 2020)

30	Main Protease Inhibitors and Drug Surface Hotspots for the Treatment of COVID-19: A Drug Repurposing and Molecular Docking Approach-Main Protease Inhibitors for the Treatment of COVID-19 (Mahmudul et al., 2021)
31	SARS-CoV 3CLpro inhibitory effects of quinone-methide triterpenes from <i>Tripterygium regelii</i> (Ryu et al., 2010)
32	Potential SARS-CoV-2 main protease inhibitors (Banerjee et al., 2021)
33	Structural Basis of Potential Inhibitors Targeting SARS-CoV-2 Main Protease (Mengist et al., 2021)
34	Biochemical screening for SARS-CoV-2 main protease inhibitors (Coelho et al., 2020)
35	High-throughput screening identifies established drugs as SARS-CoV-2 PLpro inhibitors (Zhao et al., 2021)
36	Virtual high throughput screening: Potential inhibitors for SARS-CoV-2 PLPRO and 3CLPRO proteases (Jade et al., 2021)
37	Discovery of human coronaviruses pan-papain-like protease inhibitors using computational approaches (Alamri et al., 2020)
38	Ginkgolic acid and anacardic acid are specific covalent inhibitors of SARS-CoV-2 cysteine proteases (Chen et al., 2021)
39	Computational molecular docking and virtual screening revealed promising SARS-CoV-2 drugs (Hosseini et al., 2021)
40	Cyanobacterial metabolites as promising drug leads against the Mpro and PLpro of SARS-CoV-2: an in silico analysis (Naidoo et al., 2020)
41	In silico Exploration of Inhibitors for SARS-CoV-2's Papain-Like Protease (Huynh et al., 2021)
42	Black tea bioactives as inhibitors of multiple targets of SARS-CoV-2 (3CLpro, PLpro and RdRp): a virtual screening and molecular dynamic simulation study (Gogoi et al., 2021)
43	An update review of emerging small-molecule therapeutic options for COVID-19 (Tian et al., 2021)
44	Repurposing Known Drugs as Covalent and Non-covalent Inhibitors of the SARS-CoV-2 Papain-Like Protease (Delre et al., 2020)
45	In silico ADMET and molecular docking study on searching potential inhibitors from limonoids and triterpenoids for COVID-19 (Vardhan y Sahoo, 2020)
46	Identify potent SARS-CoV-2 main protease inhibitors via accelerated free energy perturbation-based virtual screening of existing drugs (Li et al., 2020)
47	Rapid Identification of Potential Inhibitors of SARS-CoV-2 Main Protease by Deep Docking of 1.3 Billion Compounds (Ton et al., 2020)
48	Flavonoids with inhibitory activity against SARS-CoV-2 3CLpro (Jo et al., 2020)
49	In silico prediction of potential inhibitors for the Main protease of SARS-CoV-2 using molecular docking and dynamics simulation based drug-repurposing (Kumar et al., 2020)
50	Identifying structural–functional analogue of GRL0617, the only well-established inhibitor for papain-like protease (PLpro) of SARS-CoV2 from the pool of fungal metabolites using docking and molecular dynamics simulation (Rao et al., 2021)

