



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
FACULTAD DE QUÍMICA

**SELECCIÓN ASISTIDA POR COMPUTADORA DE
POSIBLES FÁRMACOS PARA EL TRATAMIENTO DE
COVID-19**

TESIS

QUE PARA OBTENER EL TÍTULO DE
QUÍMICA FARMACÉUTICA BIÓLOGA

PRESENTA

KARINA EURÍDICE JUÁREZ MERCADO



Ciudad Universitaria, CDMX

2021



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

JURADO ASIGNADO:

PRESIDENTE: Profesor: PEDRAZA CHAVERRI JOSÉ
VOCAL: Profesora: TIRADO MENDOZA ROCÍO GABRIELA
SECRETARIO: Profesor: MEDINA FRANCO JOSÉ LUIS
1er. SUPLENTE: Profesora: DEL VALLE PÉREZ PAULINA
2° SUPLENTE: Profesora: CASTREJON CARRILLO SOL

SITIO DONDE SE DESARROLLÓ EL TEMA:

EL PRESENTE TRABAJO SE DESARROLLÓ EN EL GRUPO DIFACQUIM DEL DEPARTAMENTO DE FARMACIA, UBICADO EN EL CUBÍCULO 117 DEL EDIFICIO F DE LA FACULTAD DE QUÍMICA DE LA UNAM.

ASESOR DEL TEMA:

DR. JOSÉ LUIS MEDINA FRANCO

SUSTENTANTE:

KARINA EURÍDICE JUÁREZ MERCADO

Parte de los resultados de este proyecto se difundieron en el siguiente congreso y *workshop*:

Congreso:

First Latin American Congress of Women in Bioinformatics & Data Science Virtual Edition 2020. Septiembre, 2020. Título del proyecto: “Computational Filtering of Compound Databases to Identify Potential Drug Candidates for the Treatment of COVID-19” (cartel).

Workshop:

Selección asistida por computadora de posibles fármacos para el tratamiento de COVID-19. Webinar “Posibles tratamientos de COVID por el mundo”. ACS Student Chapter, Universidad Nacional Autónoma, Costa Rica, Octubre, 2020.

Agradecimientos

A mis padres, Karina y Víctor por su amor, guía y apoyo incondicional a lo largo de mi vida. Gracias por creer en mí, sin ustedes nada de esto sería posible.

A mis hermanos, Lizzy y Jorge por ser parte de mi vida. Por todo su cariño, pero sobre todo por soportar y estar siempre en todas mis locuras. Hacen que cada día sea mejor.

A Norma por todo el amor y cariño. Por estar siempre y en todo de manera incondicional. Gracias por compartir conmigo todos esos días y noches de estudio.

A Vicky por ser mi confidente, cuidarme, apoyarme y aconsejarme. Por ser una hermana durante todos estos años.

A Chio por involucrarme en este mundo científico. Por todas las explicaciones, paciencia y motivación para estudiar esta carrera.

A Lupita, Lilí y George; por todo su amor y enseñanzas a lo largo de la vida, gracias por ser mis abuelos.

A José Luis por permitirme ser parte de tu grupo de investigación, por tu apoyo y ser mi guía. En especial, por brindarme tu amistad y dejarme compartir contigo tus vivencias y alegrías. Gracias por estar siempre ahí.

A mis compañeros del grupo DIFACQUIM por sus consejos, ayuda y todos esos instantes que hicieron más placentera mi estancia. En especial a Barbie por su paciencia, tiempo, enseñanzas y sobre todo por su amistad.

A mis amigos de la facultad y de la vida, por hacer más agradable todos los momentos en esta realidad compartida.

A los miembros del jurado: Dra. Rocío Tirado y Dr. José Pedraza; por aceptar participar en este trabajo de tesis. Gracias por su tiempo y consejos.

A la Facultad de Química de la Universidad Nacional Autónoma de México por ofrecerme el espacio y herramientas necesarias para obtener mi formación académica.

Al Taller de Ciencia para Jóvenes del Centro de Geociencias de la UNAM, proyecto PAPIIME-**PE100915**: “Simiente de Investigadores, de la DGAPA, por haberme inducido hacia la investigación.

Al Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT), proyecto **IV200121**.

A la dirección General de Cómputo y de Tecnologías de Información y Comunicación (DGTIC) UNAM por proveer los recursos computacionales para usar la supercomputadora Miztli, proyecto **LANCAD-UNAM-DGTIC-335**).

Se agradece la colaboración de la Dra. Clara Inés Espitia Pinzón (Instituto de Investigaciones Biomédicas, UNAM) y del Dr. Marco Antonio Velasco Velázquez (Facultad de Medicina, UNAM), co-responsables del proyecto PAPIIT **IV200121**.

Se agradece la colaboración para efectuar las predicciones de aprendizaje de máquina con el Dr. Sean Ekins (Collaborations Pharmaceuticals Inc.).

Se agradece la colaboración para realizar las predicciones QSAR con el Dr. Vinicius Alves y el Dr. Alexander Tropsha (Universidad de Carolina del Norte en Chapel Hill Estados Unidos).

Al CONACyT por el apoyo para ayudante de investigador Nivel III (No. **16231**).

Der Unterschied zwischen Vergangenheit, Gegenwart und Zukunft ist nur eine Illusion, wenn auch eine hartnäckige...

Albert Einstein.

Índice

Abreviaturas	9
Resumen	13
1. Antecedentes	15
1.1 Virus de RNA (+)	15
1.2 Características de los coronavirus	15
1.3 Características del SARS-CoV-2	17
1.4 Proteínas virales	18
1.4.1 Proteína S	18
1.4.2 Proteína N	19
1.4.3 Proteína M	19
1.4.4 Proteína E	19
1.4.5 Proteína no estructural nsp5	20
1.5 Ciclo replicativo	20
1.6 Patogénesis e inmunidad	22
1.6.1 Respuesta inmune innata	22
1.6.2 Respuesta inmune adaptativa	23
1.7 Epidemiología	24
1.8 Infección	25
1.8.1 Signos y síntomas clínicos	26
1.8.2 Diagnóstico	27
1.9 Tratamientos	28
1.9.1 Búsqueda de moléculas pequeñas como antivirales	28
1.9.2 Reposicionamiento de fármacos	28
1.9.3 Uso de anticuerpos	28
1.9.4 Empleo de citocinas	29
1.9.5 Vacunas	29
2.1 General	30
2.2 Particulares	30
3. Hipótesis	31
4. Métodos	32
4.1 Generar y preparar una amplia base de datos molecular (MEGA Db)	33
4.2 Crear y curar la lista de moléculas como referencias (queries)	34
4.3 Filtrar la base de datos con análisis de similitud molecular vs. las moléculas de referencia	38
4.4 Análisis consenso de los resultados de similitud molecular	39
4.5 Búsqueda de disponibilidad comercial	39
4.6 Determinar el perfil ADME/Tox de las moléculas disponibles comercialmente	39
4.6 Realizar predicciones de aprendizaje de máquina y QSAR	41
4.7 Acoplamiento molecular con la proteasa principal de SARS-CoV-2	42
4.8 Adquisición comercial de compuestos y envío a evaluación biológica	43

5. Resultados y discusión	44
5.1 Generación y preparación de la MEGA Db	44
5.2 Lista de moléculas de referencia (queries)	44
5.3 Filtrado de la base de datos con similitud molecular vs las moléculas de referencia	45
5.4 Análisis consenso de los resultados de similitud molecular	53
5.5 Búsqueda de disponibilidad comercial	56
5.6 Perfil ADME/Tox de las moléculas disponibles comercialmente	56
5.6.1 Descriptores moleculares de absorción	57
5.6.1.1 Solubilidad	60
5.6.1.2 Lipofilicidad	61
5.6.1.3 Absorción intestinal humana	63
5.6.2 Descriptores moleculares de distribución	65
5.6.2.1 Permeabilidad de la barrera hematoencefálica	65
5.6.2.2 Sustrato de la glicoproteína P	67
5.6.2.3 Inhibición de la glicoproteína P	68
5.6.3 Descriptores moleculares de metabolismo	70
5.6.3.1 Inhibición del CYP1A2	71
5.6.3.2 Inhibición del CYP2C19	72
5.6.3.3 Inhibición del CYP2C9	72
5.6.3.4 Inhibición del CYP2D6	73
5.6.3.5 Inhibición del CYP3A4	73
5.6.4 Descriptores moleculares de eliminación	74
5.6.4.1 Eliminación total	74
5.6.5 Descriptores moleculares de toxicidad	75
5.6.5.1 Prueba in silico de AMES	76
5.6.5.2 Hepatotoxicidad	77
5.7 Predicciones de aprendizaje de máquina y QSAR	78
5.8 Acoplamiento molecular con la proteasa principal de SARS-CoV-2	79
5.9 Adquisición comercial de compuestos y envío a evaluación biológica	81
6. Conclusiones	85
7. Perspectivas	86
8. Referencias	87
9. Apéndice	109

Abreviaturas

ΔG : Energía Libre de Gibbs.

3CLpro, nsp5: Proteasa Tipo 3CL, *3C Like protease*.

ABC: Área Bajo la Curva.

ABL1: Oncogén Homólogo 1 del Abensol de Leucemia Murina, *ABL Proto-Oncogene 1, Non Receptor Tyrosine Kinase*.

AdaBoost: *Adaptative Boosting*.

ADME/Tox: Perfil de Administración, Distribución, Metabolismo, Excreción y Toxicidad.

ADN: Ácido desoxirribonucleico.

AP-1: Proteína Activadora 1, *Activador Protein 1*.

API: Interfaz de Programación de Aplicaciones, *Application Programming Interfaces*.

ARN: Ácido ribonucleico.

ARNm: Ácido ribonucleico mensajero.

ARNmc: Ácido ribonucleico monocatenario.

ATP: Adenosin Trifosfato, *Adenosin Triphosphate*.

AutoDock Vina: *Autodock Vina*, versión 1.1.2.

BBB: Barrera Hematoencefálica, *Blood Brain Barrier*.

CAS: Biblioteca libre enfocada hacia el COVID-19 publicada por la *American Chemical Society*.

CD: Antígeno de Diferenciación, *Cluster of Differentiation*.

ChEMBL: Biblioteca *ChEMBL*, versión 25.

CI₅₀: Concentración Inhibitoria 50.

COCONUT: Biblioteca de productos naturales de la base de datos pública *ColleCtion of Open NatUral producTs*.

CoVs: Coronavirus.

CRT: Complejo de Replicación-Transcripción.

CYPs: Citocromos.

DC: Células Dendríticas, *Dendritic Cells*.

DCM: Biblioteca de compuestos obtenida de la base de datos *Dark Chemical Matter*.

DiFAC: Diseño de Fármacos Asistido por Computadora.

Drug-like: Propiedades tipo fármaco.

DTT: Ditiotreitól.

ECA2: Enzima Convertidora de Angiotensina 2, *Angiotensin-converting enzyme 2*.

ECFP4: Huellas Digitales de Conectividad Extendida radio 2, *Extended Connectivity Fingerprints radio 2*.

ECFP6: Huellas Digitales de Conectividad Extendida radio 3, *Extended Connectivity Fingerprints radio 3*.

Enamine: Biblioteca de compuestos de la base de datos *Readily Accessible (REAL)-Enamine*.

EPOC: Enfermedad Pulmonar Obstructiva Crónica.

ESR1: Receptor de Estrógenos 1, *Estrogen Receptor 1*.

FDA PEP: Biblioteca de péptidos aprobados por la *Food and Drug Administration*.

FoodDB: Biblioteca de datos pública de compuestos dietéticos.

G-SCF: Factor Estimulante de Colonias de Granulocitos, *Granulocyte Colony-Stimulating Factor*.

HCoV-NL63: coronavirus humano NL63, *Human Coronavirus NL63*.

HCoV-OC43: coronavirus humano OC43, *Human Coronavirus OC43*.

HIA: Absorción Intestinal Humana, *Human Intestinal Absorption*.

Hits: Lista de compuestos con resultados positivos en las predicciones QSAR y de Aprendizaje Automático.

ICTV: Comité Internacional de Taxonomía de Virus, *International Committee on Taxonomy of Viruses*.

IFNs: Interferones, *Interferons*.

IgA: Inmunoglobulina A, *Immunoglobulin A*.

IgG: Inmunoglobulina G, *Immunoglobulin G*.

IgM: Inmunoglobulina M, *Immunoglobulin M*.

IL: Interleucina, *Interleukin*.

IRF: Factor Regulador de Interferón, *Interferon Regulatory Factor*.

LogP: Logaritmo decimal del cociente de la solubilidad de un compuesto en n-octanol y agua.

LogPS: Logaritmo decimal de la permeabilidad cuantitativa en la Barrera Hematoencefálica.

LogS: Logaritmo decimal de la solubilidad molar de un compuesto.

MACCSKeys: *Molecular ACCess System Keys*.

MCP-1: Proteína Quimioatrayente de Quimocitos 1, *Monocyte Chemoattractant Protein-1*.

MEGA Db: Base de datos molecular extensa de > 17 millones de compuestos.

MERS-CoV: Coronavirus causante del Síndrome Respiratorio del Oriente Medio, *Middle East Respiratory Syndrome coronavirus*.

MERS: Síndrome Respiratorio del Oriente Medio, *Middle East Respiratory Syndrome*.

MHV-A59: Coronavirus Murino del Virus de la Hepatitis de ratón A-59, *Murine coronavirus mouse Hepatitis Virus A-59*.

MOE: *Molecular Operating Environment*, version 2018.

MP1A: Proteína Inflamatoria de Macrófagos 1A, *Macrophage Inflammatory Protein 1A*.

Mpro: Proteasa principal, *Main protease*.

NF-κB: Factor Nuclear kappa Beta, *Transcription Nuclear Factor kappa Beta*.

NK: Células Asesinas, *Natural Killers*.

NLR: Dominio de Oligomerización de Unión a Nucleótidos, *NOD-Like Receptors*.

Nsp: Proteína no estructural, *non structural protein*.

OMS: Organización Mundial de la Salud.

ORFs: Marcos de Lectura Abiertos, *Open Reading Frames*.

PDB: Base de datos de proteínas, *Protein Data Bank*.

PLpro, nsp3: Proteasa Similar a la Papaína, *Papain Like protease*.

pp1a: Poliproteína 1a.

pp1ab: poliproteína 1ab.

Proteína E: Proteína de envoltura.

Proteína M: Proteína transmembranal.

Proteína N: Proteína de nucleocápside.

Proteína S: Proteína de espiga.

PRRs: Receptores de Reconocimiento de Patrones, *Pathogen Recognition Receptors*.

QPNB: Biblioteca de productos naturales obtenidos de la base de datos Colombiana *QuiProNaB*.

QSAR: Relaciones Cuantitativas Estructura-Actividad, *Quantitative Structure-Activity Relationship*.

RBD: Dominio de Unión al Receptor, *Receptor Binding Domain*.

RdRp, nsp12: ARN polimerasa dependiente de ARN, *RNA dependent RNA polymerase gene*.

RIG-I: Gen I inducible por Ácido Retinoico, *Retinoic Acid-Inducible Gen I*.

RT-PCR: Reacción de la Cadena de la Polimerasa empleando una Transcriptasa Inversa en Tiempo Real, *Reverse Transcription Polymerase Chain Reaction*.

SARS-CoV-2: Coronavirus tipo 2 del Síndrome Respiratorio Agudo Grave, *Severe Acute Respiratory Syndrome coronavirus 2*.

SARS-CoV: Coronavirus tipo 1 del Síndrome Respiratorio Agudo Grave, *Severe Acute Respiratory Syndrome coronavirus*.

SARS: Síndrome Respiratorio Agudo Grave, *Severe Acute Respiratory Syndrome*.

Scaffolds: Núcleos base empleando la definición de Bemis y Murcko.

SDRA: Síndrome de Dificultad Respiratoria Aguda Grave.

SiRMS: 2D simplex Representation of Molecular Structure.

SMILES: *Simplified Molecular-Input Line-Entry System*.

SNC: Sistema Nervioso Central.

TLR: Receptor Tipo Toll, *Toll Like Receptor*.

TMPRSS2: Proteasa Transmembranal 2 de la Serina, *Transmembranal Serine Protease 2*.

TNF- α : Factor de Necrosis Tumoral Alfa, *Tumor Necrosis Factor-Alpha*.

TPSA: Área Polar Superficial, *Topological Polar Surface Area*.

VIH: Virus de la Inmunodeficiencia Humana.

VLP: Partículas Similares a Virus, *Virus Like Particle*.

ZINC: Base de datos pública *ZINC 15*.

Resumen

Los coronavirus (CoVs) son una familia de virus de ácido ribonucleico (ARN) que tienen como principal característica la presencia de glicoproteínas en la superficie, las cuales aparentan tener forma de coronas. Se encuentran en la naturaleza en diferentes especies incluidos mamíferos y aves. Algunas de las principales patologías causadas por los CoVs son enfermedades respiratorias, gastrointestinales y hepáticas.

En diciembre del 2019 se descubrió en Wuhan, China, un nuevo coronavirus llamado provisionalmente 2019-nCoV y posteriormente Coronavirus tipo 2 del Síndrome Respiratorio Agudo Grave (SARS-CoV-2, por sus sigla en inglés) por el Comité Internacional de Taxonomía de Virus, el cual está relacionado con infecciones del tracto respiratorio y la enfermedad denominada como COVID-19. Esta enfermedad se dispersó rápidamente en el mundo, causando una pandemia global reconocida por la Organización Mundial de la Salud (OMS) en marzo del 2020. Por lo que diversos investigadores se encuentran explorando posibles tratamientos contra esta enfermedad.

Desde el 2020 hasta la fecha se han publicado artículos que reportan algunos compuestos antivirales como el remdesivir, baricitinib y la cloroquina como posibles tratamientos a esta enfermedad. Sin embargo, actualmente no hay agentes antivirales específicos contra el SARS-CoV-2. Por este motivo, es necesaria la investigación y desarrollo de moléculas pequeñas que puedan tener potencial inhibitorio contra proteínas claves para el SARS-CoV-2, como es el caso de la proteasa principal (Mpro), para la cual se cuenta con la información de su estructura cristalográfica reportada en complejo con el inhibidor MPI1 (código del fabricante) en el sitio catalítico.

Una estrategia común para identificar candidatos potenciales contra el COVID-19 es filtrar bases de datos de compuestos para encontrar candidatos a fármacos contra algunos de los principales blancos moleculares que puedan inactivar de manera directa al SARS-CoV-2. Una manera rápida de llevar a cabo esto es utilizando métodos establecidos dentro del Diseño de Fármacos Asistido por Computadora (DiFAC), empleando un análisis quimioinformático y un filtrado computacional de bases de datos. Este proceso ayuda a reducir el tiempo y costo en comparación con pruebas experimentales masivas de bases de datos de compuestos.

En el presente trabajo se llevó a cabo como primera estrategia un filtrado o cribado computacional a partir de un análisis de similitud molecular usando como moléculas de referencia compuestos activos contra coronavirus homólogos al SARS-CoV-2. Posteriormente se implementaron modelos de aprendizaje automático y relaciones cuantitativas de estructura-actividad (QSAR, por sus siglas en inglés), así como acoplamiento molecular empleando la estructura cristalográfica de la Mpro del SARS-CoV-2 (PDB ID: 6LU7) para obtener una lista de compuestos, denominados *hits* computacionales. Finalmente, se llevó a cabo la evaluación *in vitro* para conocer la actividad inhibitoria de los *hits* computacionales contra la Mpro del SARS-CoV-2.

1. Antecedentes

1.1 Virus de RNA (+)

Los virus de ácido ribonucleico (ARN) monocatenarios de polaridad positiva pertenecen al grupo IV de la clasificación de Baltimore. Debido a que los genomas de los virus que pertenecen a esta clasificación tienen la misma polaridad del ARN mensajero del huésped, estos pueden llevar a cabo la traducción de su genoma directamente (sin necesidad de crear primero una cadena de ácido desoxirribonucleico, ADN). Este genoma codifica para diversas proteínas (Figura 1).¹

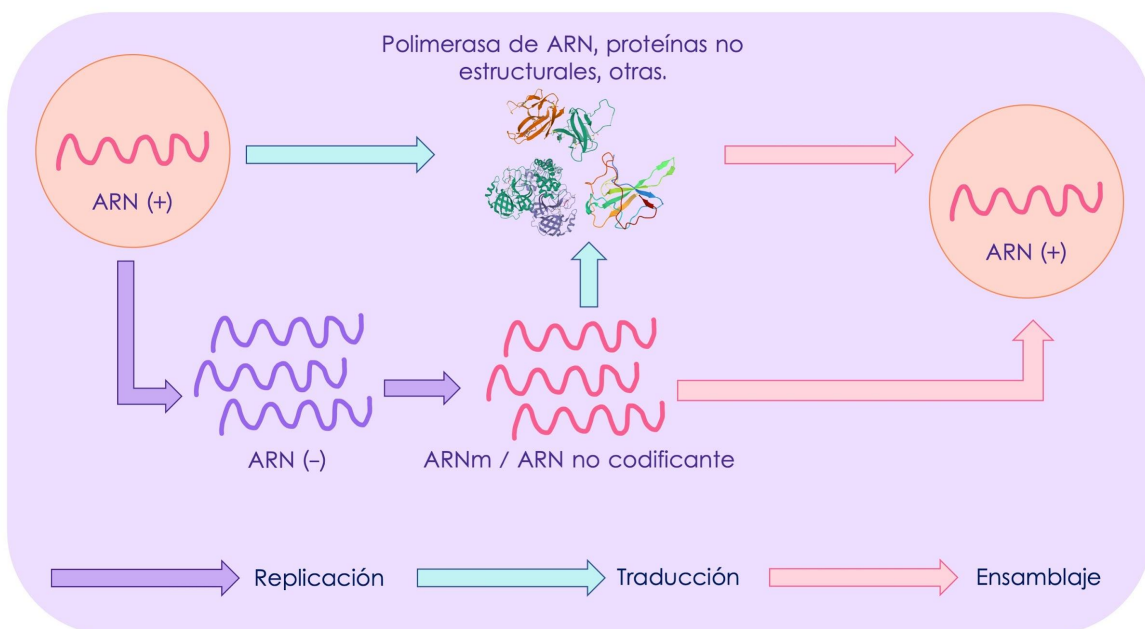


Figura 1. Proceso de replicación, traducción y ensamblaje del virión de los virus que pertenecen a la clasificación IV de Baltimore a partir del ácido ribonucleico (ARN) y el ácido ribonucleico mensajero (ARNm).

1.2 Características de los coronavirus

Los coronavirus (CoVs) son una familia de virus envueltos de ARN de cadena positiva que pertenecen a la familia *Coronaviridae* y al orden de los *Nidovirales*. Los CoVs se encuentran en diferentes especies, como es el caso de las aves y los mamíferos. Los CoVs (Figura 2) son los virus de ARN con los genomas más

grandes, teniendo un tamaño entre 26 y 32 kb. Están formados principalmente por cuatro proteínas estructurales; la proteína de nucleocápside (N), la proteína transmembranal (M), la proteína de envoltura (E) y la proteína de espiga (S).^{2,3}

Los CoVs se pueden diferenciar en cuatro géneros: *alfa*, *beta*, *gamma* y *delta*; siendo los *alfa* y *beta* los causantes de las enfermedades en humanos⁴. Algunas de las principales patologías causadas por los CoVs van desde las gripes comunes, infecciones intestinales, infecciones respiratorias, e incluso en algunos casos se pueden llegar a presentar desórdenes neurológicos o la muerte.^{5,6}

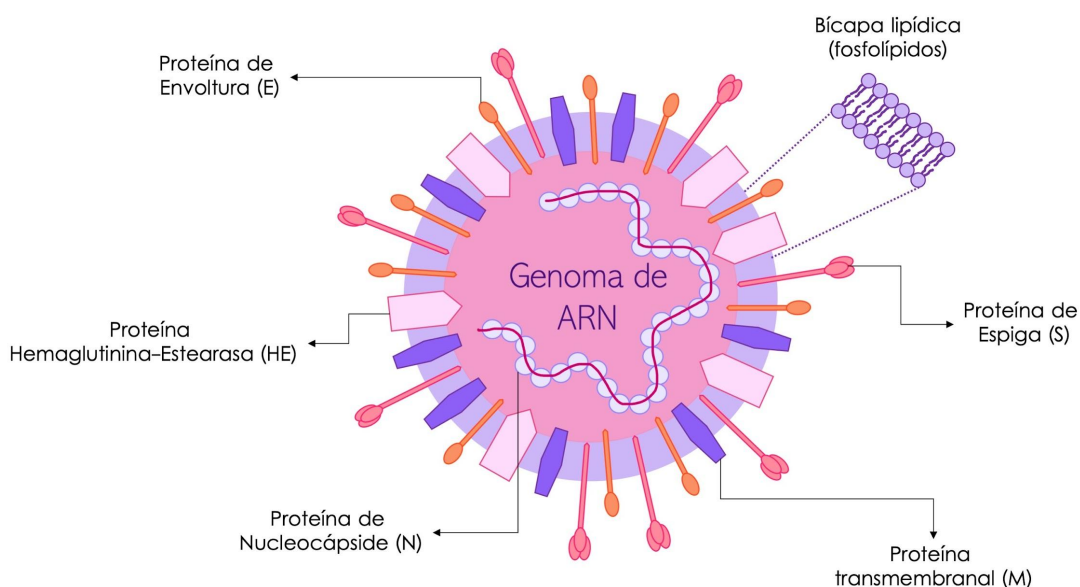


Figura 2. Representación de β-coronavirus de ARN monocatenario de sentido positivo y sus proteínas estructurales.