51	Coronavirus Main Proteinase (3CL ^{Pro}) Structure: Basis for Design of Anti-SARS Drugs (Anand et al., 2003)
52	Molecular docking identifies the binding of 3-chloropyridine moieties specifically to the S1 pocket of SARS-CoV Mpro (Niu et al., 2008)
53	Characterization of SARS-CoV main protease and identification of biologically active small molecule inhibitors using a continuous fluorescence-based assay (Kao et al., 2004)
54	Discovery of unsymmetrical aromatic disulfides as novel inhibitors of SARS-CoV main protease: Chemical synthesis, biological evaluation, molecular docking and 3D-QSAR study (Wang et al., 2017)
55	Drug Design Targeting the Main Protease, the Achilles Heel of Coronaviruses (Yang et al., 2006)
56	Structure-Based Drug Design and Structural Biology Study of Novel Nonpeptide Inhibitors of Severe Acute Respiratory Syndrome Coronavirus Main Protease (Lu et al., 2006)
57	Synthesis, modification and docking studies of 5-sulfonyl isatin derivatives as SARS-CoV 3C-like protease inhibitors (Liu et al., 2014)
58	Synthesis and evaluation of phenylisoserine derivatives for the SARS-CoV 3CL protease inhibitor (Konno et al., 2017)
59	Screening of electrophilic compounds yields an aziridiny peptide as new active-site directed SARS-CoV main protease inhibitor (Martina et al., 2005)
60	High-Throughput Screening Identifies Inhibitors of the SARS Coronavirus Main Proteinase (Blanchard et al., 2004)
61	Discovery, Synthesis, And Structure-Based Optimization of a Series of N-(tert-Butyl)-2-(N-arylamido)-2-(pyridin-3-yl) Acetamides (ML188) as Potent Noncovalent Small Molecule Inhibitors of the Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) 3CL Protease (Jacobs et al., 2013)
62	Discovery of N-(benzo[1,2,3]triazol-1-yl)-N-(benzyl)acetamido)phenyl carboxamides as severe acute respiratory syndrome coronavirus (SARS-CoV) 3CLpro inhibitors: Identification of ML300 and noncovalent nanomolar inhibitors with an induced-fit binding Mark (Turlington et al., 2013)
63	Tanshinones as selective and slow-binding inhibitors for SARS-CoV cysteine proteases (Park et al., 2012)

64	Tanshinones as selective and slow-binding inhibitors for SARS-CoV cysteine proteases (Park, Kim, et al., 2012)
65	Phenolic phytochemical displaying SARS-CoV papain-like protease inhibition from the seeds of <i>Psoralea corylifolia</i> (Kim et al., 2014)
66	Geranylated flavonoids displaying SARS-CoV papain-like protease inhibition from the fruits of <i>Paulownia tomentosa</i> (Cho et al., 2013)
67	A noncovalent class of papain-like protease/deubiquitinase inhibitors blocks SARS virus replication (Ratia et al., 2008)
68	Structure-Based Design, Synthesis, and Biological Evaluation of a Series of Novel and Reversible Inhibitors for the Severe Acute Respiratory Syndrome-Coronavirus Papain-Like Protease (Ghosh et al., 2009)

69	Severe Acute Respiratory Syndrome Coronavirus Papain-like Novel Protease Inhibitors: Design, Synthesis, Protein-Ligand X-ray Structure and Biological Evaluation (Ghosh et al., 2010)
70	X-ray Structural and Biological Evaluation of a Series of Potent and Highly Selective Inhibitors of Human Coronavirus Papain-like Proteases (Báez-Santos et al., 2014)
71	Diarylheptanoids from <i>Alnus japonica</i> Inhibit Papain-Like Protease of Severe Acute Respiratory Syndrome Coronavirus (Park, Jeong, et al., 2012)
72	Chalcones isolated from <i>Angelica keiskei</i> inhibit cysteine proteases of SARS-CoV (Park et al., 2016)
73	Thiopurine analogues inhibit papain-like protease of severe acute respiratory syndrome coronavirus (Chou et al., 2008)

74	Targeting SARS-CoV-2: a systematic drug repurposing approach to identify promising inhibitors against 3C-like proteinase and 2'-O-ribose methyltransferase (Khan et al., 2021)
75	High-throughput screening and evaluation of repurposed drugs targeting the SARS-CoV-2 main protease (Li et al., 2021)
76	Molecular Docking Study of drug molecules from Drug Bank database against COVID-19 Mpro protein (Joshi et al., 2022)
77	Computational Evaluation of the COVID-19 3c-like Protease Inhibition Mechanism, and Drug Repurposing Screening (Hao et al., 2020)
78	Potential inhibitors against 2019-nCoV coronavirus M protease from clinically approved medicines (Liu & Wang, 2020)
79	Potentially highly potent drugs for 2019-nCoV (Nguyen et al., 2020)
80	High throughput virtual screening to discover inhibitors of the main protease of the coronavirus SARS-CoV-2 (Olubiy et al., 2020)
81	Discovery of potent Covid-19 main protease inhibitors using integrated drug-repurposing strategy (Muthu Kumar et al., 2021)
82	Fast Identification of Possible Drug Treatment of Coronavirus Disease-19 (COVID-19) through Computational Drug Repurposing Study (Wang, 2020)
83	Screening of Clinically Approved and Investigation Drugs as Potential Inhibitors of SARS-CoV-2: A Combined in silico and in vitro Study (Durdagi et al., 2022)
84	A Search for Medications to Treat COVID-19 via in silico Molecular Docking Models of the SARS-CoV-2 Spike Glycoprotein and 3CL Protease (Hall & Ji, 2020)
85	Virtual high throughput screening: Potential inhibitors for SARS-CoV-2 PLPRO and 3CLPRO proteases (Jade et al., 2021)

86	Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods (Wu et al., 2020)
87	Targeting multiple conformations of SARS-CoV2 Papain-Like Protease for drug repositioning: An in-silico study (Ismail et al., 2021)
88	Identification of SARS-CoV-2 inhibitors targeting Mpro and PLpro using in-cell-protease assay (Narayanan et al., 2022)
89	High-Throughput Screening for Drugs That Inhibit Papain-Like Protease in SARS-CoV-2 (Smith et al., 2020)
90	Drug design and repurposing with DockThor-VS web server focusing on SARS-CoV-2 therapeutic targets and their non-synonym variants (Guedes et al., 2021)

91	Repurposing clinically approved drugs for COVID-19 treatment targeting SARS-CoV-2 papain-like protease (Xu et al., 2021)
92	Hepatitis C virus drugs that inhibit SARS-CoV-2 papain-like protease synergize with remdesivir to suppress viral replication in cell culture (Bafna et al., 2021)
93	Virtual high throughput screening: Potential inhibitors for SARS-CoV-2 PLPRO and 3CLPRO proteases (Jade et al., 2021)
94	Repurposing of FDA-approved antivirals, antibiotics, anthelmintics, antioxidants, and cell protectives against SARS-CoV-2 papain-like protease (Kandeel et al., 2021)
95	Target identification for repurposed drugs active against SARS-CoV-2 via high-throughput inverse docking (Ribone et al., 2022)
96	Combined deep learning and molecular docking simulations approach identifies potentially effective FDA approved drugs for repurposing against SARS-CoV-2 (Anwaar et al., 2022)

Tabla 2 (MS). Código DOI de los 97 artículos revisados para la formación de la base de datos del presente trabajo.

No. de Artículo	DOI
1	https://doi.org/10.1016/j.sjbs.2021.01.040
2	https://doi.org/10.3390/md18040225
3	https://doi.org/10.1007/s11030-020-10151-w
4	https://doi.org/10.1080/07391102.2020.1796802
5	https://doi.org/10.1080/07391102.2020.1802347
6	https://doi.org/10.1080/07391102.2020.1842806
7	https://doi.org/10.3390/microorganisms8070970
8	https://doi.org/10.1016/j.virusres.2020.198102
9	https://doi.org/10.3390/molecules26020406
10	https://doi.org/10.1016/j.ijbiomac.2020.05.184
11	https://doi.org/10.1021/acs.jcim.0c00491
12	https://doi.org/10.1080/07391102.2020.1763202
13	https://doi.org/10.13005/BPJ/1962
14	https://doi.org/10.1016/j.joim.2020.02.005
15	https://doi.org/10.1016/j.apsb.2020.02.008
16	https://doi.org/10.1080/07391102.2020.1810778
17	https://doi.org/10.1080/07391102.2020.1764868