En las dos décadas pasadas los CoVs han sido los responsables de causar dos pandemias importantes en el mundo. La primera, denominada como “Síndrome respiratorio agudo grave” (SARS) en 2002-2003 originada en Guangdong, una provincia en China, causada por el virus Coronavirus tipo 1 del Síndrome Respiratorio Agudo Grave (SARS-CoV, por sus siglas en inglés). La segunda como “Síndrome respiratorio de oriente medio” (MERS) en 2012; teniendo principal incidencia en países del medio oeste, y siendo originada por el virus Síndrome Respiratorio del Oriente Medio (MERS-CoV, por sus siglas de inglés). Acumulando más de 10 mil casos entre ambas; teniendo el SARS-CoV una mortalidad del 10% y el MERS-CoV del 37%.^{6,7}

En diciembre del 2019 inició en China un brote de neumonía con causa desconocida en Wuhan (una provincia de Hubei). A este brote se le atribuyó un nuevo coronavirus a principios de enero del 2020. Este nuevo CoV fue llamado provisionalmente como nCoV-2019, y posteriormente como Coronavirus tipo 2 del Síndrome respiratorio agudo grave (SARS-CoV-2, por sus siglas en inglés) por el Comité Internacional de Taxonomía de Virus (ICTV, por sus siglas en inglés).^{8,9} Este virus es el agente causal de la enfermedad denominada como COVID-19.

Para el 5 de marzo del 2020, fueron diagnosticados más de 96 mil casos, y ya habían 3,300 muertes confirmadas por SARS-CoV-2.¹⁰ Debido a que este virus se diseminó rápidamente a través de China y otros países de Asia, del medio Este, África, Oceanía, América y Europa en tan solo unas semanas, el 11 de marzo del 2020 la enfermedad de COVID-19 se declaró como pandemia mundial por la Organización Mundial de la Salud (OMS).¹¹ Al 14 de agosto del 2021 se han reportado más de 206 millones de casos y más de 4 millones de muertes.¹²

1.3 Características del SARS-CoV-2

El SARS-CoV-2 es un betacoronavirus envuelto de ARN monocatenario de sentido positivo con un cap en el extremo 5' y una cola de poli-A en el extremo 3'.¹³ El genoma de SARS-CoV-2 (Figura 3) tiene aproximadamente 28.8 kb de longitud y contiene 14 marcos de lectura abiertos (ORFs) que codifican para 27 proteínas.

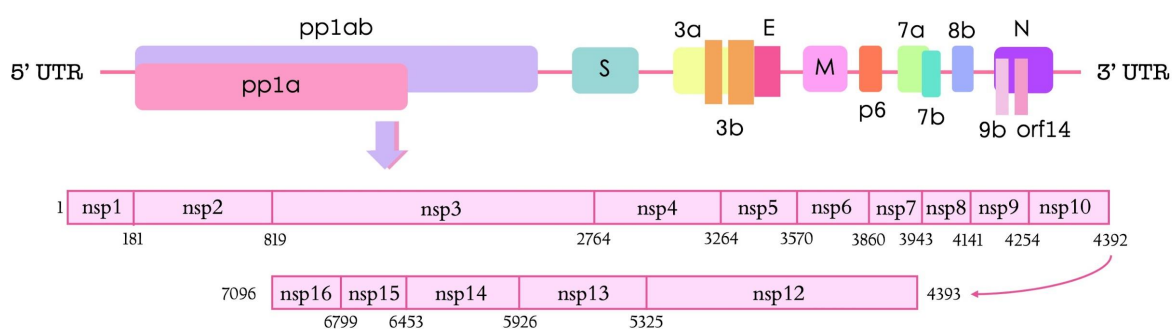


Figura 3. Organización del genoma de ácido ribonucleico mensajero (RNAm) monocatenario del Coronavirus tipo 2 del Síndrome Respiratorio Agudo Grave (SARS-CoV-2). En el que se representan las regiones que codifican para las diferentes proteínas. En sentido 5' a 3' se encuentra el segmento de RNAm que codifica para los genes de cada una de las 27 proteínas. Comenzando con las poliproteínas pp1a (pp1a) y pp1ab (pp1ab) las cuales a su vez se encuentran

estructuradas por las proteínas no estructurales (nsp) de la nsp1-nsp16, los números indican el número de nucleótidos en la secuencia en el que inicia y termina cada una de las nsp. Posteriormente se encuentran las regiones para la proteína de espiga (S), la proteína 3a, 3b, la proteína de envoltura (E), la proteína de membrana (M), la proteína p6, 7a, 7b, 8b, 9b y la proteína de nucleocápside (N).

1.4 Proteínas virales

Los genes *orf1ab* y *orf1a* que se localizan en el extremo 5' terminal del genoma codifican respectivamente para las poliproteínas 1ab (pp1ab) y 1a (pp1a), las cuales posteriormente se dividen en 15 proteínas no estructurales (nsp), las cuales incluyen la nsp1, nsp2, nsp3, nsp4, nsp5, nsp6, nsp7, nsp8, nsp9, nsp10, nsp12, nsp 13, nsp14, nsp15 y nsp16. Por otra parte, el extremo terminal 3' del genoma codifica para las cuatro proteínas estructurales: S, E, M y N, y para ocho proteínas accesorias: 3a, 3b, p6, 7a, 7b, 8b, 9b y el orf14.^{14,15}

Las proteínas nsp3 (proteasa similar a la papaína; PLpro, por sus siglas en inglés), nsp5 (proteasa tipo 3CLpro, proteasa principal: Mpro), nsp12 (ARN polimerasa dependiente de ARN; RdRp, por sus siglas en inglés) y la nsp13 (helicasa) (Figura 3) son proteínas indispensables para el ciclo de vida del virus. Por otra parte, la proteína S es la responsable de la interacción virus-receptor para llevar a cabo la entrada por parte del virus al hospedero. De esta manera, las cinco proteínas mencionadas son dianas biológicas atractivas para la búsqueda de posibles compuestos contra el SARS-CoV-2.^{13,14} En las siguientes subsecciones se describe cada una de ellas.

1.4.1 Proteína S

Es una proteína de fusión homotrimérica de clase I conformada por 1273 aminoácidos. Es la proteína más grande de las cuatro proteínas estructurales. La proteína S se puede dividir en dos subunidades funcionales, llamadas S1 y S2. La subunidad S1 está formada por una cabeza globular o con forma de pétalo y el dominio de unión al receptor (RBD, por sus siglas en inglés). Por otra parte, la subunidad S2, se encuentra formada por dos regiones heptadas (HR1 y HR2), un dominio transmembranal y una cola citosólica. La proteína S, es la encargada de

llevar a cabo las interacciones entre el virus y las células del huésped (fusión de sus membranas).¹⁶

1.4.2 Proteína N

Es una glicoproteína de 422 aminoácidos, compuesta de dos dominios estructurales. Es la única proteína encargada de formar la nucleocápside. Una de sus principales funciones es llevar a cabo el ensamblaje del genoma de ARN de los coronavirus. Además, está involucrada con procesos relacionados con el genoma viral. También tiene funciones importantes en el ciclo de replicación del ARN viral y en el proceso de la respuesta celular del huésped contra la infección viral.^{3,17}

1.4.3 Proteína M

Es una glicoproteína de 222 aminoácidos formada por tres dominios transmembranales. La glicosilación de esta proteína en el Aparato de Golgi es de vital importancia para que pueda llevar a cabo sus funciones; entre las cuales se encuentran permitir que se lleve a cabo la fusión con la célula del huésped y que tenga antigenicidad. Esta proteína también interactúa en el Retículo Endoplásmico y el Aparato de Golgi para que se lleve a cabo la formación de nuevos viriones.¹⁸

1.4.4 Proteína E

Es una glicoproteína pequeña, de 76 a 109 aminoácidos, la cual se encuentra anclada en la envoltura viral. Las principales funciones de esta proteína son: regular las posibles interacciones que puedan conllevar a la agregación entre proteínas M, cambiar la organización del Aparato de Golgi para que este sea capaz de producir vesículas de mayor tamaño que sean capaces de transportar viriones e interactuar con diversos factores del huésped que pudieran interferir con la patogénesis viral; además de tener un papel importante morfogénesis y formación del virión.^{18,19}

1.4.5 Proteína no estructural nsp5

Es una cisteína proteasa de 306 aminoácidos. También es conocida como la “proteasa principal” del SARS-CoV-2 (Mpro). Es una proteína altamente conservada y está formada por un dedo conformado por los residuos 1 a 8, el cual contiene el N terminal, el dominio catalítico conformado por los residuos 8 al 184 y el dominio con el C terminal formado por los residuos 201 al 306. Tiene una estructura similar a la de la quimotripsina. Debido a lo anterior también se le conoce como proteasa de tipo 3C o 3CLpro (proteasa tipo 3C, por sus siglas en inglés). Su función es escindir a las proteínas nsp: de la nsp4 a la nsp11 de la poliproteína pp1a, y de la nsp4 a nsp16 de la poliproteína pp1ab. Esto con el fin de generar proteínas funcionales individuales entre las que se encuentran la polimerasa de ARN dependiente de ARN, una helicasa, una proteína de unión a ARN monocatenario, una exoribonucleasa, una endorribonucleasa y una 2'-O-ribosa metiltransferasa.^{20,21}

1.5 Ciclo replicativo

El ciclo replicativo del SARS-CoV-2 (Figura 4) se encuentra mediado por el receptor humano que permite la entrada de éste. Dicho receptor es la enzima convertidora de angiotensina II (ECA2), el cual es también conocido por ser el receptor del SARS-CoV. La ECA2 es una carboxipeptidasa de membrana de 805 residuos de aminoácidos. Esta se expresa en células epiteliales respiratorias y en las células alveolares pulmonares del tipo I y II. Por lo que se encuentra en la mayoría de los tejidos, y del epitelio pulmonar.²²

La entrada del virus depende de la activación proteolítica de la proteína S. El primer paso en el proceso de entrada a las células del huésped por parte del virus se encuentra mediado por el RBD de la subunidad 1 de la proteína S el cual se une a la ECA2, esto conlleva a exponer el sitio de escisión sobre la subunidad S2 de la proteína S, sobre el que actúan las proteasas de la célula del huésped, como es el caso de la proteasa transmembranal 2 de la serina (TMPRSS2, por las siglas en inglés).¹⁶

Una vez que se ha llevado a cabo la fusión y escisión, el virus entra a la célula y el ARN del genoma del virus se libera en el citoplasma de la célula del huésped. Donde posterior a esto las dos poliproteínas codificadas en el genoma

(pp1a y pp1ab) son capaces de traducirse de manera directa usando la maquinaria celular del huésped. Estas poliproteínas a su vez son las que codifican para las proteínas nsp y son capaces de formar el complejo de replicación-transcripción (CRT) en forma de vesícula de doble membrana. El CRT tiene como principal función replicar y sintetizar otro conjunto de ARN subgenómicos, los cuales codifican para las proteínas estructurales. Así mismo, el Aparato de Golgi y Retículo Endoplásmico continúan sintetizando a las proteínas N y E y el ARN genómico para formar nuevas partículas virales. Finalmente las vesículas que contienen a los viriones se pueden fusionar con las membrana y liberar a las partículas virales.^{23,24}

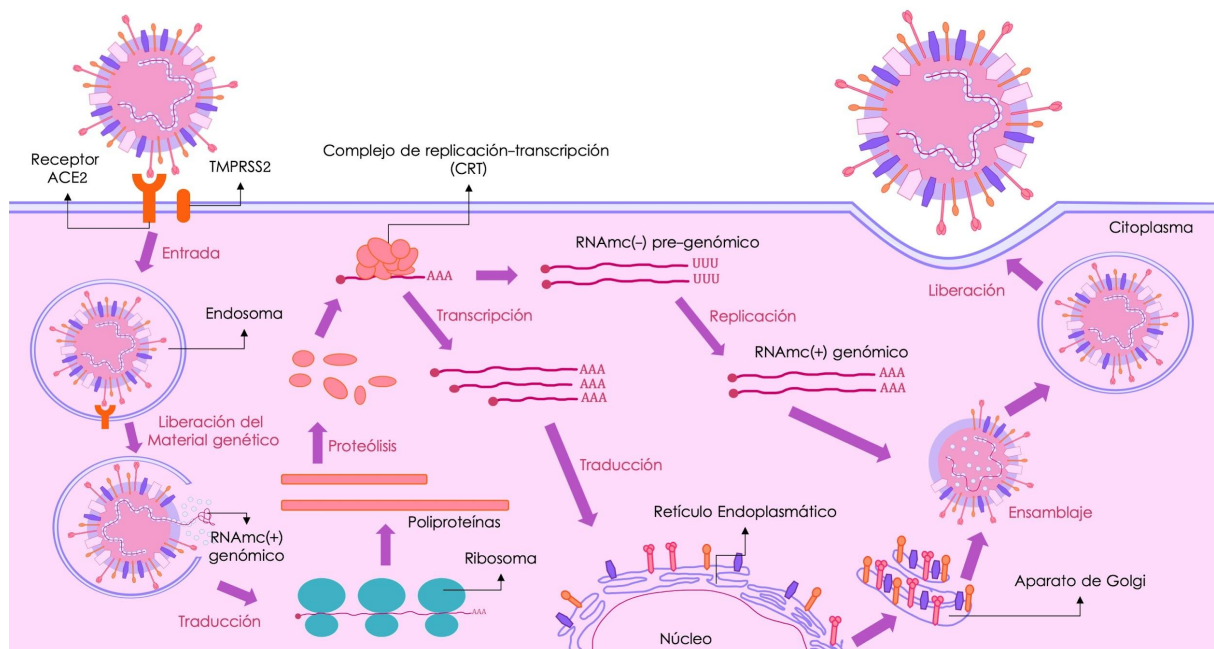


Figura 4. Ciclo de replicación del coronavirus tipo 2 del Síndrome Respiratorio Agudo Grave (SARS-CoV-2). Se muestra la etapa inicial del ciclo de vida del SARS-CoV-2, que comienza con la unión de la proteína de espiga (S) al receptor enzima convertidora de angiotensina 2 (ACE2, por sus siglas en inglés) del huésped que es seguida por la entrada viral mediada por endocitosis o fusión directa. La entrada por endocitosis da lugar a los endosomas, mientras que la fusión directa requiere ser mediado por la proteasa transmembranal 2 de la serina (TMPRSS2, por sus siglas en inglés). Posteriormente se lleva a cabo una proteólisis en la cual el material genético, en este caso ácido ribonucleico monocatenario (ARNmc) del SARS-CoV-2 es liberado. Lo cual permite que se lleve a cabo el proceso de

replicación del virus a través de la síntesis de proteínas virales, la maduración, el ensamblaje y la liberación de las nuevas partículas virales.

1.6 Patogénesis e inmunidad

El SARS-CoV-2 se transmite principalmente por medio de microgotas o gotas de Flugge, por la vía fecal-oral, o por contacto directo. Ya sea por medio del contacto entre las mucosas (transmisión directa) o por contacto de una superficie infectada con alguna mucosa nasal, bucal o conjuntival (transmisión indirecta).²⁵

La replicación viral primaria ocurre en el epitelio de la mucosa del tracto respiratorio superior (la cavidad nasal y la faringe), mientras que la multiplicación viral se da principalmente en el tracto respiratorio inferior y en el tracto gastrointestinal, lo que conlleva a una viremia leve; sin embargo, son pocas las infecciones que se pueden controlar en este punto o que permanecen asintomáticas.²

El periodo de incubación puede ir de uno a catorce días. Los casos más graves pueden presentar diarrea, daño hepático, daño cardiaco, insuficiencia renal e incluso la muerte. Debido a que la ECA2 se expresa en la mucosa nasal, los bronquios, el esófago, los pulmones, el corazón, los riñones, el estómago, la vejiga y el intestino son órganos que pueden verse afectados por el SARS-CoV-2.^{2,26,27,28}

1.6.1 Respuesta inmune innata

Después de que las partículas virales ingresan a las células del huésped, el virus es reconocido por receptores de reconocimiento de patrones (PRRs, por sus siglas en inglés) como el receptor de tipo Toll 7 y 8 (TLR7 y TLR8, por sus siglas en inglés), receptores del gen I inducible por ácido retinoico (RIG-I, por sus siglas en inglés) y receptores similares al dominio de oligomerización de unión a nucleótidos (NLR, por sus siglas en inglés), los cuales se expresan tanto en células epiteliales como en células locales de la respuesta inmune innata, como es el caso de los macrófagos alveolares.²⁹

Posterior a la unión de su ligando, los PRRs reclutan a proteínas que activan diversos factores de transcripción, entre los que se encuentran el factor regulador de interferón (IRF, por sus siglas en inglés), factor nuclear kappa beta (NF-κB, por sus

siglas en inglés) y la proteína activadora 1 (AP-1, por sus siglas en inglés). Estos PRRs conllevan a la activación de diferentes quimiocinas y de interferones (IFNs, por sus siglas en inglés) antivirales del tipo I y III. Dichas quimiocinas atraen a otras células de la respuesta inmune innata como lo son los leucocitos, monocitos, células polimorfonucleares, células asesinas naturales (NK, por sus siglas en inglés) y células dendríticas (DC, por sus siglas en inglés). Que a su vez son capaces de reconocer a los antígenos virales presentados por las DC.^{30,31}

Diversos estudios sugieren que el SARS-CoV-2 se diferencia de otros coronavirus por su capacidad de evitar los efectos antivirales de los IFNs I y III, además de activar a la respuesta inmune innata y de inducir la producción de citocinas gracias a la activación de células de la respuesta inmune adaptativa.^{32,33}

Así mismo, estudios en niños que estuvieron enfermos de COVID-19 por SARS-CoV-2 muestran una disminución significativa en células T con antígeno de diferenciación 4 positivo, 8 positivo y 3 positivo (CD4+, CD8+ y CD3+, por sus siglas en inglés) en sangre. Esto sugiere que la inmunidad celular se ve gravemente afectada posterior a la infección.³⁴

1.6.2 Respuesta inmune adaptativa

Varios estudios han demostrado que la respuesta humoral se da principalmente por anticuerpos contra la proteína S, los cuales impiden la unión entre el virus y la ECA2. La detección de anticuerpos del tipo de inmunoglobulina M y A (IgM e IgA, por sus siglas en inglés) pueden detectarse a partir de la primera semana del inicio de la sintomatología, mientras que los del tipo de inmunoglobulina G (IgG, por sus siglas en inglés) se detectan aproximadamente 14 días después. Sin embargo, aún no se sabe por cuánto tiempo son capaces de permanecer activos a niveles que confieran protección.^{35,36,37}

La respuesta protectora proviene principalmente por parte de las células T con ayuda de células B cooperadoras CD4+, para producir anticuerpos neutralizantes específicos y células CD8+ citotóxicas; las cuales tienen la capacidad de eliminar las células infectadas.^{38,39}

Una respuesta inmune adaptativa disfuncional no es capaz de inhibir la replicación del virus ni de eliminar a las células infectadas. Esto conduce a una sobreproducción y sobreexpresión de citocinas, lo que se puede manifestar en

consecuencias sistémicas como la diseminación de coagulación intravascular o el síndrome de dificultad respiratoria aguda (SDRA) grave.⁴⁰

A pesar de que las células B, las células T, los macrófagos y las DC no expresan a la ECA2; hay evidencia que el receptor DC-SIGN puede servir como receptor trans entre las partículas virales y las DC.^{31,41,42} Además se ha reportado que la molécula proteína complejante de adenosina desaminasa 2, también conocida como dipeptidil peptidasa 4 o antígeno de diferenciación 26 (CD26, por sus siglas en inglés) es una aminopeptidasa capaz de unirse a la proteína S del SARS-CoV-2 lo cual resulta en una infección de células T no productivas.⁴³ Por otra parte, se sabe que el inductor de metaloproteínasa de matriz extracelular o antígeno de diferenciación 147 (CD147, por sus siglas en inglés) es una proteína de la superfamilia de las inmunoglobulinas que es capaz de inducir metaloproteinasas de la matriz celular. A su vez, estas son capaces de unirse a la subunidad S1 de la proteína S, lo que ayuda a la entrada por parte del virus a la célula del huésped.⁴⁴ Debido a lo anterior se sugiere que la interacción entre la proteína S del SARS-CoV-2 y las moléculas CD26 y CD147 puede conllevar a una infección por células T no productivas y resulta en la muerte celular inducida por activación.³⁰

1.7 Epidemiología

Hay evidencia de que toda la población es susceptible a la enfermedad de COVID-19. La severidad de los casos se encuentra correlacionada de manera positiva con factores como la edad y algunas enfermedades preexistentes como la hipertensión, diabetes, afecciones cardiovasculares, obesidad, etc.^{45,46,47}

Se estima que entre el 56% y 86% de personas infectadas por SARS-CoV-2 desarrollan síntomas.⁴⁸ Así mismo, la enfermedad es capaz de progresar rápidamente al SDRA; siendo esta la principal causa de muerte de los pacientes infectados por SARS-CoV-2.⁴⁵

A pesar de la alta incidencia de infecciones respiratorias en niños causadas por coronavirus comunes, la incidencia por el SARS-CoV-2 es baja (< 5%). A la fecha se desconocen las posibles razones de esto. Se ha encontrado que causas ocasionadas por la enfermedad como los cuidados hospitalarios, la severidad y la mortalidad son menores en niños comparado contra los adultos. De igual manera

los hombres presentan una mayor incidencia (60%) de contraer la infección sobre las mujeres.⁴⁹

1.8 Infección

La infección causada por el SARS-CoV-2 se divide en tres etapas, mostradas en forma esquemática en la Figura 5.⁴⁹

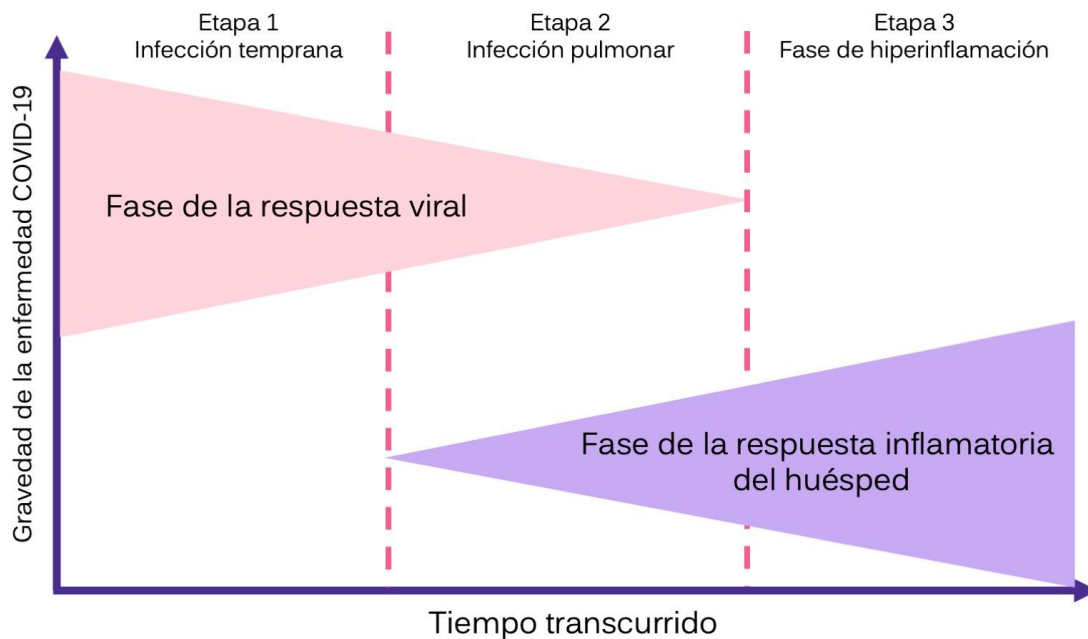


Figura 5. Principales etapas del COVID-19 causadas por la infección del coronavirus tipo 2 del Síndrome Respiratorio Agudo Grave.