18	https://doi.org/10.1080/07391102.2020.1817787
19	https://doi.org/10.1080/07391102.2020.1790425
20	https://doi.org/10.1080/07391102.2020.1779818
21	https://doi.org/10.3390/computation8030079
22	https://doi.org/10.1371/journal.pone.0240653
23	https://doi.org/10.1080/07391102.2020.1841680
24	https://doi.org/10.1080/07391102.2020.1858164
25	https://doi.org/10.1021/acs.jproteome.0c00836
26	https://doi.org/10.1016/j.jmgm.2021.107851
27	https://doi.org/10.1038/s41467-021-22362-2
28	https://doi.org/10.1126/science.abg5827
29	https://doi.org/10.1021/acsptsci.0c00074
30	https://doi.org/10.1016/j.biopha.2021.111742
31	https://doi.org/10.1016/j.bmcl.2010.01.152
32	https://doi.org/10.1016/j.drudis.2020.12.005
33	https://doi.org/10.3389/fchem.2021.622898
34	https://doi.org/10.1371/journal.pone.0240079
35	https://doi.org/10.1007/s13238-021-00836-9
36	https://doi.org/10.1016/j.ejphar.2021.174082
37	https://doi.org/10.1016/j.jpha.2020.08.012
38	https://doi.org/10.1186/s13578-021-00564-x
39	https://doi.org/10.1093/pcmedi/pbab001
40	https://doi.org/10.1080/07391102.2020.1794972
41	https://doi.org/10.3389/fchem.2020.624163
42	https://doi.org/10.1080/07391102.2021.1897679
43	https://doi.org/10.1016/j.biopha.2021.111313
44	https://doi.org/10.3389/fchem.2020.594009
45	https://doi.org/10.1016/j.combiomed.2020.103936

46	https://doi.org/10.1073/pnas.2010470117
47	https://doi.org/10.1002/minf.202000028
48	https://doi.org/10.1080/14756366.2020.1801672
49	https://doi.org/10.1016/j.jiph.2020.06.016
50	https://doi.org/10.1007/s11030-021-10220-8
51	https://doi.org/10.1126/science.1085658
52	https://doi.org/10.1016/j.bmc.2007.09.034
53	https://doi.org/10.1016/j.febslet.2004.09.026
54	https://doi.org/10.1016/j.ejmech.2017.05.045
55	https://doi.org/10.2174/138161206779010369
56	https://doi.org/10.1021/jm060207o
57	https://doi.org/10.1016/j.bmc.2013.11.028
58	https://doi.org/10.1016/j.bmcl.2017.04.056
59	https://doi.org/10.1016/j.bmcl.2005.09.012
60	https://doi.org/10.1016/j.chembiol.2004.08.011
61	https://doi.org/10.1021/jm301580n
62	https://doi.org/10.1016/j.bmcl.2013.08.112
63	https://doi.org/10.1016/j.bmc.2012.07.038
64	https://doi.org/10.1016/j.bmc.2012.07.038
65	https://doi.org/10.3109/14756366.2012.753591
66	https://doi.org/10.1016/j.bmc.2013.09.005
67	https://doi.org/10.1073/pnas.0805240105
68	https://doi.org/10.1021/jm900611t
69	https://doi.org/10.1021/jm1004489
70	https://doi.org/10.1021/jm401712t
71	https://doi.org/10.1248/bpb.b12-00623
72	https://doi.org/10.3109/14756366.2014.1003215
73	https://doi.org/10.1016/j.bcp.2008.01.005
74	https://doi.org/10.1080/07391102.2020.1753577

75	https://doi.org/10.1038/s41392-021-00763-5
76	https://doi.org/10.4213/mzml3256
77	https://doi.org/10.26434/chemrxiv.12090426.v1
78	https://doi.org/10.1016/j.jgg.2020.02.001
79	https://doi.org/10.1101/2020.02.05.936013
80	https://doi.org/10.3390/molecules25143193
81	https://doi.org/10.1002/bab.2159
82	https://doi.org/10.1021/acs.jcim.0c00179
83	https://doi.org/10.1002/minf.202100062
84	https://doi.org/10.1016/j.tmaid.2020.101646
85	https://doi.org/10.1016/j.ejphar.2021.174082
86	https://doi.org/10.1016/j.apsb.2020.02.008
87	https://doi.org/10.1016/j.compbimed.2021.104295
88	https://doi.org/10.1038/s42003-022-03090-9
89	https://doi.org/10.1177/2472555220963667
90	https://doi.org/10.1038/s41598-021-84700-0
91	https://doi.org/10.1016/j.ijbiomac.2021.07.184
92	https://doi.org/10.1016/j.celrep.2021.109133
93	https://doi.org/10.1016/j.ejphar.2021.174082
94	https://doi.org/10.1080/07391102.2020.1784291
95	https://doi.org/10.1007/s10822-021-00432-3
96	https://doi.org/10.1016/j.compbimed.2021.105049
97	https://doi.org/10.1126/science.abl4784