La primera etapa (leve o infección temprana), involucra el periodo de incubación. En esta etapa el SARS-CoV-2 entra al hospedero a través de los receptores ECA2 y se establece. En esta etapa los síntomas suelen ser leves e inespecíficos y se desarrolla en el sistema respiratorio, principalmente en los pulmones y el epitelio vascular.⁵⁰

En la segunda etapa (pulmonar o moderada), la enfermedad se encuentra establecida en los pulmones y el virus es capaz de multiplicarse y llevar a cabo su ciclo celular. Los pacientes pueden desarrollar neumonía viral con tos y fiebre y en algunos casos, hipoxia. En esta fase la mayoría de los pacientes necesitan ser hospitalizados.⁵¹

La tercera etapa (grave), se manifiesta como el síndrome de hiperinflamación sistémica extrapulmonar. Hay disminución significativa de las células T cooperadoras, supresoras y reguladoras; y la afectación sistémica de órganos.⁵¹

1.8.1 Signos y síntomas clínicos

La infección por SARS-CoV-2 puede ser asintomática, leve, moderada, grave o crítica. Se estima que la infección tiene un periodo de incubación promedio de 5.1 a 6.4 días, y el virus tiene un periodo de replicación entre 2.2 y 3.6 días. Así mismo, la mayoría de los pacientes desarrollan síntomas a los 11.5 días.⁵²

Los pacientes enfermos de COVID-19 pueden presentar una gran variedad de signos y síntomas, que pueden ser de leves a graves. Entre algunos de los signos y síntomas se encuentran la tos, el dolor de garganta, diarrea, dolor de cabeza, mialgia, artralgia, fatiga, pérdida del olfato y del gusto, dificultad para respirar, pérdida del apetito, confusión, dolor o sensación de presión en el pecho, temperatura corporal alta (>38°C) o baja, hipertensión o hipotensión, producción de esputo, disnea, hipoxia, hemoptisis, escalofríos, resfriado, náuseas, vómito y taquicardia.^{53,54,55}

Debido a que la infección grave ocasionada por el SARS-CoV-2 es poco frecuente que se presente en los niños y jóvenes, los niños y adolescentes suelen presentar síntomas leves. Sin embargo, los casos reportados en menores de 1 año se han asociado con altos índices de morbilidad y mortalidad. Con base en los patrones de mortalidad, además de los adultos; los niños y adolescentes con afecciones preexistentes como enfermedad pulmonar obstructiva crónica (EPOC), asma, trastornos cardiovasculares y antecedentes de tabaquismo y/o vapeo tienen un mayor riesgo de presentar una infección grave.. La distribución promedio de los casos de COVID-19 por edades en el mundo se muestra en la Figura 6.^{52,56}

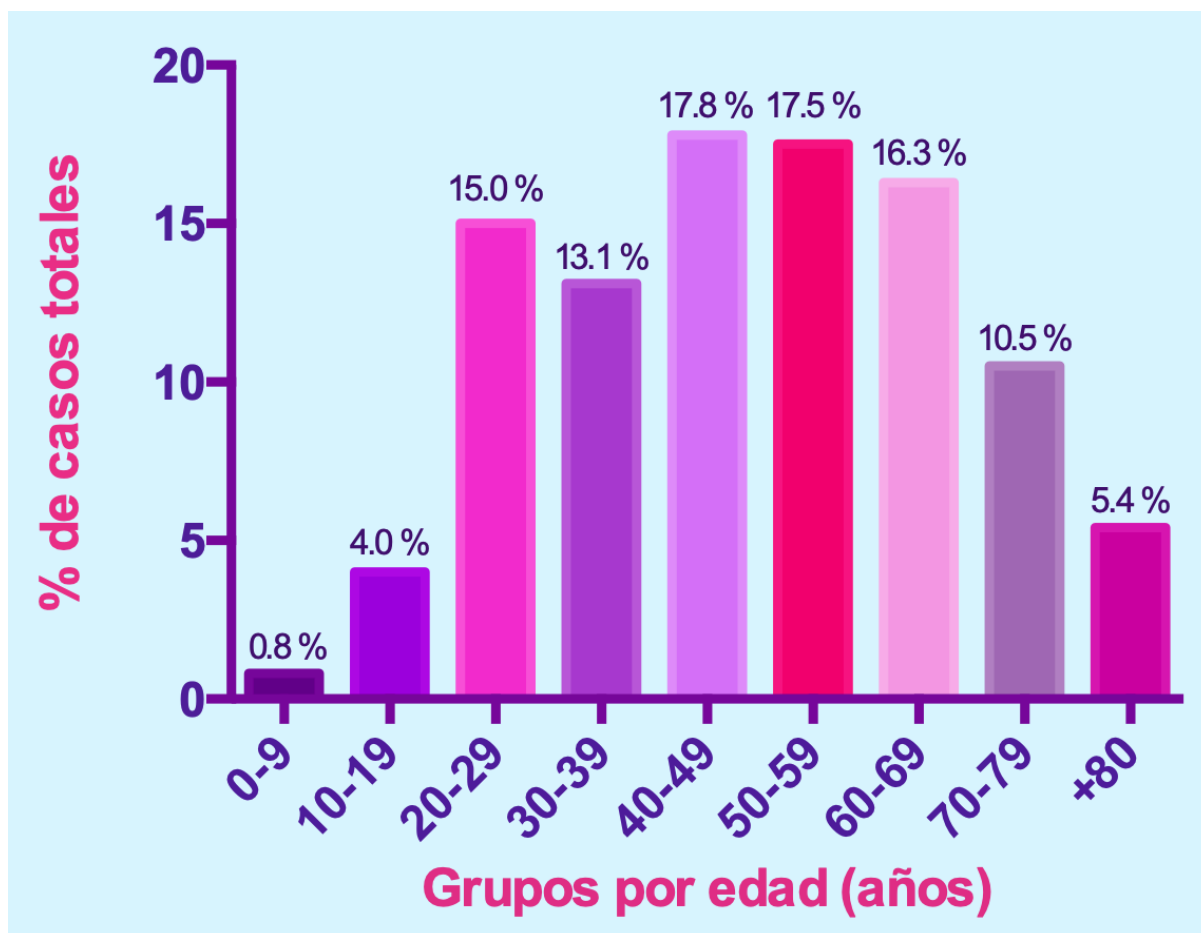


Figura 6. Distribución promedio de los casos por edad de COVID-19 en el mundo.

1.8.2 Diagnóstico

Actualmente, las técnicas empleadas para el diagnóstico de COVID-19 ocasionada por SARS-CoV-2 incluyen pruebas moleculares y pruebas rápidas.⁵⁷ En las pruebas moleculares se utiliza la reacción de la cadena de la polimerasa empleando una transcriptasa inversa (RT-PCR, por sus siglas en inglés) en tiempo real; esta prueba es la utilizada como estándar para confirmación de la enfermedad de COVID-19.⁵⁸ Las pruebas rápidas, también conocidas como pruebas serológicas, son una alternativa ampliamente utilizada debido a la creciente dispersión de la pandemia. Estas se basan principalmente en identificar anticuerpos del tipo IgM e IgG producidos por los pacientes como respuesta a la infección por SARS-CoV-2.^{59,60} Sin embargo, la OMS no recomienda el uso de pruebas rápidas para la toma de decisiones clínicas debido a la baja especificidad y sensibilidad que estas tienen.^{61,62} Así mismo, el diagnóstico se puede acompañar por una tomografía

computarizada de tórax, en las cuales se puede observar los daños a nivel pulmonar.⁶³

1.9 Tratamientos

Considerando la gravedad de la pandemia actual, diversos grupos de investigación se encuentran desarrollando tratamientos potenciales a través del descubrimiento de diversos agentes terapéuticos en conjunto con el uso de potenciales vacunas. Algunas de las principales estrategias para realizar la búsqueda son el reposicionamiento de fármacos a partir de moléculas eficaces contra otros coronavirus como el SARS-CoV y el MERS-CoV, el uso de anticuerpos, citocinas y terapias génicas.^{64,65,66,67} En las siguientes secciones se describe brevemente cada una de ellas.

1.9.1 Búsqueda de moléculas pequeñas como antivirales

Para la búsqueda y desarrollo de compuestos que puedan actuar como antivirales específicos contra el SARS-CoV-2, es importante considerar la diana biológica de interés. Algunas proteínas clave son la 3CLpro, la PLpro, la RdRp, la proteína S, el TMPRSS2 y la ECA2.^{13,64,68,69,70}

1.9.2 Reposicionamiento de fármacos

A pesar de que al momento no se han identificado moléculas con potencial antiviral contra el SARS-CoV-2, una estrategia relativamente económica y rápida es el reposicionamiento de fármacos.⁶⁴ Esta estrategia es menos costosa y más rápida que el desarrollo típico de fármacos. Algunos de los fármacos propuestos al momento son el baricitinib,⁷¹ el lopinavir,⁶⁸ el ritonavir,^{68,72} el favipiravir,^{73,74} el remdesivir,^{68,74,75} el arbidol,⁷⁶ la cloroquina,^{74,75} y la nitazoxanida.^{64,77}

1.9.3 Uso de anticuerpos

Debido a que la infección causada por el SARS-CoV-2 es capaz de inducir respuesta inmunológica del tipo humoral y celular, los pacientes pueden presentar inmunidad posterior a la infección. Así mismo, se ha detectado que los niveles de

anticuerpos, como interleucinas 2, 7 y 10 (IL-2, IL-7 e IL-10, por sus siglas en inglés), el factor estimulante de colonias de granulocitos (G-CSF, por sus siglas en inglés), proteína quimioatrayente de quimocitos 1 (MCP-1, por sus siglas en inglés), la proteína inflamatoria de macrófagos 1a (MIP1A, por sus siglas en inglés) y el factor de necrosis tumoral alfa (TNF- α , por sus siglas en inglés) son más altos en pacientes que presentan la enfermedad grave.⁷⁸ Esto sugiere que el plasma de convalecencia de pacientes recuperados de COVID-19, el cual contiene anticuerpos específicos contra el virus, puede ser útil para tratar la infección por SARS-CoV-2 en pacientes que se encuentran en las primeras etapas de la enfermedad.^{79,80,81}

1.9.4 Empleo de citocinas

La producción de diversas citocinas específicas como quimiocinas, IFNs, interleucinas y linfocinas son parte de la respuesta inmune a la invasión de patógenos, como es el caso de los virus. Estas citocinas son capaces de interferir con la replicación viral. Por esto, se ha propuesto usar algunas citocinas patentadas para tratar la infección por SARS-CoV como una estrategia para el tratamiento de la infección ocasionada por el SARS-CoV-2.^{64,82}

1.9.5 Vacunas

El desarrollo de vacunas eficaces, seguras e inmunogénicas es indispensable para disminuir la propagación y la recurrencia en el futuro ocasionada por el virus SARS-CoV-2. Las vacunas contra agentes virales pueden ser de virus atenuados, de partículas similares a virus (VLP, por sus siglas en inglés), de vectores virales, basadas en proteínas, ADN o ácido ribonucleico mensajero (ARNm).^{64,83}

2. Objetivos

2.1 General

Encontrar posibles inhibidores de la proteasa principal del SARS-CoV-2 mediante el filtrado sistemático de una amplia base de datos molecular, utilizando herramientas quimioinformáticas y de modelado molecular.

2.2 Particulares

- Generar y curar una base de datos molecular extensa (> 17 millones de compuestos, *MEGA Db*).
- Filtrar la *MEGA Db* con cálculos de similitud molecular.
- Realizar análisis computacionales adicionales a los *hits* seleccionados.
- Evaluar la actividad inhibitoria enzimática de los *hits* computacionales.

3. Hipótesis

Bajo la premisa de que es probable que “estructuras químicas similares pueden tener actividad biológica similar” se espera que sean candidatos como compuestos antivirales para la enfermedad de COVID-19 compuestos similares a moléculas con actividad conocida contra coronavirus homólogos al SARS-CoV-2.

4. Métodos

La Figura 7 resume las principales etapas para realizar este trabajo.

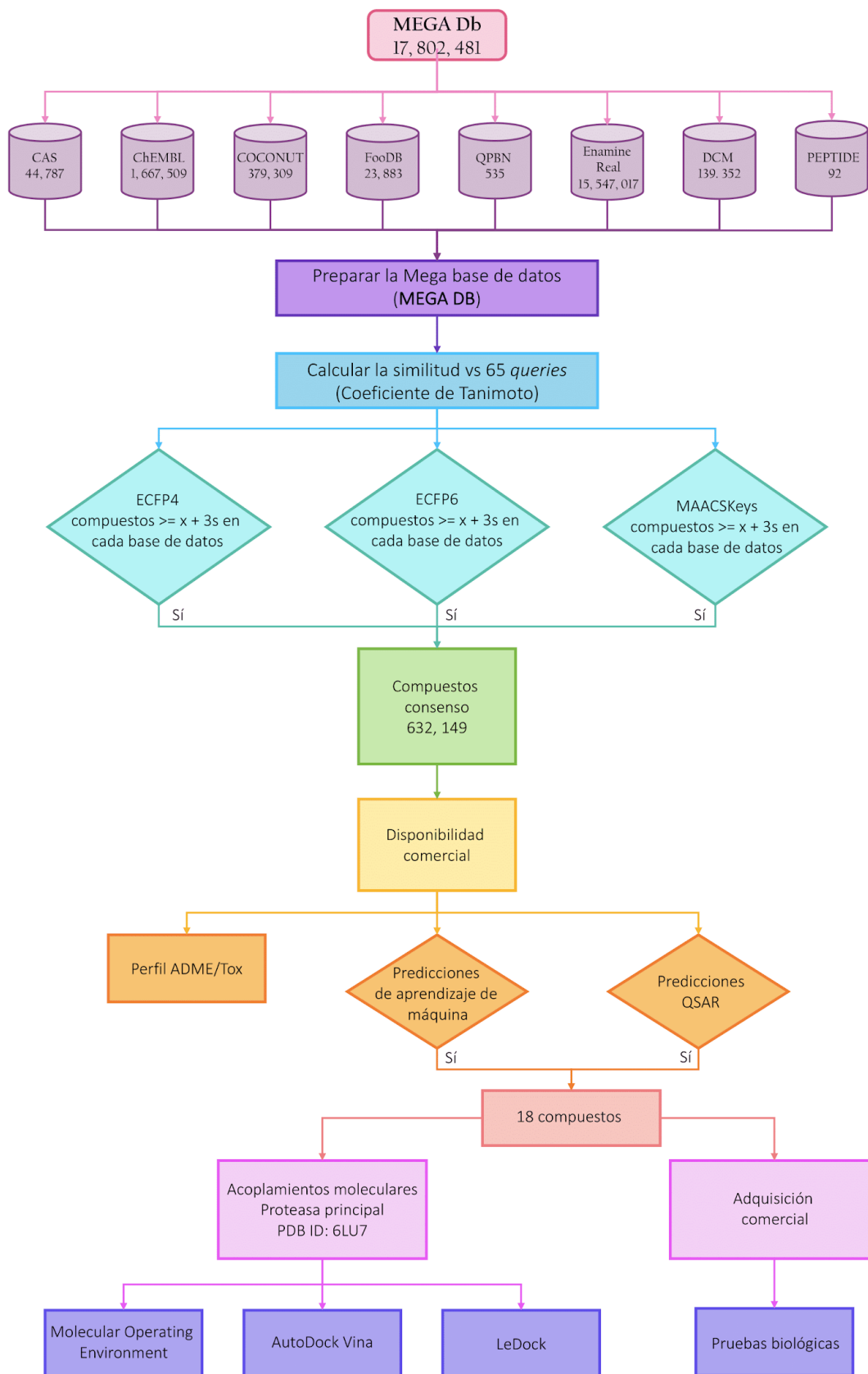


Figura 7. Diagrama de flujo del cribado virtual empleado en el presente trabajo.

4.1 Generar y preparar una amplia base de datos molecular (*MEGA Db*)

Como primer paso se llevó a cabo la elaboración de una base de datos molecular extensa o quimioteca denominada *MEGA Db*. Esta quimioteca se generó a partir de ocho bibliotecas moleculares que contienen compuestos con alta diversidad molecular. Entre estos se encuentran principalmente compuestos bioactivos, productos naturales y péptidos. La Tabla 1 resume la información de las ocho bibliotecas moleculares empleadas para la generación de la *MEGA Db*, la cual contiene un total de 17,802,401 moléculas.

Tabla 1. Compuestos utilizados para hacer la selección de compuestos asistida por computadora.

Biblioteca	Descripción	Tamaño	Referencia
CAS	Biblioteca libre enfocada hacia el COVID-19 publicada por la American Chemical Society	44,787	84
ChEMBL	ChEMBL, versión 25	1,667,509	85
COCONUT	Productos naturales obtenidos de la base de datos pública COLleCtion of Open NatUral producTs (COCONUT)	379,309	86
FooDB	Compuestos dietéticos obtenidos de la base de datos pública FooDB	23,883	87
QPNB	Productos naturales obtenidos de la base de datos Colombiana QuiProNaB	535	ND
Enamine	Compuestos de la base de datos REadily AccessibLe (REAL) - Enamine	15,547,017	88
DCM	Compuestos obtenidos de la base de datos Dark Chemical Matter	139,352	89
FDA PEP	Conjunto de péptidos aprobados por la FDA	92	90
TOTAL		17,802,481	

Todos los compuestos fueron curados utilizando un protocolo desarrollado previamente en el grupo de investigación.⁹¹ El protocolo consiste en la estandarización de las moléculas químicas a partir de su respectiva notación lineal SMILES. En este proceso, se conserva el fragmento más largo de cada compuesto y se descartan las moléculas que contienen elementos diferentes a H, B, C, N, O, F, Si, P, S, Cl, Se, Br y I, y que tienen errores en su valencia. Así mismo, se eliminan las moléculas duplicadas en cada base de datos. Finalmente, los compuestos restantes se neutralizan y reionizan para generar el tautómero correspondiente. Dicho protocolo fue realizado a partir de herramientas públicas y una interfaz de programación de aplicaciones (API, por sus siglas en inglés) en python empleando la paquetería de RDKit,⁹² versión 2019.09.1 y las funciones “*Standardizer*”, “*LargestFragmentChoser*”, “*Uncharger*”, “*Reionizer*” y “*TautomerCanonicalizer*” implementadas en la sección de herramientas de estandarización y validación de moléculas MolVS.⁹³

4.2 Crear y curar la lista de moléculas como referencias (*queries*)

Se creó una lista de 65 moléculas de referencia, denominadas como *queries* o plantillas de búsqueda. La lista se generó a partir de moléculas reportadas en la literatura con actividad antiviral conocida contra diferentes dianas biológicas de otros coronavirus. La hipótesis es que las moléculas de la Tabla 1 con estructura semejante a las moléculas de referencia, tengan actividad similar; en específico que sean activas contra el virus SARS-CoV-2.

La Tabla 2 resume la información de las 65 moléculas de referencia empleadas. Ejemplos de los compuestos de referencia son inhibidores del oncogen homólogo 1 del abensol de leucemia murina (ABL1), el cual se encuentra involucrado en la vía de replicación de coronavirus como el SARS-CoV y MERS-CoV;⁹⁴ antagonistas del receptor de estrógenos 1 (ESR1), lo cual evita el transporte de vesículas virales al endosoma tardío y por lo tanto también se afecta la fusión viral;⁹⁵ inhibidores de endocitosis mediada por clatrina; antagonistas del receptor H₁, los cuales en algunos casos inhiben la replicación viral; inhibidores de canales iónicos; inhibidores de ARN polimerasas dependientes de ARN (análogos de nucleótidos);⁹⁶ agonistas de receptores opioides; inhibidores de la Mpro del SARS-CoV; inhibidores de la catepsina; inhibidores del metabolismo de los

esteroles; inhibidores del virus de la inmunodeficiencia humana (VIH) e inhibidores del metabolismo del ADN.^{97,98}

Tabla 2. Compuestos de referencia utilizados para la selección de compuestos asistida por computadora. Coronavirus tipo 1 del Síndrome Respiratorio Agudo Grave (SARS-CoV), Coronavirus tipo 2 del Síndrome Respiratorio Agudo Grave (SARS-CoV-2), Síndrome Respiratorio del Oriente Medio (MERS-CoV), Coronavirus Humano OC43 (HCoV-OC43), Coronavirus Humano NL63 HCoV-NL63, Coronavirus Murino del Virus de la Hepatitis de ratón A-59 MHV-A59. *Inhibición indirecta a través de la proteína AAK1.

Compuesto	Coronavirus sobre el que tiene actividad antiviral reportada	Referencia
Remdesivir	MERS-CoV	96
Ivermectina	SARS-CoV-2	98
Clorhidrato de Emetina	MERS-CoV, SARS-CoV	94,97
Fosfato de cloroquina	MERS-CoV, SARS-CoV	97
Sulfato de hidroxicloroquina	MERS-CoV, SARS-CoV	97
Mefloquina	MERS-CoV, SARS-CoV	94,97
Clorhidrato de amodiaquina	MERS-CoV, SARS-CoV	94,97
Aloxistatina	MERS-CoV, SARS-CoV	94,97
Clorhidrato de gemcitabina	MERS-CoV, SARS-CoV	94,97
Citrato de tamoxifeno	MERS-CoV, SARS-CoV	94,97
Citrato de toremifeno	MERS-CoV, SARS-CoV	94,97
Terconazol	MERS-CoV, SARS-CoV	97
Triparanol	MERS-CoV, SARS-CoV	97

Anisomicina	MERS-CoV, SARS-CoV	97
Cicloheximida	MERS-CoV, SARS-CoV	97
Mepesuccinato de omacetaxina	MERS-CoV, SARS-CoV	97
Mesilato de benztropina	MERS-CoV, SARS-CoV	97
Fluspirileno	MERS-CoV, SARS-CoV	97
Tiotixeno	MERS-CoV, SARS-CoV	97
Clorhidrato de flufenazina	MERS-CoV, SARS-CoV	94,97
Clorhidrato de prometazina	MERS-CoV, SARS-CoV	94,97
Astemizol	MERS-CoV, SARS-CoV	94,97
Clorhidrato de clorfenoxamina	MERS-CoV, SARS-CoV	94,97
Clorhidrato de clorpromazina	MERS-CoV, SARS-CoV	97
Clorhidrato de trifluoperazina	MERS-CoV, SARS-CoV	94,97
Maleato de tietilperazina	MERS-CoV, SARS-CoV	97
Clorhidrato de clomipramina	MERS-CoV, SARS-CoV	94,97
Mesilato de imatinib	MERS-CoV, SARS-CoV	94,97
Dasatinib	MERS-CoV, SARS-CoV	94,97
Clorpromazina	MERS-CoV	66,94,97
Cloroquina	MERS-CoV	66,94,97
Toremifeno	MERS-CoV	66
Baicalin	SARS-CoV	99
Glicirricina	SARS-CoV	99

Ácido clorogénico	SARS-CoV	99
Niclosamida	SARS-CoV	67
Cinanserin	SARS-CoV	67
Valinomicina	SARS-CoV	67
Citarabina	SARS-CoV	67
Ribavirina	SARS-CoV	67
Loperamida	MERS-CoV	94,97
Lopinavir	MERS-CoV	94,97
Nutlin	MERS-CoV	94
Hidroxizina	MERS-CoV	94
Saracatinib	MERS-CoV	94
Sotrastaurina	MERS-CoV	94
Acetofenazina	MERS-CoV	94
Dosulepin	MERS-CoV	94
Metotrimeprazina	MERS-CoV	94
Licorina	HCoV-OC43, HCoV-NL63, MERS-CoV, MHV-A59	94
Emetina	HCoV-OC43, HCoV-NL63, MERS-CoV, MHV-A59	94
Micofenolato de mofetilo	HCoV-OC43, HCoV-NL63, MERS-CoV, MHV-A59	94
Fenazopiridina	HCoV-OC43, HCoV-NL63, MERS-CoV, MHV-A59	94
Ácido micofenólico	HCoV-OC43, HCoV-NL63, MERS-CoV, MHV-A59	94

Pamoato de pirvinio	HCoV-OC43, HCoV-NL63, MERS-CoV, MHV-A59	94
Monensina de sodio	HCoV-OC43, HCoV-NL63, MERS-CoV, MHV-A59	94
6-mercaptopurina	MERS-CoV	94
6-tioguanina	MERS-CoV	94
N-metilmaleimida	MERS-CoV	94
Disulfiram	MERS-CoV	94
Baricitinib*	SARS-CoV-2	71
Favipiravir	SARS-CoV-2	64
Darunavir	SARS-CoV-2	64
Arbidol	SARS-CoV-2	64
Nitazoxanida	SARS-CoV-2	64

Las moléculas de referencia de la Tabla 2 fueron preparadas con el mismo protocolo empleado para la preparación de la *MEGA Db* descrita en la sección 4.1.

4.3 Filtrar la base de datos con análisis de similitud molecular vs. las moléculas de referencia

Los valores de similitud molecular entre cada una de las moléculas en la quimioteca *Mega DB* y los compuestos de referencia se calcularon utilizando tres huellas digitales moleculares en dos dimensiones: *Extended Connectivity Fingerprints*, 2048 bits (ECFP4 y ECFP6) y *Molecular ACCess System Keys* (MACCSKeys), 166 bits. Para cuantificar la similitud se empleó el coeficiente de Tanimoto.¹⁰⁰ Los cálculos para el análisis de similitud molecular se realizaron a partir de una API desarrollada en python, utilizando los módulos “*Fingerprinting and Molecular Similarity*”, “*Topological Fingerprints*”, “*MACCS Keys*” y “*Morgan*

Fingerprints (Circular Fingerprints)” de la paquetería de RDKit⁹² y la supercomputadora Miztli de la UNAM. En total se realizaron 3,471,483,795 comparaciones, resultantes del análisis de los 17,802,401 compuestos de la *MEGA Db* contra las 65 moléculas de referencia empleando las tres diferentes huellas digitales moleculares. Los cálculos se hicieron en un tiempo aproximado de 40 horas.

Posteriormente, se obtuvo la lista de compuestos como posibles candidatos computacionales. Los compuestos seleccionados fueron aquellos que tenían un valor de similitud mayor o igual al promedio más tres desviaciones estándar de la distribución para cada biblioteca y huella digital molecular con cada una de las 65 moléculas de referencia.

4.4 Análisis consenso de los resultados de similitud molecular

Para cada molécula de referencia se obtuvo una lista de compuestos conformada por las moléculas que se identificaron como candidatos computacionales con los tres descriptores moleculares (huellas digitales): ECFP4, ECFP6 y MACCSKeys. De esta manera, se obtuvo un conjunto denominado “compuestos consenso”, el cual incluyó aquellas moléculas no repetidas de la *MEGA Db*.

4.5 Búsqueda de disponibilidad comercial

La disponibilidad comercial de las moléculas de la lista de compuestos consenso se verificó en la base de datos de ZINC (ZINC 15)^{101,102} utilizando las palabras clave “For Sale”, “In Stock” y “Now”.

4.6 Determinar el perfil ADME/Tox de las moléculas disponibles comercialmente

Las propiedades para describir el perfil de administración, distribución, metabolismo, excreción y toxicidad (ADME/Tox) de los compuestos de la lista de

compuestos disponibles comercialmente se calcularon utilizando dos servidores en línea de acceso libre: *SwissADME*¹⁰³ y *pkCSM-pharmacokinetics*.¹⁰⁴

Para obtener el perfil ADME/Tox de dichos compuestos se seleccionaron 18 descriptores de absorción, distribución, metabolismo, eliminación y toxicidad; los cuales se encuentran descritos en la Tabla 3.

Tabla 3. Descriptores moleculares empleados para determinar el perfil ADME/Tox de los compuestos disponibles comercialmente.

Descriptor molecular	Tipo de resultado	Etapas que describe	Servidor
LogS (<i>Silicos IT LogSw</i>)	Numérico	Absorción	<i>SwissADME</i>
Lipofilicidad (<i>Consensus LogP</i>)	Numérico	Absorción	<i>SwissADME</i>
Absorción intestinal humana (<i>HIA</i>)	Numérico	Absorción	<i>pkCSM</i>
Violaciones a las reglas de Lipinski	Número de violaciones a la regla	Absorción	<i>SwissADME</i>
Violaciones a las reglas de Veber	Número de violaciones a la regla	Absorción	<i>SwissADME</i>
Permeabilidad de la barrera hematoencefálica (<i>BBB permeability</i>)	Numérico	Distribución	<i>pkCSM</i>
Sustrato de la glicoproteína P	Categorico (sí o no)	Distribución	<i>SwissADME</i>
Inhibición de la glicoproteína P-I	Categorico (sí o no)	Distribución	<i>pkCSM</i>
Inhibición de la glicoproteína P-II	Categorico (sí o no)	Distribución	<i>pkCSM</i>
Inhibición del citocromo 1A2 (CYP1A2)	Categorico (sí o no)	Metabolismo	<i>SwissADME</i>
Inhibición del citocromo 2C19 (CYP2C19)	Categorico (sí o no)	Metabolismo	<i>SwissADME</i>
Inhibición del citocromo 2C9 (CYP2C9)	Categorico (sí o no)	Metabolismo	<i>SwissADME</i>
Inhibición del citocromo	Categorico (sí o no)	Metabolismo	<i>SwissADME</i>

2D6 (CYP2D6)			
Inhibición del citocromo 3A4 (CYP3A4)	Categorico (sí o no)	Metabolismo	<i>SwissADME</i>
Eliminación total	Numérico	Eliminación	<i>pkCSM</i>
Prueba <i>in silico</i> del ensayo de AMES	Categorico (sí o no)	Toxicidad	<i>pkCSM</i>
Hepatotoxicidad	Categorico (sí o no)	Toxicidad	<i>pkCSM</i>

4.6 Realizar predicciones de aprendizaje de máquina y QSAR

Las predicciones de aprendizaje de máquina se llevaron a cabo en colaboración con el grupo *Collaborations Pharmaceuticals, Inc.*, dirigido por el Dr. Sean Ekins. Empleando algoritmos adicionales al algoritmo Assay Central™, desarrollados por el mismo grupo.¹⁰⁵

El algoritmo reportado en Assay Central™ utiliza modelos de aprendizaje de máquina Bayesianos a partir de resultados obtenidos con el descriptor molecular ECFP6 para así asignar una puntuación de predictibilidad y aplicabilidad. Algunos de los algoritmos adicionales de aprendizaje de máquina implementados son clasificadores de *Bernoulli Naive Bayes*, de *AdaBoost* (por su nombre en inglés, *Adaptive Boosting*), de bosques aleatorios (*random forest*), entre otros. Las moléculas con resultados mayores a 0.5 para estas predicciones son consideradas como activas.¹⁰⁶

Por su parte, las predicciones cuantitativas de relaciones estructura-actividad (QSAR) se realizaron en colaboración con el Dr. Vinicius Alves y el Dr. Alexander Tropsha de la Universidad de Carolina del Norte en Chapel Hill, Estados Unidos.

El algoritmo QSAR se basa en el uso de clasificadores de bosques aleatorios y tres tipos de descriptores: Huellas digitales de Morgan, SiRMS (por su nombre en inglés, *2D Simplex Representation of Molecular Structure*) y la versión 7.0 de *Dragon* (un programa computacional diseñado para el código de descriptores desarrollado por *Kode Chemoinformatics srl*). Este modelo contempla diversas características de los compuestos y la actividad inhibitoria contra la Mpro de coronavirus que comparten homología en el sitio catalítico de esta proteína (> 96%). Algunas características consideradas son la lipofilicidad, cargas parciales, índice de

refracción, y la capacidad de que puedan formar puentes de hidrógeno (ya sean de tipo aceptores o donadores).¹⁰⁷

Los resultados del modelo QSAR se evalúan a partir de un análisis consenso promedio de los tres descriptores a partir de los resultados obtenidos en cada una de las predicciones llevadas a cabo. La actividad de cada predicción se escala y evalúa del cero al uno, donde los valores más cercanos a uno tienen una predicción de actividad más alta. Aquellos compuestos con un resultado consenso mayor a 0.7 (70%) se consideran como positivos. Los resultados de este modelo son de tipo categórico donde el 0 es un resultado negativo y el 1 es un resultado positivo.¹⁰⁷

Los compuestos con resultados positivos para ambas predicciones se denominaron *hits computacionales*.

4.7 Acoplamiento molecular con la proteasa principal de SARS-CoV-2

Se realizaron acoplamientos moleculares para determinar las posibles interacciones entre la Mpro de SARS-CoV-2 y los *hits computacionales*. Los acoplamientos moleculares se realizaron empleando tres diferentes programas: *Molecular Operating Environment* (MOE) versión 2018,¹⁰⁸ *LeDock*¹⁰⁹ y *AutoDock Vina* versión 1.1.2.¹¹⁰ La proteína utilizada fue la estructura cristalográfica de la Mpro del SARS-CoV cocristalizada con el inhibidor con código del fabricante MPI1 (PDB ID: 7JPZ).¹¹¹ Antes de realizar el acoplamiento de los compuestos en el sitio catalítico se quitó la estructura de inhibidor co-cristalizado.

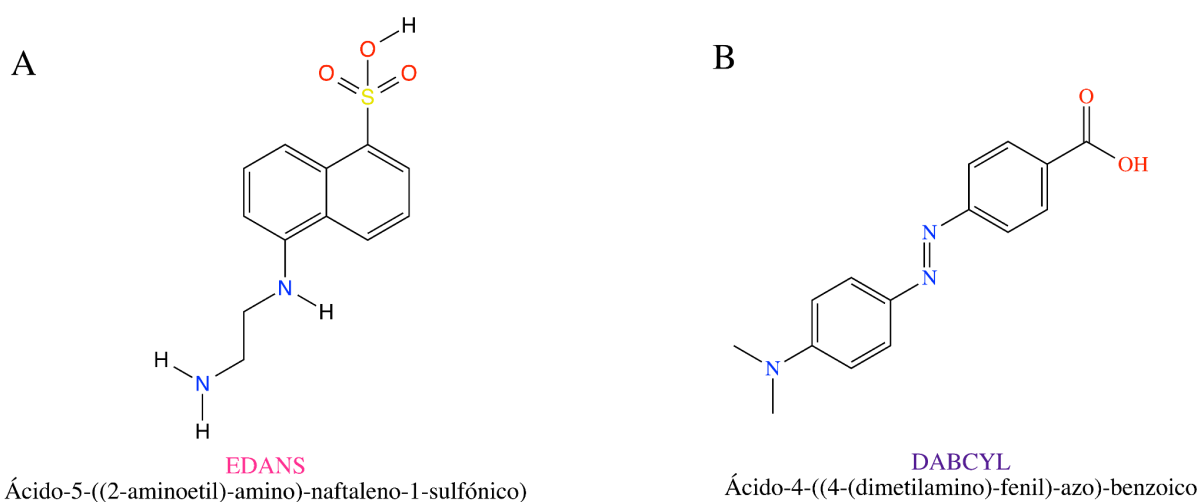
Para realizar los acoplamientos moleculares en MOE y *LeDock* se preparó la proteína en MOE utilizando el módulo “*Quick prepare*” con los parámetros predeterminados y el campo de fuerza AMBER10:EHT. Por otra parte, la preparación de la proteína para utilizar en *AutoDock Vina* se llevó a cabo en *AutoDock Tools* versión 1.5.2,¹¹² y la construcción de la caja (*grid*) se realizó en el sitio de unión del inhibidor MPI1. Las dimensiones de la caja utilizadas fueron de 60 unidades por eje (22.5 Angstroms). Las coordenadas del centro de la caja fueron:

-16.585 1.792 -1.724

La construcción de la estructura de los ligandos se hizo en MOE a partir de los canonical SMILES de los *hits computacionales* y las funciones “*wash*”, “*partial charges*” y “*energy minimize*”. Los acoplamientos moleculares realizados en MOE se hicieron con el método “*Triangle Matcher*” y la función “*London dG*.”

4.8 Adquisición comercial de compuestos y envío a evaluación biológica

Los *hits computacionales* fueron adquiridos comercialmente a la empresa MolPort para realizar pruebas enzimáticas *in vitro* con la Mpro del SARS-CoV-2 en los laboratorios Reaction Biology Corp. (Malvern, Pensilvania, EUA) para evaluar la actividad inhibitoria. Esta prueba consiste en detectar cuantitativamente la actividad enzimática de la Mpro del SARS-CoV-2 a partir de un ensayo de actividad de proteasa basado en fluorescencia, la cual permite conocer la actividad proteolítica de la proteasa a partir de la escisión de sustratos de péptidos fluorogénicos. Para el caso de la Mpro, el ensayo utiliza Ditioneitol (DTT) como agente reductor, el compuesto con código **GC376** denominado por la biofarmacéutica Anivive¹¹³ utilizado como control y el sustrato: [NH₂-C(EDANS)VNSTQSGLRK(DABCYL)M-COOH] mostrado en la Figura 8.^{114,115}



C

Cys-(EDANS)-Val-Asn-Ser-Thr-Gln-Ser-Gly-Leu-Arg-Lys-(DABCYL)-Met

Figura 8. A) Estructura del ácido-5-((2-aminoetil)-amini)-naftaleno-1-sulfónico, también conocido con el nombre de EDANS, B) Estructura del ácido-4-((4-(dimetilamino)-fenil)-azo)-benzoico, también conocido con el nombre de DABCYL, C) Secuencia de la estructura del sustrato [NH₂-C(EDANS)VNSTQSGLRK(DABCYL)M-COOH] empleado en los ensayos fluorogénicos de la proteasa principal del coronavirus tipo 2 del Síndrome Respiratorio Agudo Grave .

5. Resultados y discusión

5.1 Generación y preparación de la *MEGA Db*

La Tabla 1 resume el contenido de *MEGA Db* (ver sección 4.1). La *MEGA Db* contiene 17,802,481 compuestos en total, entre los cuales contiene moléculas de diferentes orígenes y de relevancia en la Química Farmacéutica para el descubrimiento de fármacos. Esta incluye productos naturales, fármacos aprobados para uso clínico, moléculas de origen sintético y péptidos. La colección de compuestos que tiene la mayor proporción de moléculas es la biblioteca Enamine (moléculas “bajo demanda”) con 15,547,017 (87.33%), seguida de la biblioteca ChEMBL con 1,667,509 (9.37%).

5.2 Lista de moléculas de referencia (*queries*)

Las 65 moléculas de referencia empleadas para el cálculo de similitud se encuentran en la Tabla 2 (ver sección 4.2). Estas se seleccionaron tomando como criterio el que tuvieran como principal característica actividad biológica conocida contra diferentes blancos moleculares de coronavirus homólogos al SARS-CoV-2. En esta lista se encuentran principalmente compuestos derivados de origen natural y de origen sintético. En la Figura 9 se muestran los núcleos base (*scaffolds*) más frecuentes de las moléculas de referencia. Los seis núcleos base moleculares mostrados en la figura contienen el 27.7 % del total de los 65 compuestos que conforman la lista de moléculas de referencia.

Los núcleos base se calcularon empleando la definición de Bemis y Murcko.¹¹⁶ En esta definición, el núcleo base consiste en obtener los anillos centrales y enlaces que unen a estos, a partir de la eliminación de las cadenas laterales de los compuestos. Para la obtención de estos, se realizó un código en el lenguaje de programación de Python a partir de la nomenclatura en SMILES de las moléculas de referencia y del módulo *MurckoScaffold* de RDKit.⁹²

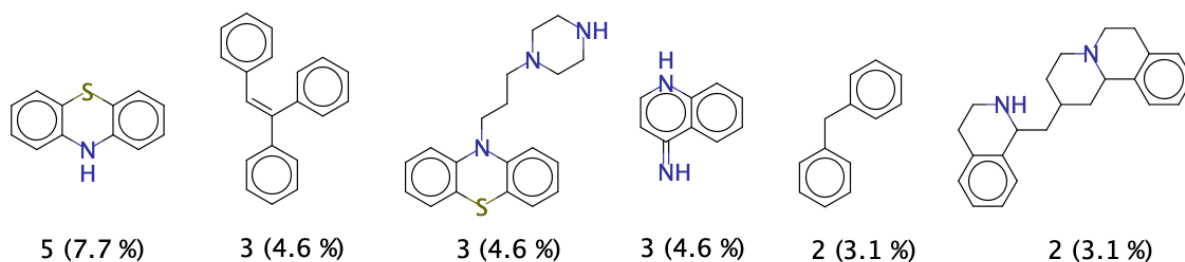


Figura 9. Núcleos base más frecuentes de las moléculas de referencia. Se muestra la frecuencia y el porcentaje que representan en la lista conformada por las 65 moléculas.

5.3 Filtrado de la base de datos con similitud molecular vs las moléculas de referencia

Las huellas digitales moleculares son una manera de representar estructuras químicas como un objeto matemático, para así poder realizar análisis estadísticos que permitan obtener predicciones de los compuestos en estudio. Por este motivo, las huellas digitales son una herramienta quimioinformática relevante para llevar a cabo la búsqueda de moléculas por medio de análisis de similitud molecular y cribados virtuales.¹¹⁷ Para este trabajo se emplearon diversas huellas digitales: ECFP4, ECFP6 (ECFP de radio 2 y radio 3) y MACCSKeys.

En el caso de ECFP,¹¹⁷ se utiliza una variación derivada del algoritmo de Morgan, en la cual consiste en tres etapas secuenciales:

- 1) Etapa de asignación inicial: a cada átomo se le asigna un número entero como identificador. Para asignar el identificador se consideran el número de átomos vecinos que no sean hidrógenos, número de enlaces unidos al átomo, número atómico, masa atómica, número de hidrógenos enlazados al átomo y si el átomo se encuentra en un anillo. Este proceso se repite con cada átomo, hasta que a todos los átomos de la molécula se les asigna un identificador.
- 2) Etapa de actualización iterativa: cada identificador asignado previamente, se actualiza para considerar los identificadores de átomos vecinos. Para llevar a cabo esta actualización se realiza la creación de una matriz que considere el nivel de iteración, el número de identificador del átomo, el orden de enlace a átomos vecinos diferentes a hidrógeno y el número de identificador del átomo

vecino. Posteriormente se aplica una función *hash* a esta matriz y se genera un nuevo identificador. La huella digital obtenida es la concatenación de los identificadores de las iteraciones. El número de iteraciones termina cuando se consideran a todos los átomos vecinos que se encuentran dentro del radio definido.

- 3) Etapa de eliminación de identificadores duplicados: apariciones múltiples de identificadores para las mismas características se reducen a un solo identificador representativo. Finalmente se crea una matriz para cada identificador de cada átomo para cada nivel de iteración, y se crea un vector de 2048 *bits* utilizando una función *hash*.

Como resultado se tiene una representación matemática para cada molécula. Por otra parte *MACCSKeys* es una huella digital conformada por un diccionario o preguntas a las cuales se le asignan como respuesta valores binarios (0 o 1). Existen dos versiones de este diccionario, uno de 960 *bits* o preguntas y el de 166 *bits*. Actualmente solo el de 166 *bits* se encuentra disponible al público en diferentes paqueterías o programas quimioinformáticos de código abierto.¹¹⁸

Las Figuras 10 a 12 muestran los mapas de calor de la distribución de los valores promedio de similitud calculados con el coeficiente de Tanimoto para cada biblioteca molecular vs. cada una de las moléculas de referencia para las tres huellas digitales empleadas.

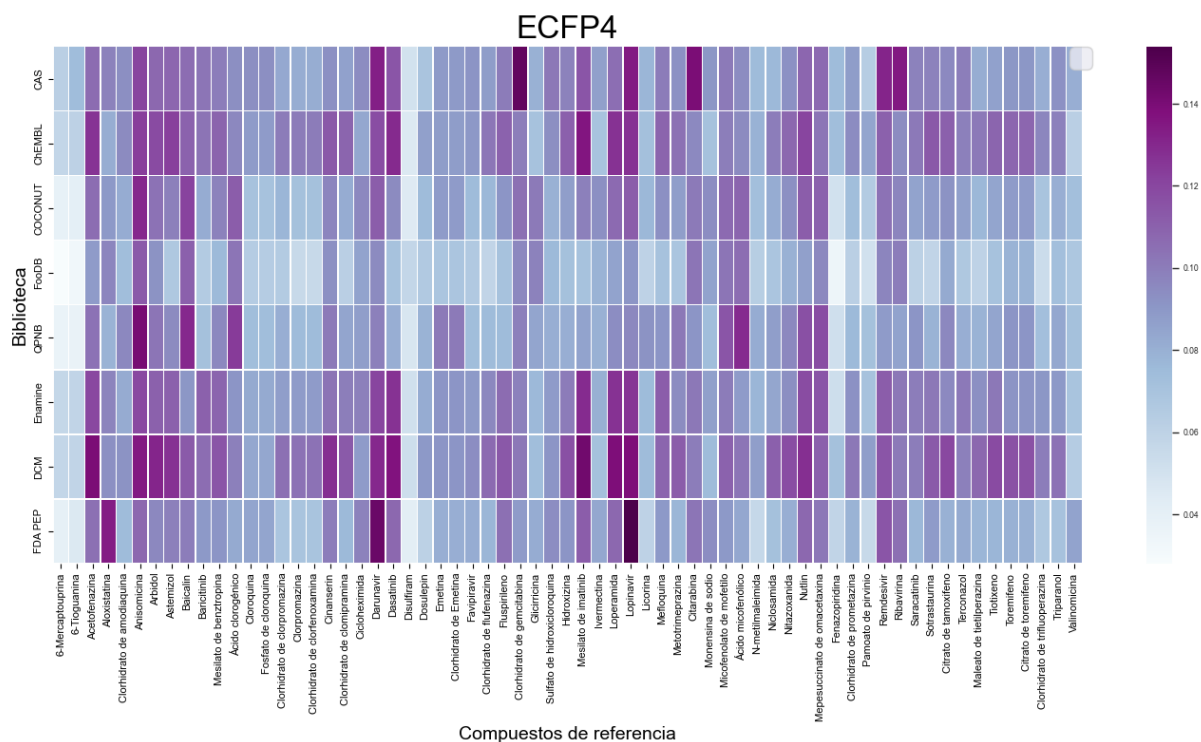


Figura 10. Promedio de los valores de similitud obtenidos para cada biblioteca molecular vs. cada molécula de referencia utilizando *Extended Connectivity Fingerprints* de radio 2 (ECFP4) como huella digital y el coeficiente de Tanimoto. Los rectángulos en color morado oscuro presentan los valores de similitud más altos, mientras que aquellos en color azul claro son aquellos con valores de similitud bajos.

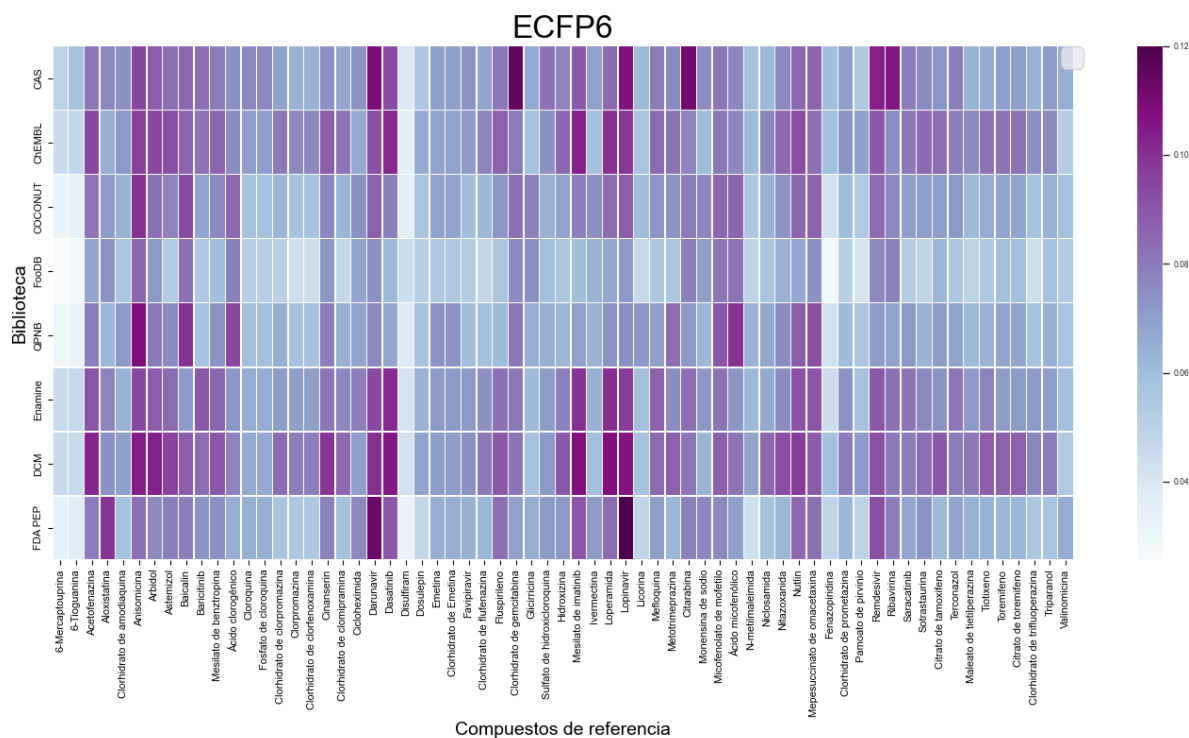


Figura 11. Promedio de los valores de similitud obtenidos para cada biblioteca molecular vs. cada molécula de referencia utilizando *Extended Connectivity Fingerprints* de radio 3 (ECFP6) como huella digital y el coeficiente de Tanimoto. Los rectángulos en color morado oscuro presentan los valores de similitud más altos, mientras que aquellos en color azul claro son aquellos con valores de similitud bajos.

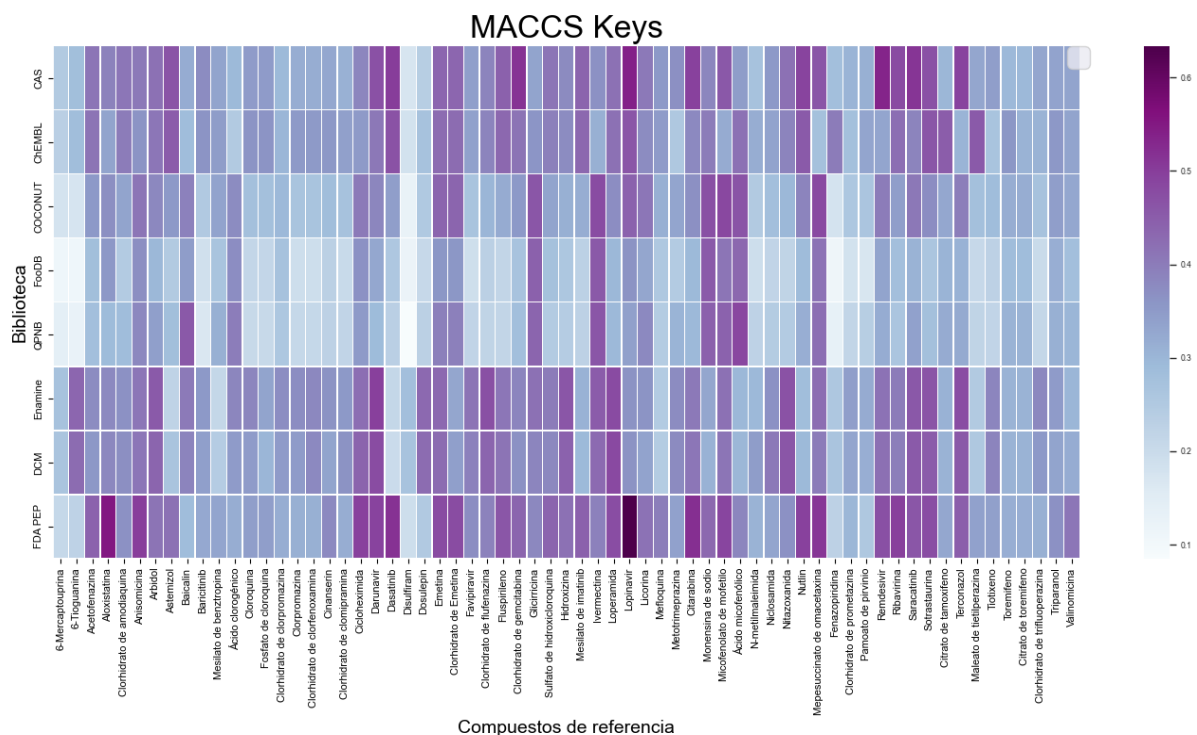


Figura 12. Promedio de los valores de similitud obtenidos para cada biblioteca molecular vs. cada molécula de referencia utilizando *Molecular ACCess System Keys* (MACCSKeys) como huella digital y el coeficiente de Tanimoto. Los rectángulos en color morado oscuro presentan los valores de similitud más altos, mientras que aquellos en color azul claro son aquellos con valores de similitud bajos.

Como se observa en la Figura 10, los valores de similitud más altos (representados con el color morado) para la huella digital ECFP4 se presentan entre el clorhidrato de gemcitabina y la biblioteca CAS (0.1476), y entre el lopinavir y la biblioteca FDA PEP (0.1540). Los valores de similitud más bajos son entre la fenazopiridina y la biblioteca FooDB (0.0350).

Para ECFP6 (Figura 11) los valores de similitud más altos son entre el clorhidrato de gemcitabina y la biblioteca CAS (0.1156) y el lopinavir y la biblioteca FDA PEP (0.1194). A su vez, los valores de similitud más bajos empleando esta huella digital son entre el disulfiram y la biblioteca FDA PEP (0.0320) y la fenazopiridina y la biblioteca FooDB (0.0288).

El valor más alto de similitud para MACCSKeys (Figura 12) es entre el lopinavir y la biblioteca FDA PEP (0.6339). Por el contrario, los valores de similitud

más bajos son entre el disulfiram (0.0853) y la fenazopiridina (0.1340) con la biblioteca QPNB y la fenazopiridina y la biblioteca FooDB (0.1140).

Con base en lo anterior, las moléculas de referencia que, en general, presentaron valores de similitud más altos, independientemente de la huella digital, fueron el clorhidrato de gemcitabina y el lopinavir. Las moléculas de referencia que presentaron valores de similitud más bajos fueron la fenazopiridina y el disulfiram.

El comportamiento general de los valores de similitud obtenidos para las ocho bibliotecas que conforman la MEGADb vs. las 65 moléculas de referencia muestra una tendencia similar entre las bibliotecas QPNB, FooDB y COCONUT. Lo cual se puede deber a que aunque las tres bibliotecas contienen compuestos diferentes, estos compuestos son productos naturales y dietéticos y a su vez es posible que presenten una similitud estructural entre estos y por esto se obtengan resultados similares. Así mismo, también se observa cierta similitud entre las bibliotecas DCM, Enamine y ChEMBL, entre las cuales puede también existir relación estructural entre los compuestos que conforman estas bibliotecas; en estas últimas bibliotecas moleculares se encuentran moléculas sintéticas, con bioactividad y con propiedades tipo fármaco (*drug-like*).

La finalidad de emplear diversas bibliotecas moleculares fue ampliar el espacio químico para así poder generar una amplia base molecular con compuestos diversos entre los que se encuentran compuestos sintéticos, tipo fármaco, fármacos aprobados, péptidos, productos naturales, dietéticos, moléculas pequeñas, etc. El objetivo de esto es el aumentar a su vez la probabilidad de encontrar moléculas que pudieran llegar a presentar mejores valores de similitud contra las moléculas de referencia.

Así mismo, es de esperarse que los rangos de los valores de similitud obtenidos para cada huella digital sean diferentes. En el caso de MACCSKeys, en general tienen una tendencia a tener valores más altos por sobre ECFP debido a que el algoritmo de esta huella digital es más simple y a su vez, se encuentra conformado por una menor cantidad de *bits*. Por lo que al ser menos específico al calcular la similitud está es “relativamente” mayor. Por otro lado, en el caso de ECFP6 el cual tiene un radio mayor (3) que el de ECFP4 (2), este toma en cuenta la consideración de un número mayor de átomos. Por esta razón, la especificidad de la huella digital generada será mayor, y por lo tanto, se espera que al calcular la similitud los valores obtenidos sean menores a los de ECFP4. Tendencias similares

sobre la magnitud de los valores de similitud cuantificados con el índice de Tanimoto y huellas digitales de diferente origen se han reportado para otras bases de datos.¹¹⁹

Con base en la distribución de los valores estadísticos obtenidos durante el cálculo de similitud para cada biblioteca vs. cada una de las 65 moléculas de referencia y cada huella digital, se obtuvieron listas de conformadas por aquellas moléculas con valores de similitud favorecidas (ver sección 4.3). La Tabla 4 resume el número de moléculas únicas en la *MEGA Db* después de seleccionar a aquellas con valores de similitud mayores al promedio más tres desviaciones estándar para cada biblioteca y molécula de referencia.

Tabla 4. Moléculas únicas con valores de similitud mayores al promedio más tres desviaciones estándar para cada molécula de referencia en la biblioteca denominada *MEGA Db* utilizando 3 diferentes descriptores moleculares: *Extended Connectivity Fingerprints* de radio 2 (ECFP4), radio 3 (ECFP6) y *Molecular ACCess System Keys* (MACCSkeys).

Molécula de referencia	Descriptor molecular		
	ECFP4	ECFP6	MACCSkeys
Remdesivir	1,621,844	4,348,212	3,009,455
Ivermectina	1,621,844	4,348,212	3,009,455
Clorhidrato de Emetina	1,621,844	4,348,173	3,009,401
Fosfato de cloroquina	1,621,844	4,348,097	3,009,351
Sulfato de hidroxiclороquina	1,621,844	4,348,212	3,009,455
Mefloquina	1,621,828	4,348,212	3,009,455
Clorhidrato de amodiaquina	1,621,844	4,348,212	3,009,455
Aloxistatina	1,621,844	4,348,645	3,009,972
Clorhidrato de gemcitabina	1,621,844	4,348,206	3,009,446
Citrato de tamoxifeno	1,621,844	4,348,212	3,009,455
Citrato de toremifeno	1,621,844	4,348,212	3,009,455
Terconazol	1,621,844	4,348,183	3,009,416
Triparanol	1,621,844	4,348,212	3,009,455

Anisomicina	1,621,844	4,348,212	3,009,455
Cicloheximida	1,621,844	4,348,212	3,009,454
Mepesuccinato de omacetaxina	1,621,844	4,348,212	3,009,455
Mesilato de benzotropina	1,621,844	4,348,113	3,009,330
Fluspirileno	1,621,844	4,348,131	3,009,370
Tiotixeno	1,621,844	4,348,198	3,009,442
Clorhidrato de flufenazina	1,621,844	4,348,203	3,009,442
Clorhidrato de prometazina	1,621,844	4,348,051	3,009,294
Astemizol	1,621,844	4,348,189	3,009,424
Clorhidrato de clorfenoxamina	1,621,844	4,348,142	3,009,352
Clorhidrato de clorpromazina	1,621,844	4,348,039	3,009,290
Clorhidrato de trifluoperazina	1,621,844	4,348,088	3,009,326
Maleato de tietilperazina	1,621,844	4,348,153	3,009,373
Clorhidrato de clomipramina	1,621,844	4,348,151	3,009,375
Mesilato de imatinib	1,621,844	4,348,212	3,009,455
Dasatinib	1,621,844	4,348,212	3,009,455
Clorpromazina	1,621,844	4,348,039	3,009,290
Cloroquina	1,621,844	4,348,097	3,009,351
Toremifeno	1,621,844	4,348,212	3,009,455
Baicalin	1,621,316	4,348,046	3,009,622
Glicirricina	1,621,843	4,348,209	3,009,450
Ácido clorogénico	1,621,630	4,348,298	3,009,765
Niclosamida	1,621,838	4,348,632	3,009,899
Cinanserin	1,621,844	4,348,211	3,009,455
Valinomicina	1,621,843	4,348,068	3,009,347
Ribavirina	1,621,844	4,348,205	3,009,444
Loperamida	1,621,844	4,348,212	3,009,455
Lopinavir	1,621,844	4,348,812	3,009,455

Nutlin	1,621,844	4,348,212	3,009,455
Hidroxizina	1,621,844	4,348,209	3,009,448
Saracatinib	1,621,844	4,348,177	3,009,418
Sotrastaurina	1,621,844	4,348,867	3,009,277
Acetofenazina	1,621,844	4,348,212	3,009,455
Dosulepin	1,621,844	4,348,170	3,009,396
Metotrimeprazina	1,621,844	4,348,122	3,009,347
Licorina	1,621,836	4,348,810	3,009,444
Emetina	1,621,844	4,348,173	3,009,401
Micofenolato de mofetilo	1,621,844	4,348,212	3,009,455
Fenazopiridina	1,621,815	4,348,714	3,009,552
Ácido micofenólico	1,621,844	4,348,212	3,009,454
Pamoato de pirvinio	1,621,839	4,348,427	3,009,865
Monensina de sodio	1,621,844	4,348,108	3,009,780
6-mercaptopurina	1,621,824	4,348,550	3,009,572
6-tioguanina	1,621,825	4,348,739	3,009,083
N-metilmaleimida	1,621,844	4,348,583	3,009,900
Disulfiram	1,621,152	4,348,954	3,009,296
Baricitinib	1,621,844	4,348,211	3,009,454
Favipiravir	1,621,755	4,348,166	3,009,272
Darunavir	1,621,844	4,348,212	3,009,455
Arbidol	1,621,844	4,348,212	3,009,453
Nitazoxanida	1,621,844	4,348,172	3,009,397

5.4 Análisis consenso de los resultados de similitud molecular

El número de moléculas en la *MEGA Db* para cada una de las moléculas de referencia que se encuentran simultáneamente para los tres descriptores utilizados en la sección 4.4 se resumen en la Tabla 5.

Tabla 5. Número de compuestos obtenidos para cada molécula de referencia en la *MEGA Db*.

Molécula de referencia	No. de compuestos en la MEGA Db
Remdesivir	11,736
Ivermectina	12,089
Clorhidrato de Emetina	29,448
Fosfato de cloroquina	38,400
Sulfato de hidroxiclороquina	36,174
Mefloquina	28,302
Clorhidrato de amodiaquina	38,447
Aloxistatina	28,773
Clorhidrato de gemcitabina	23,000
Citrato de tamoxifeno	39,135
Citrato de toremifeno	37,810
Terconazol	27,125
Triparanol	50,040
Anisomicina	39,773
Cicloheximida	14,485
Mepesuccinato de omacetaxina	15,727
Mesilato de benzotropina	37,642
Fluspirileno	26,358
Tiotixeno	37,062
Clorhidrato de flufenazina	34,301
Clorhidrato de prometazina	21,421
Astemizol	33,684
Clorhidrato de clorfenoxamina	37,481
Clorhidrato de clorpromazina	27,432
Clorhidrato de trifluoperazina	33,823

Maleato de tietilperazina	30,567
Clorhidrato de clomipramina	35,869
Mesilato de imatinib	27,219
Dasatinib	13,098
Clorpromazina	35,869
Cloroquina	38,400
Toremifeno	37,810
Baicalin	21,290
Glicirricina	16,613
Ácido clorogénico	19,130
Niclosamida	45,217
Cinanserin	29,982
Valinomicina	22,193
Ribavirina	23,240
Loperamida	28,360
Lopinavir	24,210
Nutlin	27,476
Hidroxizina	39,331
Saracatinib	23,395
Sotrastaurina	29,622
Acetofenazina	21,412
Dosulepin	31,878
Metotrimeprazina	28,306
Licorina	35,933
Emetina	29,448
Micofenolato de mofetilo	28,722
Fenazopiridina	20,517
Ácido micofenólico	18,966

Pamoato de pirvinio	20,671
Monensina de sodio	8,528
6-mercaptopurina	42,300
6-tioguanina	42,031
N-metilmaleimida	27,966
Disulfiram	55,589
Baricitinib	20,706
Favipiravir	28,942
Darunavir	29,245
Arbidol	21,001
Nitazoxanida	31,481

La lista final de compuestos consenso se obtuvo a partir de las moléculas de la Tabla 4, la cual contiene un total de 632,149 moléculas únicas (es decir, no repetidas).

5.5 Búsqueda de disponibilidad comercial

Para evaluar la disponibilidad comercial de las 632,149 moléculas únicas de la lista de compuestos consenso, se llevó a cabo una búsqueda rápida (ver sección 4.5), en la cual se obtuvieron un total de 4,423 compuestos que potencialmente se pueden adquirir (muestra física) para su evaluación experimental.

5.6 Perfil ADME/Tox de las moléculas disponibles comercialmente

Uno de los principales problemas en el diseño y desarrollo de fármacos está relacionado con la farmacocinética y toxicidad. Por lo cual evaluar distintas características moleculares para poder predecir el comportamiento de moléculas tipo-fármaco o *drug-like* es de suma importancia.^{103,120,121} Por este motivo el uso de servidores o programas computacionales para el cálculo de descriptores

moleculares se ha convertido en una de las herramientas *in silico* más confiables y prácticas, gracias a su facilidad de obtención.¹²²

Sin embargo, es importante tomar en cuenta que las predicciones computacionales de las propiedades ADME/Tox no pretenden reemplazar las pruebas experimentales. Aún tomando en cuenta lo anterior, el contar con una evaluación preliminar rápida es de amplia utilidad debido a que en su mayoría las pruebas experimentales ADME/Tox son muy costosas y no se pueden realizar rápidamente; especialmente cuando se quieren evaluar a miles de compuestos.

Para calcular el perfil de ADME/Tox de las moléculas disponibles comercialmente se calcularon los descriptores moleculares en la Tabla 3 (ver sección 4.6).

5.6.1 Descriptores moleculares de absorción

Los primeros enfoques *in silico* para predecir la farmacocinética fueron propuestos por Lipinski, et. al., gracias al estudio de más de 2,000 fármacos de la relación entre sus propiedades y actividad.¹²³ A partir de sus resultados obtenidos se desarrolló el concepto conocido actualmente como “regla de los cinco de Lipinski”, las cuales son capaces de predecir qué tan buena puede llegar a ser la absorción intestinal de los compuestos. La regla de los cinco de Lipinski, propuesta en 1996 indica que para que una molécula pueda tener una buena biodisponibilidad oral el número de enlaces donadores de hidrógeno debe ser menor a 5, el número de enlaces aceptores de hidrógeno menor a 10 (5x2), el peso molecular relativo menor a 500 Da y un valor de LogP menor a 5. A su vez es importante considerar que la regla de los cinco de Lipinski es aplicable únicamente a aquellos fármacos cuya vía de administración sea la oral.

Aunque no es necesario que se cumplan los valores de los cuatro descriptores anteriores sugeridos por Lipinski para que un compuesto sea exitoso como fármaco, e incluso hay estudios que demuestran que a lo largo de los años los fármacos aprobados con buena disponibilidad oral tienen valores mayores a los propuestos por Lipinski para los descriptores de peso molecular y número de enlaces aceptores de hidrógeno;¹²⁴ éstos criterios empíricos siguen siendo considerados hasta la fecha como guías para el desarrollo de moléculas tipo-fármaco para que puedan administrarse por vía oral.^{121,123}

En la Figura 13 se observa que el 100% de los compuestos de las bibliotecas moleculares CAS, FooDB y QPNB, el 99.24% de la biblioteca DCM y el 97.93% de la biblioteca COCONUT no presentan violaciones a las reglas de Lipinski; el 0.76% de la biblioteca DCM y el 2.07% de la biblioteca COCONUT tiene una violación. Ninguna biblioteca de compuestos presenta dos o más violaciones a las reglas de Lipinski. Esto implica que en su mayoría los compuestos disponibles comercialmente, independientemente de la biblioteca molecular a la que pertenecen pueden llegar a tener una buena biodisponibilidad administrados por vía oral.

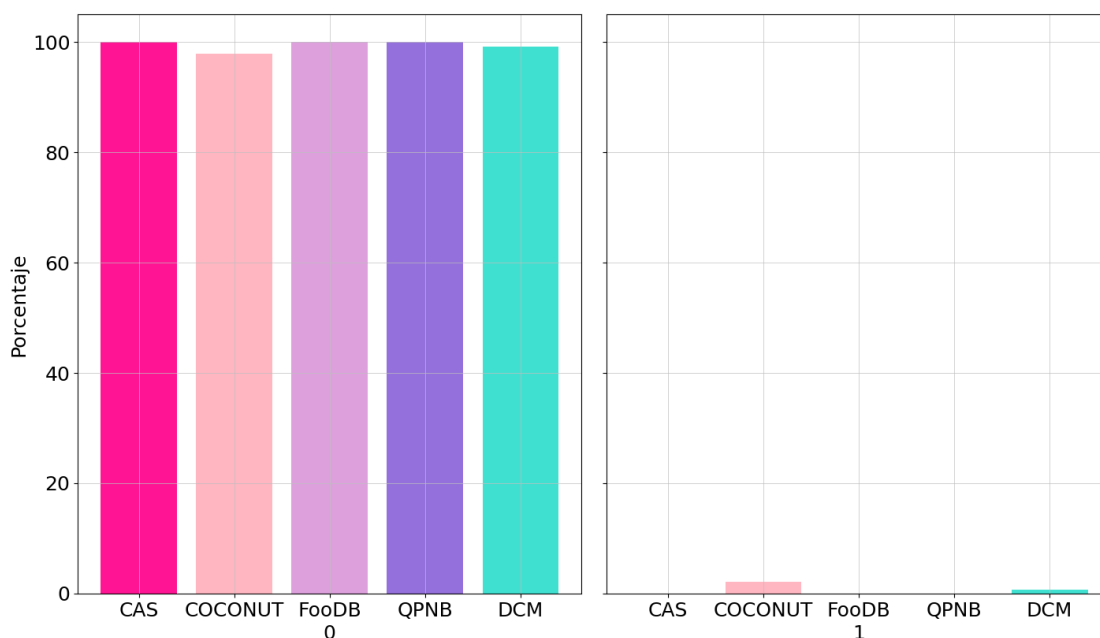


Figura 13. Porcentaje de violaciones a las reglas de Lipinski que presentan los compuestos disponibles comercialmente.

Junto con la regla empírica de Lipinski, la regla de Veber es una de las más empleadas para anticipar la biodisponibilidad oral. Veber et. al., determinaron que hay dos propiedades moleculares que influyen en la biodisponibilidad a partir de un estudio en ratas de más de 1,100 compuestos candidatos a fármacos. La primera propiedad es la flexibilidad molecular reducida, la cual puede ser medida con el número de enlaces rotables; mientras que la segunda es el área polar superficial (TPSA, por sus siglas en inglés) baja representada por la suma del número de enlaces donadores y aceptores de hidrógeno. Así mismo este estudio sugiere que

aquellos compuestos con 10 o menos enlaces rotables y un valor de TPSA igual o menor a 140 \AA^2 o la suma de 12 o menos aceptores y donadores de hidrógeno tienen una mayor probabilidad de tener una buena biodisponibilidad oral.¹²⁰

La Figura 14 muestra que el porcentaje de compuestos que no presentan violaciones a las reglas de Veber son el 100% de la biblioteca QPNB, el 99.22% de la biblioteca FooDB, el 97.02% de la biblioteca COCONUT, el 96.28% de la biblioteca DCM y el 75.00% de la biblioteca CAS. Por su parte, los compuestos que muestran una violación son el 25.00% de la biblioteca CAS, el 3.57% de la biblioteca DCM, el 2.98% de la biblioteca COCONUT y el 0.78% de la biblioteca FooDB. Así mismo el 0.15% de los compuestos de la biblioteca DCM tienen 2 violaciones.

Es importante destacar que las reglas de Veber representan una tendencia de resultados en modelos animales de compuestos candidatos. Por lo que, a pesar de que algunos compuestos de las diferentes bibliotecas moleculares presentan una o dos violaciones a las reglas de Veber; los resultados en conjunto con las reglas de Lipinski muestran que la mayoría de los compuestos en todas las quimiotecas tienen alta probabilidad de tener buena absorción por la vía oral.

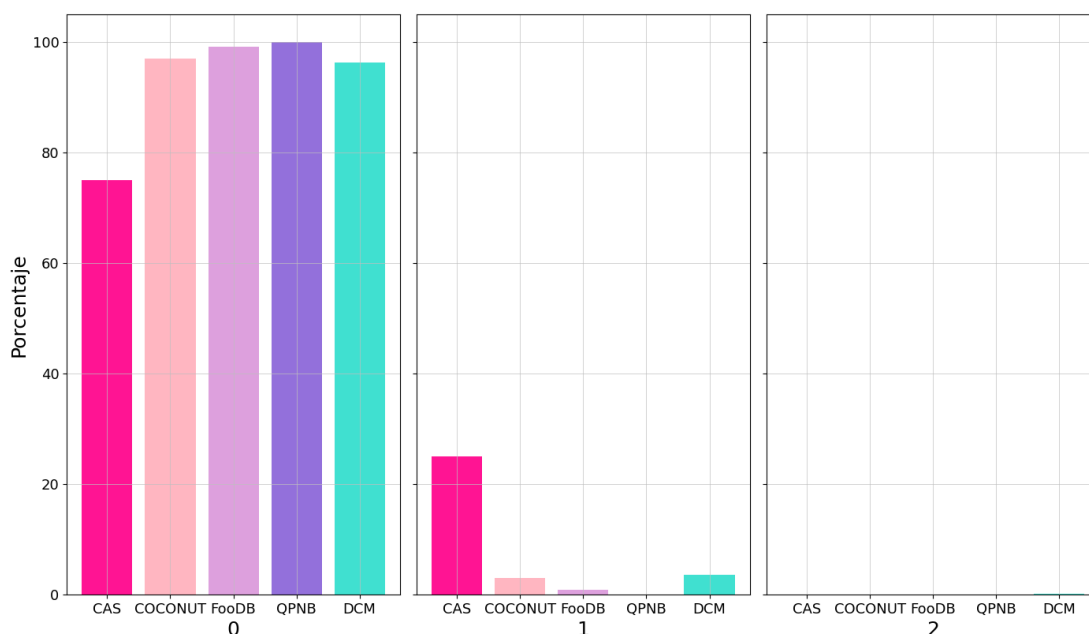


Figura 14. Porcentaje de violaciones a las reglas de Veber que presentan los compuestos disponibles comercialmente.

5.6.1.1 Solubilidad

La solubilidad en agua es una propiedad importante para el desarrollo de fármacos administrados por vía oral, ya que esta influye en el proceso de absorción. Sin embargo, también es uno de los parámetros más difíciles de determinar debido a los polimorfismos moleculares que las estructuras químicas pueden tener.^{125,126}

Para obtener el valor de solubilidad, se seleccionó el modelo *Silicos IT LogSw* del servidor *SwissADME*. Este se basa en una corrección de la predicción del LogS de Filter-it™.^{127,128} La cual considera el peso molecular y tiene un coeficiente de correlación lineal de 0.75 (el cual es mayor a los otros modelos para calcular la solubilidad implementados en SwissADME).¹⁰³ Los valores de solubilidad de un compuesto generalmente se expresan como el logaritmo decimal de S (LogS), donde S es la solubilidad molar (mol/L) en agua. Se suele considerar un valor de LogS ideal entre -5 y -1, debido a que en este rango se considera que existe la polaridad requerida entre la hidrofilia y lipofilia necesaria para el transporte a través de la membrana celular.¹²⁹ Así mismo, SwissADME predice cualitativamente a la solubilidad con base en la siguiente escala:

LogS : insoluble < -10 < escasamente soluble < -6 < moderadamente soluble < -4 < soluble < -2 < muy soluble < 0 < extremadamente soluble.

Para poder conocer de manera visual la distribución de los resultados para cada compuesto en cada base de datos se llevó a cabo la construcción del gráfico de la función de distribución acumulada empírica para los resultados obtenidos. La Figura 15 muestra el gráfico de la función de distribución acumulada de los valores obtenidos de LogS para los compuestos de cada biblioteca molecular. Los compuestos de la biblioteca CAS muestran valores centrados entre -7 y -6 por lo que se consideran escasamente solubles. Por su parte, los compuestos de las bibliotecas COCONUT, FooDB, QPNB y DCM muestran valores centrados entre -6 y -4, los cuales se pueden considerar como moderadamente solubles en agua.

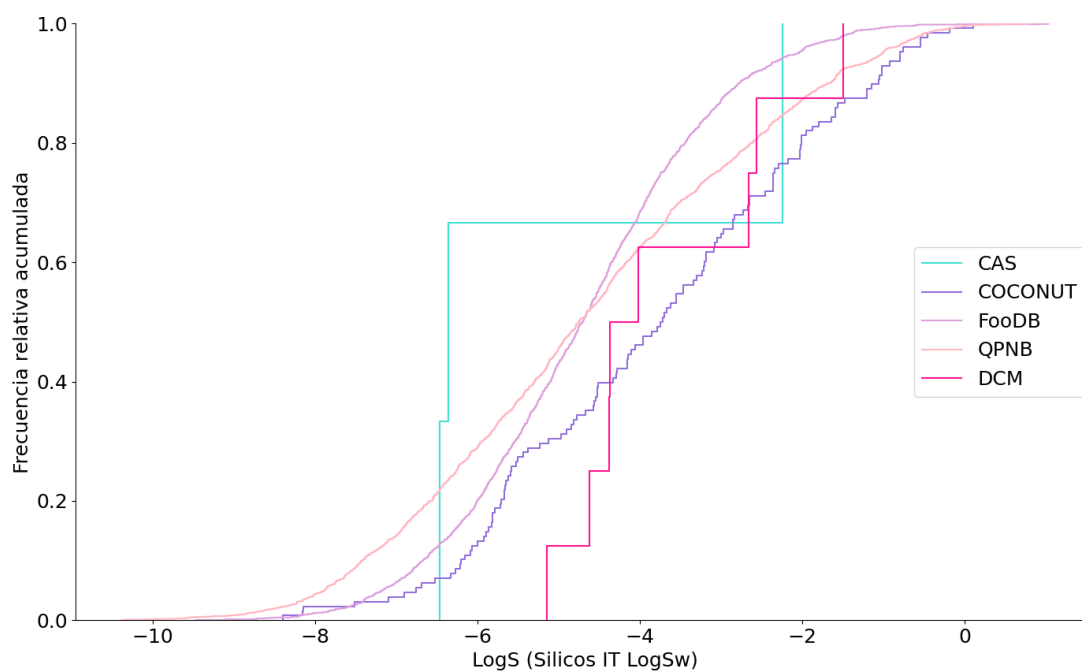


Figura 15. Gráfico de la función de distribución acumulada para la predicción de la solubilidad (LogS) de los compuestos disponibles comercialmente.

5.6.1.2 Lipofilidad

Experimentalmente la lipofilidad de un compuesto se puede cuantificar a partir del coeficiente de partición; el cual es el cociente de la solubilidad del compuesto en dos disolventes inmiscibles entre sí, generalmente n-octanol y agua ($\text{LogP}_{o/w}$).¹³⁰ Debido a que la lipofilidad se encuentra relacionada tanto en procesos farmacocinéticos como farmacodinámicos, así como en la toxicidad que los compuestos puedan presentar, la lipofilidad es una de las propiedades más importantes a considerar en el diseño de fármacos.¹³¹ Generalmente los fármacos liposolubles atraviesan la membrana celular por difusión pasiva a diferencia de los compuestos hidrosolubles, los cuales suelen quedar atrapados en la bicapa lipídica de la membrana.¹³²

Para el caso de los servidores en línea, el $\text{LogP}_{o/w}$, es el descriptor clásico para expresar la lipofilidad. SwissADME proporciona seis modelos para el cálculo de la predicción de la lipofilia:

1. XLOGP3: predice el valor de LogP a partir de un modelo aditivo basado en los átomos y un factor de corrección, el cual emplea un valor de LogP de un

compuesto de referencia a partir de la similitud estructural y un análisis de regresión lineal multivariante de 8,199 compuestos orgánicos.¹³³

2. WLOGP: es un modelo construido a partir de un entrenamiento con 9,920 moléculas basado en los átomos a partir del sistema de fragmentación propuesto por Wildman y Crippen, el cual además de considerar el valor del LogP, considera la refractividad molar.¹³⁴
3. MLOGP: este modelo se construyó con base en un análisis de regresión múltiple que considera 13 descriptores moleculares a partir de una biblioteca de 1,230 compuestos.¹³⁵
4. SILICOS-IT: se basa en un método patentado de Filter-it™, que involucra fragmentos y descriptores moleculares. El modelo se entrenó a partir de un ajuste de mínimos cuadrados de 23,455 valores experimentales.¹⁰³
5. iLOGP: este consiste en ser un modelo lineal múltiple que considera valores experimentales del $\text{LogP}_{o/w}$ y la energía libre de solvatación en solventes implícitos de un conjunto de más de 17,500 moléculas.¹³⁶
6. Consensus LogP: es la media aritmética de los valores predichos por los cinco métodos mencionados anteriormente.¹⁰³

Para los compuestos disponibles comercialmente se realizó el gráfico de la función de distribución acumulada empírica para el parámetro de lipofilicidad Consensus $\text{LogP}_{o/w}$ calculado con el servidor SwissADME. Los resultados se resumen en la Figura 16, donde se muestra que el promedio de los valores de $\text{LogP}_{o/w}$ para la biblioteca CAS se encuentran entre 1 y 3, para la biblioteca COCONUT entre 1.5 y 2.5, para las biblioteca FooDB y QPNB entre 2 y 3 y para la biblioteca DCM 1 y 2.5.

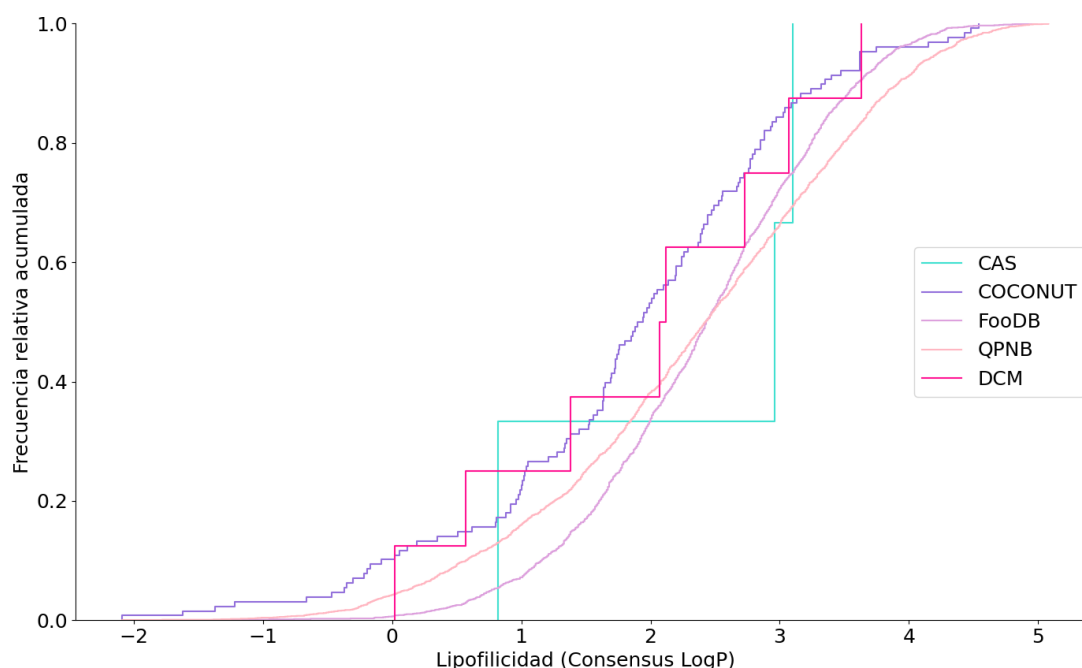


Figura 16. Gráfico de la función de distribución acumulada para la predicción de la lipofilidad (LogP) de los compuestos disponibles comercialmente.

5.6.1.3 Absorción intestinal humana

La mayoría de los xenobióticos se absorben a través de transporte pasivo o activo por medio de las células epiteliales, las cuales forman la mucosa del intestino grueso y delgado; para así poder llegar al torrente sanguíneo.¹³⁷ La absorción intestinal humana (HIA) representa una de las propiedades ADME/Tox más importantes a tener en cuenta. Sin embargo, HIA también es una de las más difíciles de predecir con modelos estadísticos. Esto se debe a que la absorción de fármacos es un proceso extremadamente complicado que involucra una gran cantidad de descriptores relacionados con el transporte mediado por proteínas transportadoras y el metabolismo de primer paso.¹³⁸

Debido a que la administración oral es la vía de administración preferida por los pacientes gracias a su facilidad y practicidad; la biodisponibilidad oral es una de las propiedades más deseables para el desarrollo de nuevos fármacos. Así mismo, se ha demostrado que la biodisponibilidad se encuentra controlada principalmente por la absorción intestinal. Es por esto que la HIA es un indicador alternativo para la biodisponibilidad oral.¹³⁹

La HIA para una sustancia se encuentra frecuentemente cuantificada como la porción de esta sustancia capaz de llegar a la vena porta o la sangre, y se puede calcular a partir de la siguiente ecuación:

$$HIA = D_{sangre}/D_{oral}$$

dónde D_{sangre} es la cantidad de sustancia capaz de llegar a la vena porta y D_{oral} es la cantidad de sustancia total administrada por la vía oral.¹⁴⁰

El cálculo de la HIA en el servidor *pkCSM-pharmacokinetics* utiliza un modelo construido a partir de cuatro descriptores: TPSA, $\text{LogP}_{o/w}$ a pH = 6,5, el número de violaciones a la regla de los cinco de Lipinski y el número de enlaces donadores de hidrógeno. Este modelo obtuvo un valor de $r = 0.84$ y un error de predicción del 11.2% para el conjunto de compuestos con el que fue entrenado. A su vez, clasifica los resultados en dos clases; aquellos compuestos con resultados de HIA menores o iguales al 30% se consideran como pobremente absorbidos intestinalmente, mientras que aquellos compuestos con resultados mayores al 30% se consideran que tienen una buena absorción intestinal.¹⁴¹

La Figura 17 muestra el gráfico de la función de distribución acumulada de los porcentajes obtenidos de HIA para los compuestos presentes en cada biblioteca molecular de los compuestos disponibles comercialmente. La gráfica indica que más del 90% de los compuestos de todas las bibliotecas se considera que tienen una buena absorción intestinal.

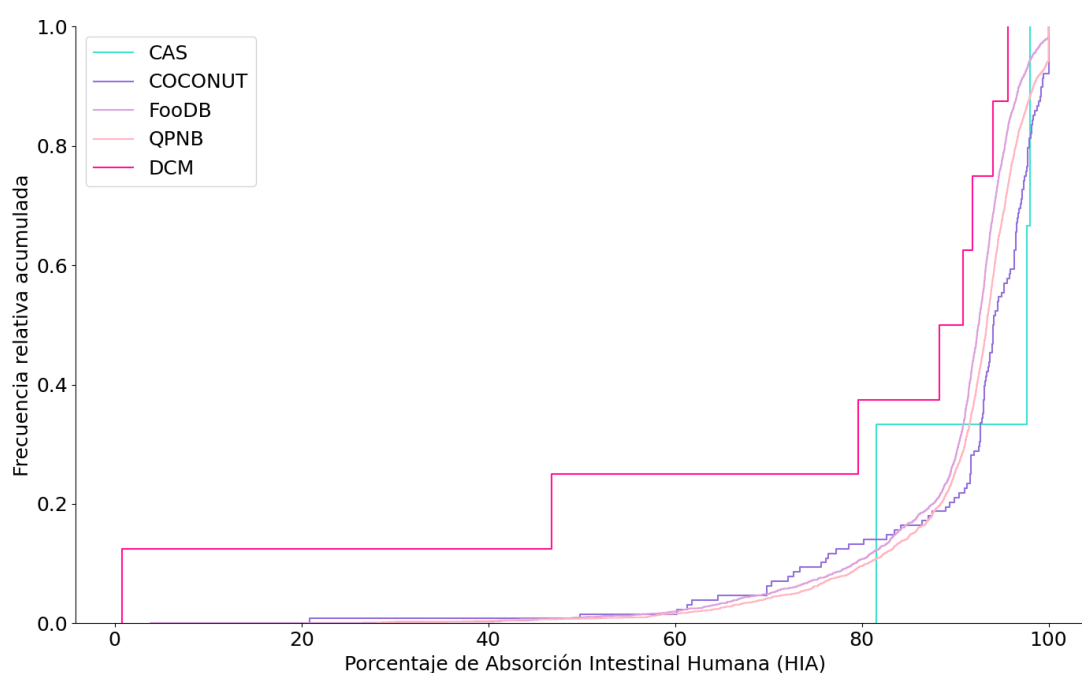


Figura 17. Gráfico de la función de distribución acumulada para la predicción de la absorción intestinal humana (HIA, por sus siglas en inglés) de los compuestos disponibles comercialmente para diferentes bibliotecas moleculares estudiadas.

5.6.2 Descriptores moleculares de distribución

En el desarrollo de moléculas pequeñas como posibles fármacos, es de suma importancia poder comprender la biodistribución de la forma activa de estas en el sitio activo terapéutico; para así poder establecer una relación entre su farmacocinética y farmacodinamia. De manera experimental, esta relación se determina midiendo la concentración del fármaco de interés en el compartimiento sanguíneo para posteriormente asociar esta concentración con la respuesta obtenida.¹⁴²

En los últimos 10 años, se han desarrollado diversos modelos matemáticos que involucran descriptores capaces de integrar procesos individuales que sean capaces de representar las etapas del proceso ADME/Tox.^{103,143,144} De esta manera el proceso de distribución puede representarse a través de algoritmos capaces de predecir el paso de moléculas por la barrera hematoencefálica (BBB, por sus siglas en inglés) y la unión a proteínas plasmáticas.

5.6.2.1 Permeabilidad de la barrera hematoencefálica

La BBB es el conjunto de vasos sanguíneos y células endoteliales microvasculares, esta tiene como principal función mantener la homeostasis del sistema nervioso central (SNC) a través de diversos mecanismos como el transporte de nutrientes, la eliminación de metabolitos del cerebro y proteger al cerebro de los efectos secundarios de diversos xenobióticos. Sin embargo, en algunos casos, como el tratamiento de enfermedades neurológicas puede llegar a ser indispensable que ciertos fármacos lleguen al cerebro en concentraciones terapéuticas.^{145,146}

La determinación experimental de la permeabilidad de la BBB para moléculas pequeñas es un proceso complicado. Generalmente, suelen utilizarse animales de experimentación como ratas o ratones para determinar el aclaramiento del tejido cerebral.¹⁴⁷ Por otra parte, los modelos de predicción de la permeabilidad de la BBB son de suma importancia en las primeras etapas del desarrollo de fármacos en un

intento de disminuir la tasa de deserción de compuestos candidatos a fármacos en ensayos preclínicos y clínicos.¹⁴⁵

Las predicciones *in silico* de la permeabilidad de la BBB se ha convertido en una herramienta económica y que requiere menos tiempo. Las predicciones se basan generalmente en resultados de estudios *in vitro* e *in vivo* realizados previamente. A pesar de esto, una de las principales desventajas es que estos modelos suelen asumir que los fármacos atraviesan la barrera BBB por difusión pasiva, por lo tanto no suelen considerar varios mecanismos de transporte que pueden llevarse a cabo en la BBB como el transporte activo mediado por transportadores y receptores.¹⁴⁸ Debido a lo anterior, es importante considerar otros descriptores como el transporte activo mediado por la glicoproteína P y el metabolismo de los compuestos de interés.

En la presente investigación el cálculo de la permeabilidad de la BBB se realizó en el servidor *pkCSM-pharmacokinetics*, el cual cuenta con un valor de $r = 0.862$ y un error de predicción del 8.27% para el conjunto de compuestos con el que fue entrenado. Este modelo clasifica los resultados a partir de la medida de permeabilidad cuantitativa (LogPS) y fue entrenado a partir de los resultados experimentales de 120 compuestos. Aquellos compuestos con valores de LogPS ≥ -2 se consideran con la capacidad de tener una penetración fácil, mientras que aquellos con valores < -2 se consideran como no penetrantes.¹⁴⁷

La Figura 18 muestra el gráfico de la función de distribución acumulada de los valores del LogPS para la permeabilidad de la BBB de los compuestos presentes en cada biblioteca molecular de los compuestos disponibles comercialmente. Los resultados indican que más del 95% de los compuestos de todas las bibliotecas se considera que tienen capacidad fácil de penetración de la BBB.

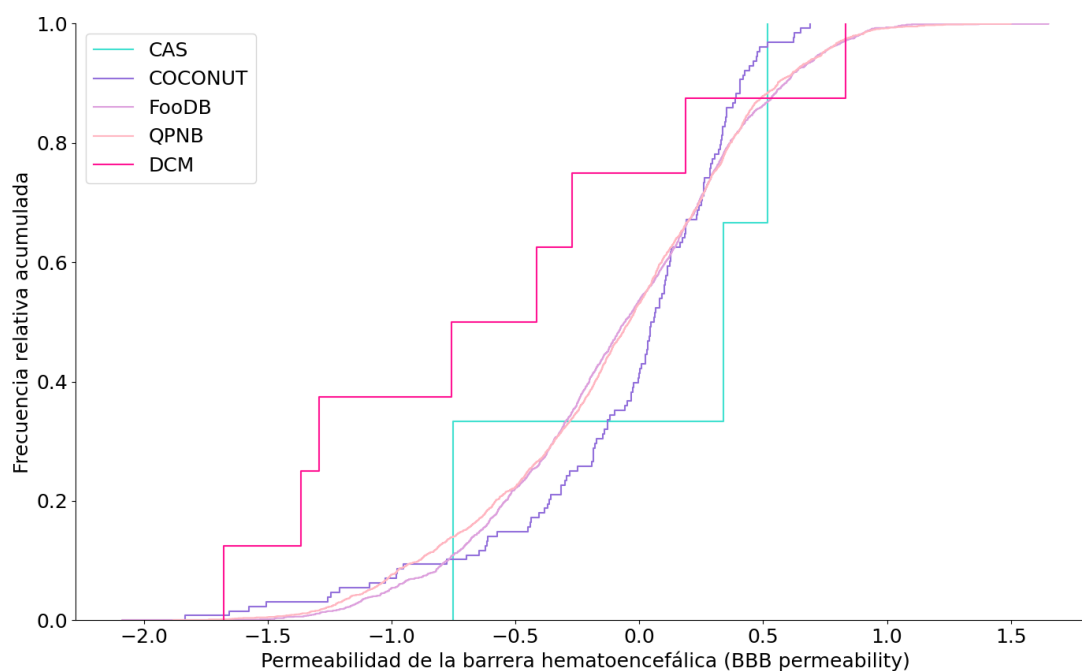


Figura 18. Gráfico de la función de distribución acumulada para la predicción de la permeabilidad de la barrera hematoencefálica (LogPS) de los compuestos disponibles comercialmente.

5.6.2.2 Sustrato de la glicoproteína P

La glicoproteína P es el transportador de unión al adenosin trifosfato (ATP, por sus siglas en inglés) más estudiado. Esta funciona como una barrera biológica de defensa debido a que es capaz de expulsar toxinas y xenobióticos de las células. Así mismo, diversos estudios han demostrado que la glicoproteína P desarrolla un papel importante en las fases de absorción y distribución durante el perfil ADME/Tox.¹⁴⁸

Debido a su localización en las células endoteliales del lumen intestinal, la glicoproteína P tiene una amplia participación en el transporte de fármacos hacia la circulación sanguínea y en el proceso de excreción de estos mismos compuestos.

La glicoproteína P tiene dos dominios citoplasmáticos de unión al ATP; y cada dominio está conformado por tres motivos: Walker A, Walker B y C. Se ha reportado que el residuo de Lys en el motivo Walker A es el que se encuentra relacionado directamente con la unión del ATP a la glicoproteína P.^{149,150} Debido a lo anterior, la unión de ATP a la glicoproteína P para su posterior hidrólisis son los

responsables del transporte de diversos xenobióticos, como es el caso de fármacos mediado por esta proteína.

Es bien conocido que análogos del ATP son capaces de unirse a los motivos Walker A de la glicoproteína P. Sin embargo, es necesario que uno de los dos motivos presentes en ambos dominios actúe catalíticamente hidrolizando ATP para que esta proteína pueda ejercer su función como transportador. La estequiometría de la hidrólisis del ATP asociada al transporte de fármacos depende del sustrato y puede ir en un rango de 0.3 a 6 moléculas de ATP por cada molécula de fármaco.¹⁵¹

A pesar de que los motivos Walker A de los dominios de unión al ATP se encuentran restringidos a reconocer ATP o sus análogos, se han identificado dominios transmembranales en la glicoproteína P (TM1, TM2, TM3, TM4, TM5, TM6, TM7, TM8, TM9, TM10, TM11 y TM12) sensibles al reconocimiento diversos sustratos.^{151,152,153,154}

A pesar de que la glicoproteína P es capaz de transportar moléculas con un amplio rango de peso molecular (250 a 1202 Da). Es importante considerar si los candidatos a fármacos pueden llegar a ser sustratos de esta proteína, ya que esto ayuda a conocer el flujo de transporte activo a través de la membrana celular y así mismo tener en cuenta que esta característica puede llegar a afectar el perfil ADME/Tox predicho del compuesto.¹⁰³

La predicción de si los compuestos seleccionados (disponibles comercialmente) pudieran ser sustratos de la glicoproteína P se realizó en el servidor *SwissADME*, el cual utiliza un algoritmo de aprendizaje automático construido a partir de 16 descriptores moleculares y 1033 moléculas de entrenamiento con una precisión de validación externa del 94%. Este modelo devuelve un resultado de tipo categórico.¹⁰³

En la Figura 19 se observa que el 62.5% de los compuestos de las biblioteca CAS, el 57.6% de la biblioteca COCONUT, el 50% de la biblioteca FooDB, el 66.7% de la biblioteca QPNB y el 54.2% de la biblioteca DCM tienen predicciones positivas a ser sustratos de la glicoproteína P.

5.6.2.3 Inhibición de la glicoproteína P

Como se discutió en la sección anterior (5.6.2.2), la función transportadora de moléculas por la glicoproteína P se encuentra mediada por los dos dominios de

unión al ATP; así mismo, la interacción de fármacos como sustratos a esta proteína tienen una acción cooperativa en la unidad funcional. Por esto, la interacción con moléculas o fármacos también puede resultar en la inhibición competitiva de la glicoproteína p, lo cual conduce a una disminución de la hidrólisis del ATP y por lo tanto en un decremento de la función transportadora de la glicoproteína P.

El cálculo de la inhibición de la glicoproteína P en el servidor *pkCSM-pharmacokinetics* cuenta con la implementación de dos diferentes predictores.¹⁰⁴ El primer predictor (denominado como, inhibición de la glicoproteína P-I) se construyó a partir de un modelo bayesiano y 1,273 moléculas que se encuentran clasificadas como inhibidores y no inhibidores de la glicoproteína P; cuenta con una precisión de validación externa del 81.2%.¹⁴¹ El segundo predictor denominado como inhibición de la glicoproteína P-II, fue construido a partir de una biblioteca de 1,200 compuestos y métodos QSAR multivariados. Este modelo cuenta con una precisión del 88%.¹⁵⁵

En la Figura 19 se observa que el 12.5% de los compuestos de la biblioteca CAS, el 28.3% de la biblioteca COCONUT, el 12.5 % de la biblioteca FooDB, el 0% de la biblioteca QPNB y el 26.4% de la biblioteca DCM tienen predicciones positivas a ser inhibidores de la glicoproteína P evaluados con el primer modelo; mientras que el 0% de los compuestos de la biblioteca CAS, el 27.2% de la biblioteca COCONUT, el 16.4 % de la biblioteca FooDB, el 33.3% de la biblioteca QPNB y el 17.1% de la biblioteca DCM tienen predicciones positivas a ser inhibidores de la glicoproteína P evaluados con el segundo modelo. Debido a los valores de precisión de ambos modelos para predecir la inhibición de la glicoproteína p se puede considerar que el segundo modelo implementado tiene un mejor rendimiento debido a que tiene una mayor precisión.

A su vez, también se puede llevar a cabo un análisis adicional en donde se obtengan los compuestos con resultados positivos para ambas predicciones, ya que como cada modelo emplea un algoritmo diferente se podrían obtener los compuestos con más alta probabilidad de ser inhibidores de la glicoproteína p que solo considerando cada modelo de manera independiente.

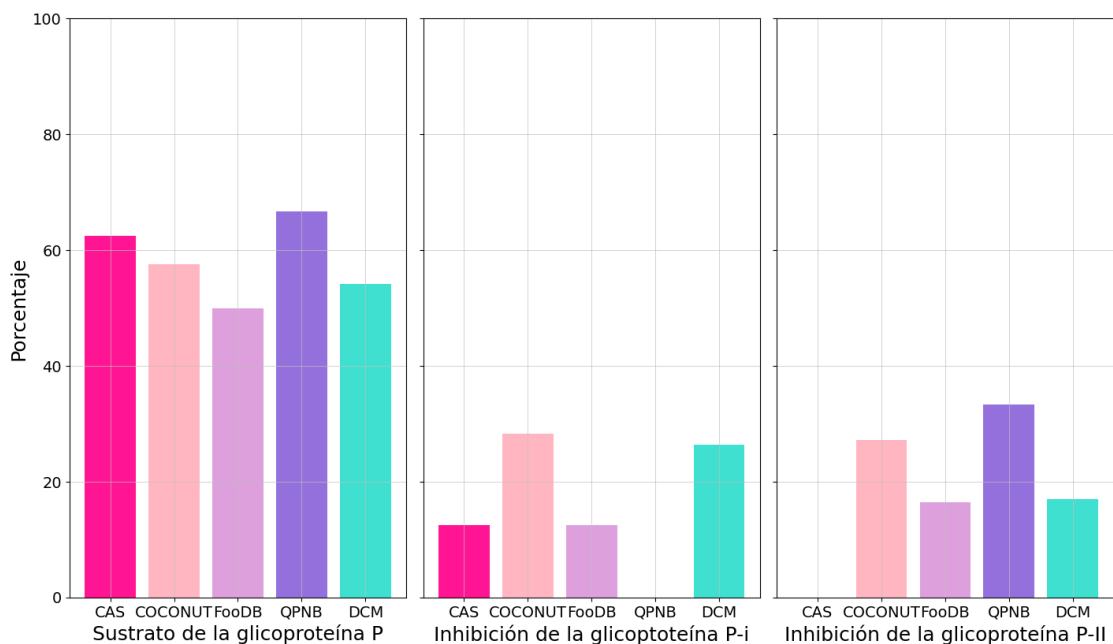


Figura 19. Porcentaje de compuestos posibles a ser sustratos de la glicoproteína P e inhibidores de la glicoproteína P a partir de los dos modelos implementados que presentan los compuestos disponibles comercialmente.

5.6.3 Descriptores moleculares de metabolismo

El metabolismo es un proceso que involucra la participación de diversas proteínas, como es el caso de las enzimas y diversos transportadores. Por lo que es de suma importancia considerar el metabolismo de los compuestos candidatos a fármacos para poder analizar su posible farmacocinética, farmacodinamia y toxicidad. Así mismo, el metabolismo se caracteriza por ser un proceso conformado por biotransformaciones complejas donde los fármacos son modificados estructuralmente.¹⁵⁶ Por tal motivo, el desarrollo de algoritmos capaces de predecir el metabolismo de moléculas es de gran utilidad para la optimización de los compuestos candidatos a fármacos.

En general, las enzimas metabolizadoras tienen sustratos específicos; sin embargo, algunos fármacos o moléculas exógenas son susceptibles a ser metabolizadas por enzimas como el citocromo (CYP) P450. Las enzimas de la familia del CYP P450 son hemoproteínas que se encuentran localizadas en su mayoría en el intestino y se les atribuye el ser las responsables de la desintoxicación de muchos compuestos tanto endógenos como exógenos,

metabolismo celular y la homeostasis. Así mismo, son capaces de ser inducidas o inhibidas por varios fármacos.¹⁵⁷

El servidor *SwissADME* cuenta con funciones para predecir la actividad de algunos de los CYPs frente a los compuestos de interés. La Figura 20 muestra el porcentaje de compuestos por biblioteca molecular de los compuestos disponibles comerciales que son predichos por *SwissADME* como inhibidores de los CYP1A2, CYP2C19, CYP2C9, CYP2D6, CYP3A4.

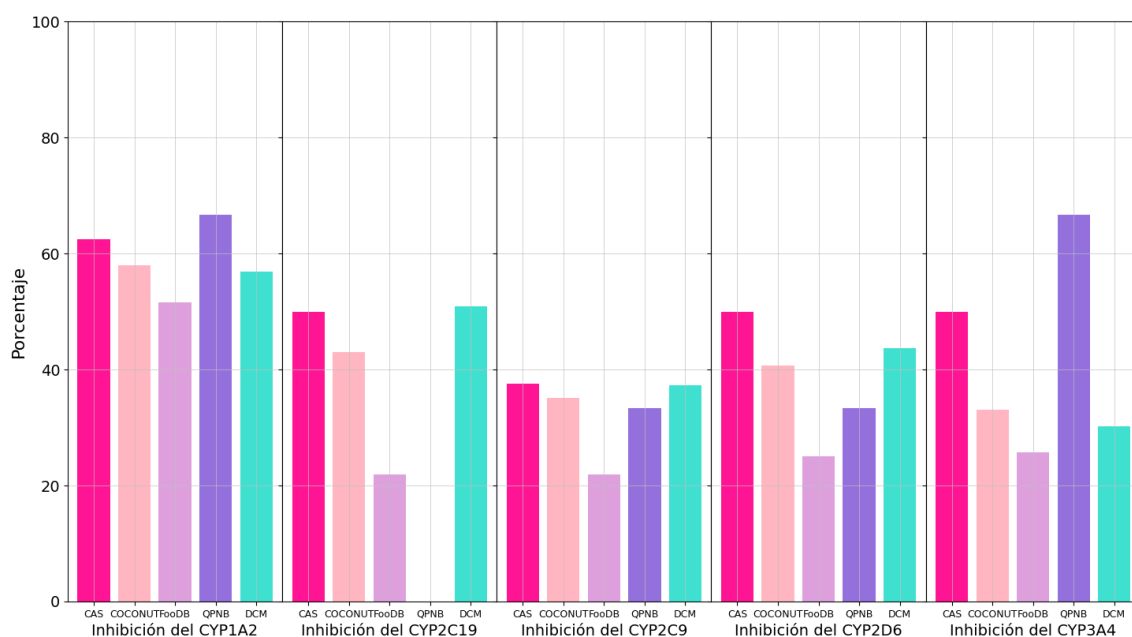


Figura 20. Porcentaje de compuestos posibles a ser inhibidores de los diferentes citocromos (CYPs) de los compuestos disponibles comercialmente.

5.6.3.1 Inhibición del CYP1A2

El CYP1A2 representa aproximadamente el 12% del contenido total de CYP P450 en el hígado en humanos. Entre los principales sustratos del CYP1A2 se encuentran medicamentos, productos químicos industriales y tóxicos ambientales que afectan el nivel de expresión y actividad de la enzima. Así mismo la expresión y regulación del CYP1A2 está altamente regulada por el receptor de hidrocarburos arilo y es conocido que participa en la activación de aminas aromáticas y heterocíclicas e hidrocarburos aromáticos policíclicos. Este CYP también se encuentra involucrado con la activación de procarcinógenos; por lo que una alta

expresión de esta enzima conduce a aumentar el riesgo de los pacientes al cáncer mientras que su inhibición disminuye la posibilidad de desarrollarlo.¹⁵⁸

La Figura 20 muestra que el porcentaje de compuestos que se predicen con actividad inhibitoria del CYP1A2 son el 62.50% de la biblioteca CAS, el 58.04% de la biblioteca COCONUT, el 51.56% de la biblioteca FooDB, el 66.67% de la biblioteca QPNB y el 56.95% de la biblioteca DCM.

5.6.3.2 Inhibición del CYP2C19

La enzima CYP2C19 es importante clínicamente debido a que es la responsable del metabolismo de ciertos fármacos como el clopidogrel, omeprazol, lansoprazol, diazepam, propranolol, entre otros. También se sabe que se encuentra involucrada con la desintoxicación de potenciales carcinógenos e inhibe la bioactivación de procarcinógenos ambientales, impidiendo la formación de metabolitos reactivos con afinidad por el ADN.¹⁵⁹

La Figura 20 muestra que el porcentaje de compuestos que se predicen con actividad inhibitoria del CYP2C19 son el 50% de la biblioteca CAS, el 42.95% de la biblioteca COCONUT, el 21.88% de la biblioteca FooDB, el 0% de la biblioteca QPNB y el 50.94% de la biblioteca DCM.

5.6.3.3 Inhibición del CYP2C9

El CYP2C9 es una de las enzimas CYP más abundantes en el hígado humano, se encuentra asociada con las reacciones de fase I en la oxidación de xenobióticos; a su vez, es la responsable del metabolismo de más del 15% de los fármacos aprobados, entre los cuales se encuentran hipoglucemiantes, antiinflamatorios no esteroideos, inhibidores selectivos de la ciclooxigenasa-2, antiepilépticos, inhibidores del receptor de angiotensina II, anticoagulantes, entre otros,¹⁶⁰ Su función metabólica hacia fármacos se ha asociado con compuestos débilmente ácidos y grupos funcionales como las aminas aromáticas y heterocíclicas. Así mismo existe una gran variedad de fármacos que actúan como inhibidores específicos del CYP2C9, como el fluconazol, voriconazol y amiodarona.^{161,162}

La Figura 20 indica que el porcentaje de compuestos que se predicen con actividad inhibitoria del CYP2C9 son el 37.50% de la biblioteca CAS, el 35.06% de la biblioteca COCONUT, el 21.88% de la biblioteca FooDB, el 33.33% de la biblioteca QPNB y el 37.29% de la biblioteca DCM.

5.6.3.4 Inhibición del CYP2D6

CYP2D6 es un gen altamente polimórfico con más de 100 variantes alélicas; lo que conlleva a una clasificación de cuatro fenotipos diferentes: metabolizadores lentos, metabolizadores intermedios, metabolizadores normales y metabolizadores ultrarrápidos. Debido a lo anterior esta enzima es la responsable de la variabilidad interindividual de las concentraciones plasmáticas de los diferentes xenobióticos; siendo los metabolizadores lentos son aquellos que capaces de mostrar concentraciones plasmáticas más altas y una vida media más larga.¹⁶³ Estas últimas características también se pueden observar con la presencia de fármacos inhibidores de esta proteína.

El análisis de la Figura 20 indica que el porcentaje de compuestos que se predicen con actividad inhibitoria del CYP2D6 son el 50% de la biblioteca CAS, el 40.75% de la biblioteca COCONUT, el 25% de la biblioteca FooDB, el 33.33% de la biblioteca QPNB y el 43.66% de la biblioteca DCM.

5.6.3.5 Inhibición del CYP3A4

El CYP3A4 es cuantitativamente la enzima más importante de la familia del P450 debido a que además de expresarse en el hígado, también se encuentra en el intestino delgado; contribuyendo así al metabolismo presistémico y sistémico de aproximadamente el 30% de los fármacos aprobados. Así mismo se sabe que el gen *CYP3A4* se encuentra altamente conservado debido a su importancia en el metabolismo tanto endógeno como exógeno.¹⁶⁴

La Figura 20 muestra que el porcentaje de compuestos que se predicen con actividad inhibitoria del CYP3A4 son el 50% de la biblioteca CAS, el 33.03% de la biblioteca COCONUT, el 25.78% de la biblioteca FooDB, el 66.67% de la biblioteca QPNB y el 30.16% de la biblioteca DCM.

Los resultados anteriores indican que todas las bibliotecas moleculares cuentan con la presencia de compuestos inhibidores de todos los CYPs mencionados anteriormente (CYP1A2, CYP2C19, CYP2C9, CYP2D6 y CYP3A4). Por lo que es importante tener en cuenta si alguno de los compuestos inhibidores es potencialmente un candidato a fármaco para tomar en cuenta el ajuste de las dosis además de considerar las posibles optimizaciones que puedan mejorar el metabolismo que puedan llegar a tener.

5.6.4 Descriptores moleculares de eliminación

El aclaramiento de un fármaco se mide a partir del aclaramiento o eliminación total (*Total clearance*). Este se expresa como una constante de proporcionalidad, la cual describe la relación entre la velocidad de transferencia de un compuesto o sustancia en cantidad por unidad de tiempo. La eliminación de un fármaco ocurre debido a la perfusión de sangre a los órganos de excreción; generalmente el hígado y riñones.¹⁶⁵

El valor de eliminación total de un fármaco es un parámetro farmacocinético relevante debido a que se encuentra directamente relacionado con la disponibilidad y la eliminación de este, lo cual puede ser de ayuda al determinar la tasa de dosificación y la concentración a la que el fármaco se encuentra en condiciones estables.¹⁶⁶

5.6.4.1 Eliminación total

Experimentalmente, el valor de eliminación total de un compuesto candidato a fármaco se obtiene a partir de estudios *in vitro* e *in vivo*; lo cual conlleva a que este proceso sea lento y costoso. Debido a lo anterior, se han explorado y construido modelos *in silico*.¹⁶⁷

Se determinó el valor de eliminación para los compuestos disponibles comercialmente a partir del modelo implementado en el servidor *pkCSM-pharmacokinetics*. Este modelo se construyó a partir de un conjunto de entrenamiento de 398 compuestos obteniendo un valor de R^2 de 0.755 y un error estándar de 0.245.¹⁶⁵

La Figura 21 muestra el gráfico de la función de distribución acumulada del Log de los valores de la eliminación total (mL/min/Kg) de los compuestos presentes

en cada biblioteca molecular de los compuestos disponibles comercialmente donde se observa que la mayoría de los compuestos de la biblioteca QPNB tiene valores cercanos a 1. Debido a estos resultados se puede predecir que en comparación con los compuestos de las otras bibliotecas (CAS, COCONUT, FooDB, DCM), los compuestos que conforman la biblioteca molecular QPNB pueden ser eliminados con mayor facilidad del organismo.

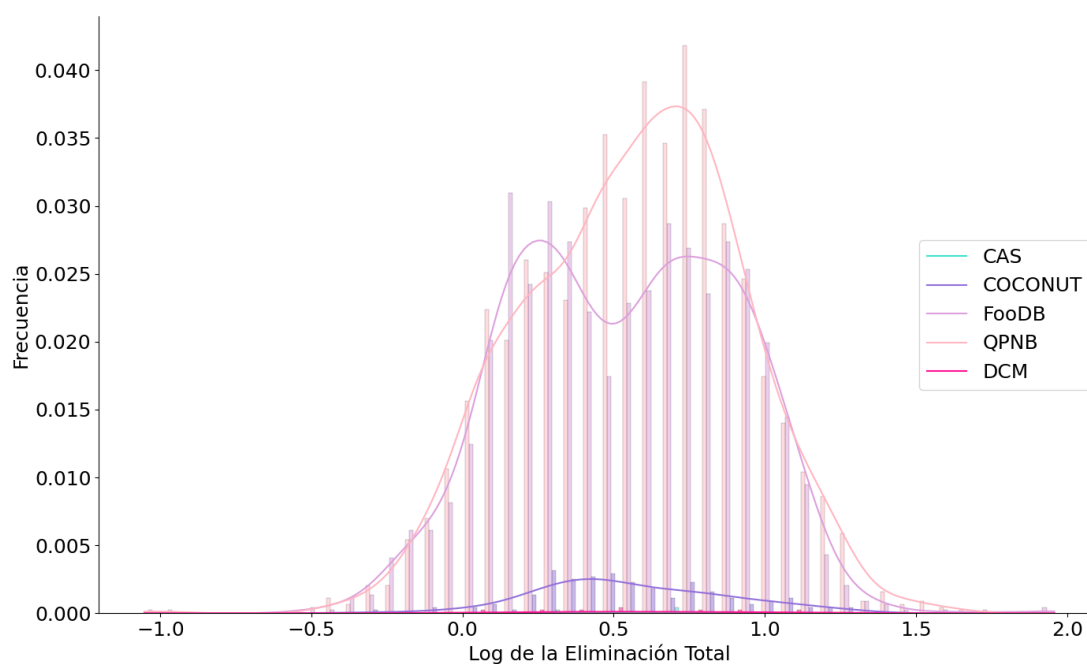


Figura 21. Gráfico de distribución para la predicción del \log_{10} de los valores obtenidos de la eliminación total en mL/min/Kg de los compuestos disponibles comercialmente.

5.6.5 Descriptores moleculares de toxicidad

La toxicidad es una de las principales causas por las que solo pocos de los candidatos fármacos son capaces de llegar al mercado como un medicamento. Hasta hace un par de décadas, aproximadamente el 50% de los compuestos terapéuticos potenciales fracasaban debido a efectos adversos inaceptables.¹⁴³ Por esta razón, el desarrollo de algoritmos predictores de descriptores relacionados con la toxicidad se han convertido en un área de especial interés.

Para los compuestos disponibles comercialmente se hizo la predicción categórica para la prueba *in silico* de AMES y la hepatotoxicidad con el servidor *pkCSM-pharmacokinetics*. Los resultados se muestran en la Figura 22.

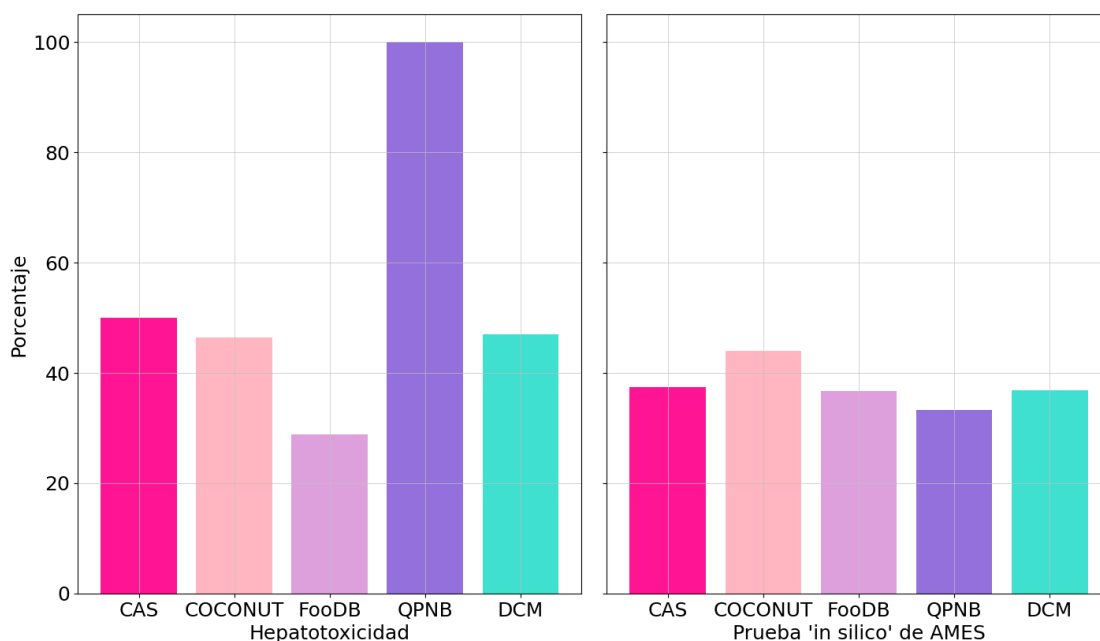


Figura 22. Porcentajes de toxicidad de pruebas *in silico* para los compuestos disponibles comercialmente.

5.6.5.1 Prueba *in silico* de AMES

La prueba de mutagenicidad de *Salmonella*, también conocida como prueba de AMES, proporciona un procedimiento rápido para determinar la mutagenicidad de sustancias químicas o metabolitos generados a partir de la sustancia química de interés durante el proceso metabólico.¹⁶⁸

La prueba de AMES consiste en utilizar cepas de bacterias mutantes a histidina, generalmente *Escherichia coli* o *Salmonella Typhimurium*. El fundamento de la prueba se basa en que la histidina es un aminoácido esencial para que las bacterias puedan crecer en el medio. Por lo tanto las cepas mutadas no podrán formar colonias en placas con medio en ausencia de histidina. Sin embargo; si estas bacterias se encuentran ante la exposición de alguna sustancia mutagénica que sea capaz de llevar a cabo un cambio, delección o adición de alguna base nucleotídica; esto conlleva a que se revierta la secuencia del gen mutado de histidina al del tipo

silvestre. Por lo que las bacterias revertidas serán capaces de llevar a cabo la síntesis de histidina y por lo tanto el crecimiento de colonias en el medio.¹⁶⁸

Al realizar la prueba de AMES, las cepas mutadas se incuban en presencia de un disco de papel filtro que contiene la sustancia de interés el cual se difunde en el agar formando un gradiente de concentración. Así mismo se realiza el ensayo añadiendo la fracción metabólica de un homogenizado de hígado de rata fortificado (fracción S9), el cual funciona como activador metabólico de la sustancia de interés para poder evaluar también la mutagenicidad de los metabolitos derivados. Un resultado positivo de mutagenicidad para la prueba de AMES se observa con el crecimiento de colonias bacterianas en medio con ausencia de histidina, mientras que la ausencia de estas conduce a un resultado negativo.¹⁶⁹

El modelo para la prueba de AMES del servidor *pkCSM-pharmacokinetics* utiliza el árbol de decisiones de Cramer. Este modelo de predicción consta de 33 preguntas con respuestas categóricas binarias (sí o no); el rendimiento de clasificación comparativa de este modelo reporta un valor de área bajo la curva (ABC) de 0.909.^{104,170}

La Figura 22 muestra los compuestos con predicciones positivas para el ensayo *in silico* de AMES, siendo estos el 37.50% de la biblioteca CAS, el 44.09% de la biblioteca COCONUT, el 36.72% de la biblioteca FooDB, el 33.33% de la biblioteca QPNB y el 36.88% de la biblioteca DCM. Con base en lo anterior, es de suma importancia tener presentes cuales son los compuestos con predicciones positivas para este ensayo debido a que al ser potencialmente mutagénicos es muy probable que sean descartados como candidatos a fármacos.

5.6.5.2 Hepatotoxicidad

La lesión hepática inducida por fármacos es una de las principales causas por las cuales se retiran los fármacos del mercado posterior a su desarrollo clínico y aprobación; por lo tanto, la eliminación en las primeras etapas del diseño de fármacos de compuestos candidatos capaces de causar daño hepático puede reducir significativamente la tasa de deserción y el costo en el desarrollo de fármacos.¹⁷¹

El modelo para predecir la hepatotoxicidad implementado en el servidor *pkCSM-pharmacokinetics* se construyó a partir de la estructura química de 951

compuestos y datos biológicos, los cuales comprenden la información de los efectos secundarios en el hígado, el rendimiento de clasificación comparativa de este modelo reporta un valor de ABC de 0.687.^{104,171}

La Figura 22 muestra los compuestos con predicciones positivas para la prueba de hepatotoxicidad, siendo estos el 50% de la biblioteca CAS, el 46.48% de la biblioteca COCONUT, el 28.91% de la biblioteca FooDB, el 100% de la biblioteca QPNB y el 47.02% de la biblioteca DCM. Con base en los resultados es importante considerar que aunque hay una alta proporción de compuestos con resultados positivos a ser hepatotóxicos. Esto no necesariamente significa que estos compuestos deban ser descartados debido a que es probable que estos resultados puedan llegar a ser falsos positivos ya que de acuerdo con el valor del rendimiento que tiene el modelo empleado, la probabilidad de tener un resultado correcto es poco confiable.

5.7 Predicciones de aprendizaje de máquina y QSAR

A partir de los resultados obtenidos de las predicciones de aprendizaje automático y QSAR en colaboración con la Universidad de Carolina del Norte en Chapel Hill y el grupo *Collaborations Pharmaceuticals Inc.*, se obtuvieron 18 *hits* computacionales mostrados en la Figura 23.

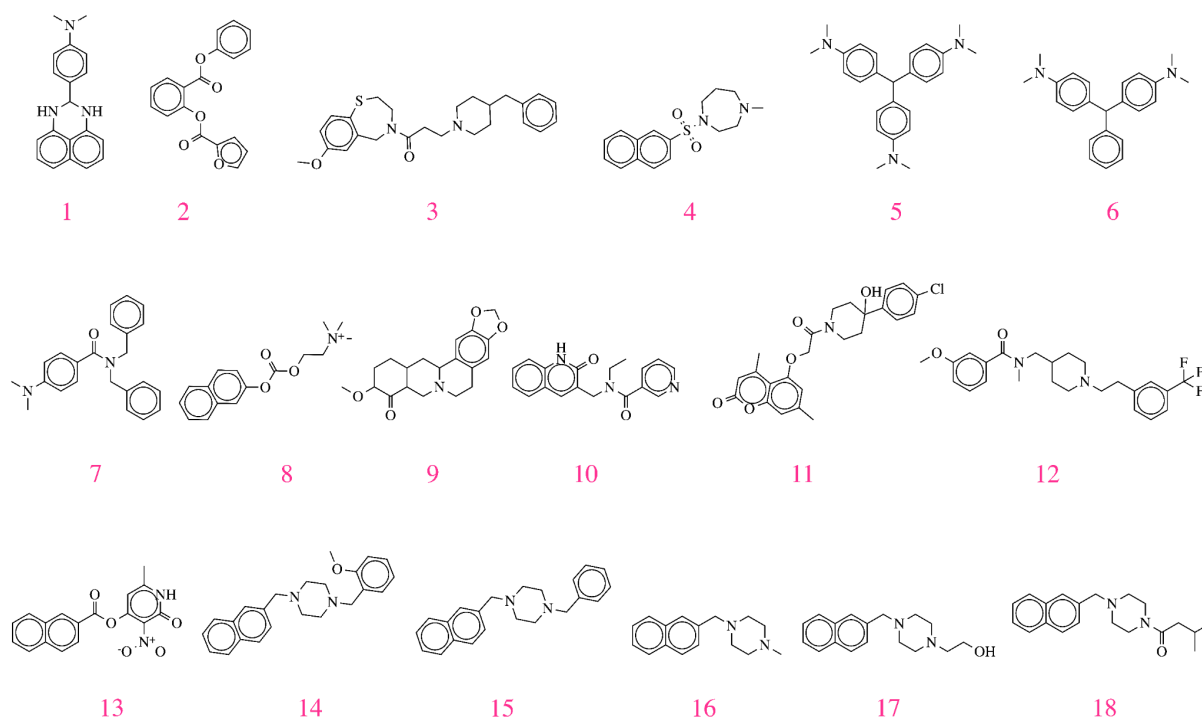


Figura 23. Estructuras químicas de los *hits* computacionales obtenidos del cribado virtual.

5.8 Acoplamiento molecular con la proteasa principal de SARS-CoV-2

Los acoplamiento molecular fueron realizados en el sitio donde interactúa el inhibidor MPI1 con la Mpro del SARS-CoV-2. La Tabla 6 resume los resultados de los valores de energía libre de Gibbs (ΔG) de unión obtenidos para los 18 *hits* computacionales y la Mpro del SARS-CoV-2 a partir de los tres programas empleados.

Tabla 6. Energía libre de Gibbs (ΔG) de unión obtenidos de los *hits* computacionales y la proteasa principal (Mpro) del coronavirus tipo 2 del Síndrome Respiratorio Agudo Grave (SARS-CoV-2), PDB ID: 7JPZ.

<i>Hit</i> computacional	ΔG de unión (kcal/mol)		
	MOE	LeDock	Autodock Vina
1	-6.08	-4.94	-7.60
2	-6.96	-5.61	-7.10

3	-7.84	-6.33	-7.20
4	-6.51	-4.61	-7.30
5	-6.98	-5.62	-6.70
6	-6.67	-5.54	-6.90
7	-6.98	-5.82	-7.40
8	-6.54	-4.55	-5.90
9	-6.81	-5.60	-7.40
10	-7.03	-6.67	-7.10
11	-7.46	-6.44	-8.30
12	-6.88	-5.72	-6.70
13	-6.77	-6.79	-7.80
14	-6.95	-4.87	-7.00
15	-6.26	-4.78	-7.10
16	-6.19	-4.46	-6.70
17	-6.31	-5.16	-6.20
18	-6.71	-5.25	-6.90

La finalidad de emplear tres diferentes programas para realizar los acoplamientos moleculares es el generar diferentes confórmers debido a que cada uno de estos programas utiliza un algoritmo diferente. Por lo tanto, se pretende obtener un resultado más confiable.¹⁷² De acuerdo con la Tabla 6, los compuestos con mejores ΔG de unión identificados para los tres programas empleados simultáneamente son los compuestos **2, 3, 7, 9, 10, 11 y 13**.

Las principales interacciones entre los 18 *hits* computacionales y la Mpro del SARS-CoV-2 se presentan con los residuos His 41, Leu 141, Gly 143 y Cys 145; por lo que se puede concluir que esta región es importante para que se puedan llevar a cabo la formación del complejo proteína-ligando, el cual en el caso de este proyecto tiene la finalidad de impedir la función de la proteína. Los diagramas en 2D de las interacciones de los *hits* computacionales y la Mpro del SARS-CoV-2 se encuentran en el apéndice.

5.9 Adquisición comercial de compuestos y envío a evaluación biológica

Todos los ensayos *in vitro* de la Mpro del SARS-CoV-2 se realizaron por duplicado en presencia y ausencia del agente reductor (DTT).

Como primera etapa de los ensayos de actividad enzimática de la Mpro del SARS-CoV-2, se evaluó la actividad de los *hits* a una sola concentración (100 μ M). Utilizando como control positivo el compuesto **GC376**, con un valor de concentración inhibitoria 50 (CI_{50}) de 19.64 nM. En esta primera etapa los *hits* **7** y **13** mostraron una reducción en la actividad de la Mpro. En los ensayos en los que se utilizó DTT, el compuesto **7** mostró tener una inhibición de la actividad del 24.13 ± 1.02 %, mientras que el compuesto **13** tuvo una inhibición de la actividad del 55.30 ± 1.22 %. A su vez, para los ensayos realizados en ausencia de agente reductor, el compuesto **7** tuvo una inhibición de la actividad del 51.88 ± 1.99 %, mientras que el compuesto **13** tuvo una inhibición de la actividad del 87.34 ± 0.04 %.

Posteriormente, se llevaron a cabo los ensayos de inhibición enzimática utilizando diez concentraciones de diluciones seriadas 1:3 de los compuestos **7** y **13** comenzando con una concentración de 100 μ M, para así obtener los valores CI_{50} . En ausencia del agente reductor el compuesto **7** presentó un valor de CI_{50} de 106.0 ± 13.45 μ M, mientras que el compuesto **13** tuvo un valor de CI_{50} de 3.554 ± 0.216 μ M. Por otro lado, en presencia de agente reductor, únicamente el compuesto **13** presentó inhibición de la Mpro con un valor de un CI_{50} de 106.4 ± 9.256 μ M.

En los ensayos enzimáticos de cisteínas proteasas es común emplear algún agente reductor como el DTT o glutatión con la finalidad de tener la certeza de que la enzima de interés, en este caso la Mpro se encuentre en su forma activa. Esto implica que la cisteína catalítica se encuentre en su forma reducida. A su vez; el agente reductor tiene la función de evitar que los compuestos pueden reaccionar con grupos tiol inespecíficos, los cuales pueden ser capaces de modificar covalentemente la cisteína catalítica.¹⁷³ Con base en lo anterior, se espera que los inhibidores específicos de la Mpro del SARS-CoV-2 sean capaces de inhibir la actividad de esta proteína en presencia y ausencia de DTT. A partir de esto, se puede concluir que solo el compuesto **13** muestra una

inhibición selectiva débil hacia la Mpro, mientras que los resultados obtenidos en los ensayos en ausencia de DTT ($CI_{50} = 3.554 \pm 0.216 \mu M$) pueden deberse a la posible alquilación u oxidación de la cisteína catalítica. Por otra parte, tomando en cuenta que la estructura del compuesto **13** carece de grupos tiol es posible saber que este no es capaz de formar puentes disulfuro con la cisteína catalítica. Por lo que se sabe que el resultado obtenido de la inhibición enzimática que se produce en presencia de DTT no se lleva a cabo debido al proceso oxidación.

Con base en los resultados de la Tabla 6, el compuesto **13** fue el que mostró el mejor valor de ΔG de unión con el programa *LeDock*. Así mismo, con los programas *MOE* y *Autodock Vina* mostró valores favorecidos sobre el resto de los compuestos (valores de ΔG de unión menores al promedio). La Figura 24 muestra las posibles principales interacciones entre el compuesto **13** y la Mpro del SARS-CoV-2. Algunas de estas son puentes de hidrógeno entre el grupo amino del anillo de hidroxipiridina y la cadena principal de la Leu141, entre uno de los oxígenos del grupo nitro y la cadena lateral de la Cys145 y entre el oxígeno del anillo de hidroxipiridina y la cadena principal de la Gly 143. Debido a lo anterior se puede concluir que los residuos 141-143 son parte del sitio catalítico de esta proteína y por lo tanto importantes para la actividad de esta. Así mismo, esto soporta que es importante el uso de algún agente reductor en los ensayos *in vitro*, debido a la presencia de un residuo de Cys en esta región.

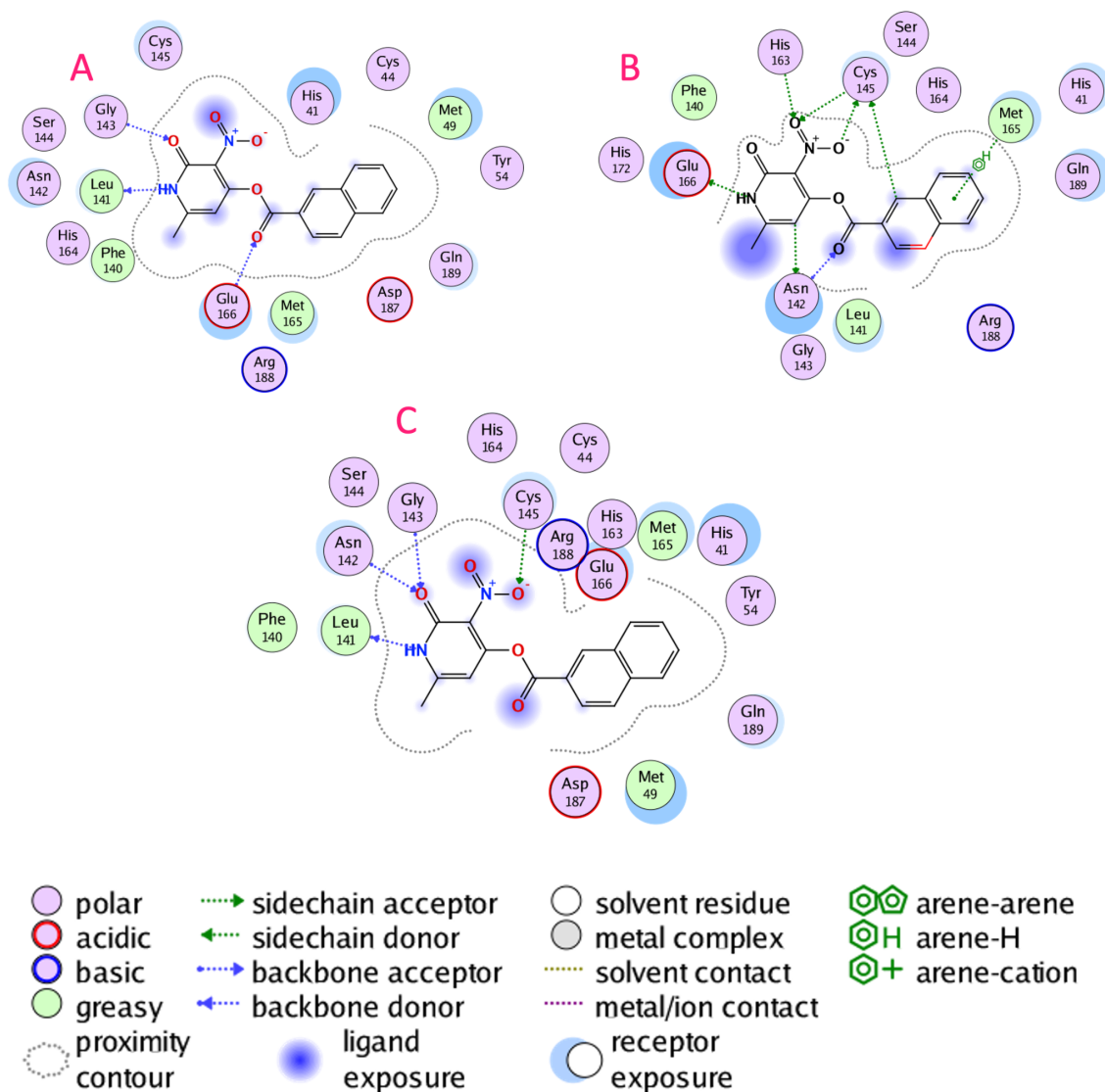


Figura 24. Interacciones predichas entre el *hit* computacional **13** y la proteasa principal (Mpro) del coronavirus tipo 2 del Síndrome Respiratorio Agudo Grave (SARS-CoV-2) en los diferentes programas computacionales A) MOE, B) LeDock y C) Autodock Vina.

En la Tabla 7 se muestra el perfil ADME/Tox del compuesto **13**. La tabla indica que las predicciones *in silico* indican que es altamente probable que este compuesto tenga una buena biodisponibilidad oral debido a los valores obtenidos de LogS, LogP, HIA; además de no presentar violaciones las reglas de Lipinski y Veber. También es posible que este compuesto pueda ser un inhibidor del CYP1A2, del CYP2C19 y del CYP2C9. Debido a que se sabe que la inhibición del CYP1A2 reduce el riesgo de desarrollar cáncer, esto no representa una desventaja para este compuesto. Por otra parte, el CYP2C19 tiene como función el metabolizar algunos

fármacos, por lo cual es importante evaluar experimentalmente si es capaz de metabolizar al compuesto **13**. Aunque el CYP2C9 se encuentra asociado a reacciones de fase 1 o de oxidación, en el caso del compuesto 13 es poco probable que tenga actividad debido a que este compuesto no cuenta con grupos funcionales susceptibles a reacciones de oxidación. Por último, este compuesto obtuvo resultados positivos para la prueba *in silico* de AMES y la hepatotoxicidad; por lo cual, es importante evaluar experimentalmente si efectivamente es capaz de causar mutagenicidad y daño hepático. Pues de ser este el caso es importante contemplar las posibles modificaciones estructurales que pudieran ayudar a mejorar su perfil ADME/Tox sin perder su actividad.

Tabla 7. Perfil ADME/Tox del compuesto **13**.

Descriptor molecular	Resultado	Descriptor molecular	Resultado
Solubilidad (<i>Silicos IT LogSw</i>)	-5.56	Inhibición del citocromo 1A2 (CYP1A2)	Sí
Lipofilicidad (<i>Consensus LogP</i>)	2.35	Inhibición del citocromo 2C19 (CYP2C19)	Sí
Absorción intestinal humana (<i>HIA</i>)	89.21%	Inhibición del citocromo 2C9 (CYP2C9)	Sí
Violaciones a las reglas de Lipinski	Ninguna	Inhibición del citocromo 2D6 (CYP2D6)	No
Violaciones a las reglas de Veber	Ninguna	Inhibición del citocromo 3A4 (CYP3A4)	No
Permeabilidad de la barrera hematoencefálica (<i>BBB permeability</i>)	-0.62	Eliminación total	0.755
Sustrato de la glicoproteína P	No	Prueba <i>in silico</i> del ensayo de AMES	Sí
Inhibición de la glicoproteína P	No	Hepatotoxicidad	Sí

6. Conclusiones

1. Se construyó y curó una base de datos molecular extensa, denominada *MEGA Db*, con 17, 802, 481 millones de compuestos.
2. Se filtró la *MEGA Db* con cálculos de similitud molecular usando tres representaciones moleculares (ECFP4, ECFP6 y MACCSKeys) y el coeficiente de Tanimoto.
3. Se obtuvo una lista de *hits* consenso a partir de los datos obtenidos en el análisis de similitud molecular y se verificó su disponibilidad comercial.
4. Se realizaron predicciones de aprendizaje automático y QSAR a los compuestos disponibles comercialmente.
5. Se obtuvieron 18 *hits* considerados como candidatos computacionales.
6. Se realizaron acoplamientos moleculares de los *hits* seleccionados con la estructura cristalográfica de la Mpro del SARS-CoV-2 (PDB ID: 6LU7) en tres programas computacionales: MOE, LeDock y Autodock Vina.
7. Se compraron los *hits* obtenidos y se llevaron a cabo evaluaciones enzimáticas *in vitro* de la Mpro del SARS-CoV-2 para evaluar la actividad inhibitoria de estos.
8. Se determinó que el compuesto **13** es un inhibidor de la Mpro del SARS-CoV-2 con un valor de CI_{50} de $106.4 \pm 9.256 \mu\text{M}$ en presencia de un agente reductor (DTT) y con un valor de $CI_{50} = 3.554 \pm 0.216 \mu\text{M}$ en ausencia de agentes reductores.

7. Perspectivas

Adquirir compuestos análogos al compuesto **13** y llevar a cabo la evaluación experimental a partir de pruebas enzimáticas *in vitro* con la Mpro del SARS-CoV-2 con el objetivo de:

- Obtener mejores valores de inhibición CI_{50} .
- Realizar un análisis cuantitativo de relaciones estructura-actividad .

8. Referencias

- (1) Koonin, E. V.; Dolja, V. V.; Krupovic, M.; Varsani, A.; Wolf, Y. I.; Yutin, N.; Zerbini, F. M.; Kuhn, J. H. Global Organization and Proposed Megataxonomy of the Virus World. *Microbiol. Mol. Biol. Rev.* **84** (2), e00061-19. <https://doi.org/10.1128/MMBR.00061-19>.
- (2) Jin, Y.; Yang, H.; Ji, W.; Wu, W.; Chen, S.; Zhang, W.; Duan, G. Virology, Epidemiology, Pathogenesis, and Control of COVID-19. *Viruses* **2020**, *12* (4), 372. <https://doi.org/10.3390/v12040372>.
- (3) Hasöksüz, M.; Kiliç, S.; Saraç, F. Coronaviruses and SARS-COV-2. *Turk J Med Sci* **8**.
- (4) Palacios Cruz, M.; Santos, E.; Velázquez Cervantes, M. A.; León Juárez, M. COVID-19, una emergencia de salud pública mundial. *Rev. Clínica Esp.* **2021**, *221* (1), 55–61. <https://doi.org/10.1016/j.rce.2020.03.001>.
- (5) Mousavizadeh, L.; Ghasemi, S. Genotype and Phenotype of COVID-19: Their Roles in Pathogenesis. *J. Microbiol. Immunol. Infect.* **2021**, *54* (2), 159–163. <https://doi.org/10.1016/j.jmii.2020.03.022>.
- (6) Lauxmann, M. A.; Santucci, N. E.; Aufrán-Gómez, A. M. The SARS-CoV-2 Coronavirus and the COVID-19 Outbreak. *Int. Braz. J. Urol.* **2020**, *46*, 6–18. <https://doi.org/10.1590/S1677-5538.IBJU.2020.S101>.
- (7) Huang, C.; Wang, Y.; Li, X.; Ren, L.; Zhao, J.; Hu, Y.; Zhang, L.; Fan, G.; Xu, J.; Gu, X.; Cheng, Z.; Yu, T.; Xia, J.; Wei, Y.; Wu, W.; Xie, X.; Yin, W.; Li, H.; Liu, M.; Xiao, Y.; Gao, H.; Guo, L.; Xie, J.; Wang, G.; Jiang, R.; Gao, Z.; Jin, Q.; Wang, J.; Cao, B. Clinical Features of Patients Infected with 2019 Novel Coronavirus in Wuhan, China. *The Lancet* **2020**, *395* (10223), 497–506. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5).
- (8) Gorbalenya, A. E.; Baker, S. C.; Baric, R. S.; de Groot, R. J.; Drosten, C.; Gulyaeva, A. A.; Haagmans, B. L.; Lauber, C.; Leontovich, A. M.; Neuman, B. W.; Penzar, D.; Perlman, S.; Poon, L. L. M.; Samborskiy, D. V.; Sidorov, I. A.; Sola, I.; Ziebuhr, J.; Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The Species Severe Acute Respiratory Syndrome-Related Coronavirus: Classifying 2019-NCoV and Naming It SARS-CoV-2. *Nat. Microbiol.* **2020**, *5* (4), 536–544. <https://doi.org/10.1038/s41564-020-0695-z>.
- (9) Lupia, T.; Scabini, S.; Mornese Pinna, S.; Di Perri, G.; De Rosa, F. G.; Corcione,

- S. 2019 Novel Coronavirus (2019-NCoV) Outbreak: A New Challenge. *J. Glob. Antimicrob. Resist.* **2020**, *21*, 22–27. <https://doi.org/10.1016/j.jgar.2020.02.021>.
- (10) Singhal, T. A Review of Coronavirus Disease-2019 (COVID-19). *Indian J. Pediatr.* **2020**, *87* (4), 281–286. <https://doi.org/10.1007/s12098-020-03263-6>.
- (11) Hemmati, F.; Saedi, S.; Hemmati-Dinarvand, M.; Zarei, M.; Seghatoleslam, A. Mysterious Virus: A Review on Behavior and Treatment Approaches of the Novel Coronavirus, 2019-NCoV. *Arch. Med. Res.* **2020**, *51* (5), 375–383. <https://doi.org/10.1016/j.arcmed.2020.04.022>.
- (12) Home - Johns Hopkins Coronavirus Resource Center <https://coronavirus.jhu.edu/> (accessed 2021 -08 -13).
- (13) Li, G.; De Clercq, E. Therapeutic Options for the 2019 Novel Coronavirus (2019-NCoV). *Nat. Rev. Drug Discov.* **2020**, *19* (3), 149–150. <https://doi.org/10.1038/d41573-020-00016-0>.
- (14) Wu, A.; Peng, Y.; Huang, B.; Ding, X.; Wang, X.; Niu, P.; Meng, J.; Zhu, Z.; Zhang, Z.; Wang, J.; Sheng, J.; Quan, L.; Xia, Z.; Tan, W.; Cheng, G.; Jiang, T. Genome Composition and Divergence of the Novel Coronavirus (2019-NCoV) Originating in China. *Cell Host Microbe* **2020**, *27* (3), 325–328. <https://doi.org/10.1016/j.chom.2020.02.001>.
- (15) Malik, Y. S.; Sircar, S.; Bhat, S.; Sharun, K.; Dhama, K.; Dadar, M.; Tiwari, R.; Chaicumpa, W. Emerging Novel Coronavirus (2019-NCoV)—Current Scenario, Evolutionary Perspective Based on Genome Analysis and Recent Developments. *Vet. Q.* **2020**, *40* (1), 68–76. <https://doi.org/10.1080/01652176.2020.1727993>.
- (16) Pillay, T. S. Gene of the Month: The 2019-NCoV/SARS-CoV-2 Novel Coronavirus Spike Protein. *J. Clin. Pathol.* **2020**, *73* (7), 366–369. <https://doi.org/10.1136/jclinpath-2020-206658>.
- (17) Chang, C.; Sue, S.-C.; Yu, T.; Hsieh, C.-M.; Tsai, C.-K.; Chiang, Y.-C.; Lee, S.; Hsiao, H.; Wu, W.-J.; Chang, W.-L.; Lin, C.-H.; Huang, T. Modular Organization of SARS Coronavirus Nucleocapsid Protein. *J. Biomed. Sci.* **2006**, *13* (1), 59–72. <https://doi.org/10.1007/s11373-005-9035-9>.
- (18) Tok, T. T.; Tatar, G. Structures and Functions of Coronavirus Proteins: Molecular Modeling of Viral Nucleoprotein. *Int. J. Virol.* **2** (1), 7.
- (19) Neuman, B. W.; Buchmeier, M. J. Chapter One - Supramolecular Architecture of the Coronavirus Particle. In *Advances in Virus Research*; Ziebuhr, J., Ed.;

- Coronaviruses; Academic Press, 2016; Vol. 96, pp 1–27. <https://doi.org/10.1016/bs.aivir.2016.08.005>.
- (20) Snijder, E. J.; Decroly, E.; Ziebuhr, J. Chapter Three - The Nonstructural Proteins Directing Coronavirus RNA Synthesis and Processing. In *Advances in Virus Research*; Ziebuhr, J., Ed.; Coronaviruses; Academic Press, 2016; Vol. 96, pp 59–126. <https://doi.org/10.1016/bs.aivir.2016.08.008>.
- (21) Muramatsu, T.; Takemoto, C.; Kim, Y.-T.; Wang, H.; Nishii, W.; Terada, T.; Shirouzu, M.; Yokoyama, S. SARS-CoV 3CL Protease Cleaves Its C-Terminal Autoprocessing Site by Novel Subsite Cooperativity. *Proc. Natl. Acad. Sci.* **2016**, *113* (46), 12997–13002. <https://doi.org/10.1073/pnas.1601327113>.
- (22) Kim, C.-H. SARS-CoV-2 Evolutionary Adaptation toward Host Entry and Recognition of Receptor O-Acetyl Sialylation in Virus–Host Interaction. *Int. J. Mol. Sci.* **2020**, *21* (12), 4549. <https://doi.org/10.3390/ijms21124549>.
- (23) Al-Horani, R. A.; Kar, S.; Aliter, K. F. Potential Anti-COVID-19 Therapeutics That Block the Early Stage of the Viral Life Cycle: Structures, Mechanisms, and Clinical Trials. *Int. J. Mol. Sci.* **2020**, *21* (15), 5224. <https://doi.org/10.3390/ijms21155224>.
- (24) Guo, Y.-R.; Cao, Q.-D.; Hong, Z.-S.; Tan, Y.-Y.; Chen, S.-D.; Jin, H.-J.; Tan, K.-S.; Wang, D.-Y.; Yan, Y. The Origin, Transmission and Clinical Therapies on Coronavirus Disease 2019 (COVID-19) Outbreak – an Update on the Status. *Mil. Med. Res.* **2020**, *7* (1), 11. <https://doi.org/10.1186/s40779-020-00240-0>.
- (25) Bonny, V.; Maillard, A.; Mousseaux, C.; Plaçais, L.; Richier, Q. COVID-19 : physiopathologie d'une maladie à plusieurs visages. *Rev. Médecine Interne* **2020**, *41* (6), 375–389. <https://doi.org/10.1016/j.revmed.2020.05.003>.
- (26) Tian, S.; Hu, N.; Lou, J.; Chen, K.; Kang, X.; Xiang, Z.; Chen, H.; Wang, D.; Liu, N.; Liu, D.; Chen, G.; Zhang, Y.; Li, D.; Li, J.; Lian, H.; Niu, S.; Zhang, L.; Zhang, J. Characteristics of COVID-19 Infection in Beijing. *J. Infect.* **2020**, *80* (4), 401–406. <https://doi.org/10.1016/j.jinf.2020.02.018>.
- (27) Lei, S.; Jiang, F.; Su, W.; Chen, C.; Chen, J.; Mei, W.; Zhan, L.-Y.; Jia, Y.; Zhang, L.; Liu, D.; Xia, Z.-Y.; Xia, Z. Clinical Characteristics and Outcomes of Patients Undergoing Surgeries during the Incubation Period of COVID-19 Infection. *EClinicalMedicine* **2020**, *21*. <https://doi.org/10.1016/j.eclinm.2020.100331>.
- (28) Lauer, S. A.; Grantz, K. H.; Bi, Q.; Jones, F. K.; Zheng, Q.; Meredith, H. R.;

- Azman, A. S.; Reich, N. G.; Lessler, J. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Ann. Intern. Med.* **2020**, *172* (9), 577–582. <https://doi.org/10.7326/M20-0504>.
- (29) Fung, T. S.; Liu, D. X. Human Coronavirus: Host-Pathogen Interaction. *Annu. Rev. Microbiol.* **2019**, *73* (1), 529–557. <https://doi.org/10.1146/annurev-micro-020518-115759>.
- (30) García, L. F. Immune Response, Inflammation, and the Clinical Spectrum of COVID-19. *Front. Immunol.* **2020**, *11*, 1441. <https://doi.org/10.3389/fimmu.2020.01441>.
- (31) Hur, S. Double-Stranded RNA Sensors and Modulators in Innate Immunity. *Annu. Rev. Immunol.* **2019**, *37* (1), 349–375. <https://doi.org/10.1146/annurev-immunol-042718-041356>.
- (32) Chu, H.; Chan, J. F.-W.; Wang, Y.; Yuen, T. T.-T.; Chai, Y.; Hou, Y.; Shuai, H.; Yang, D.; Hu, B.; Huang, X.; Zhang, X.; Cai, J.-P.; Zhou, J.; Yuan, S.; Kok, K.-H.; To, K. K.-W.; Chan, I. H.-Y.; Zhang, A. J.; Sit, K.-Y.; Au, W.-K.; Yuen, K.-Y. Comparative Replication and Immune Activation Profiles of SARS-CoV-2 and SARS-CoV in Human Lungs: An Ex Vivo Study With Implications for the Pathogenesis of COVID-19. *Clin. Infect. Dis.* **2020**, *71* (6), 1400–1409. <https://doi.org/10.1093/cid/ciaa410>.
- (33) Blanco-Melo, D.; Nilsson-Payant, B. E.; Liu, W.-C.; Uhl, S.; Hoagland, D.; Møller, R.; Jordan, T. X.; Oishi, K.; Panis, M.; Sachs, D.; Wang, T. T.; Schwartz, R. E.; Lim, J. K.; Albrecht, R. A.; tenOever, B. R. Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell* **2020**, *181* (5), 1036-1045.e9. <https://doi.org/10.1016/j.cell.2020.04.026>.
- (34) Zhou, M.-Y.; Xie, X.-L.; Peng, Y.-G.; Wu, M.-J.; Deng, X.-Z.; Wu, Y.; Xiong, L.-J.; Shang, L.-H. From SARS to COVID-19: What We Have Learned about Children Infected with COVID-19. *Int. J. Infect. Dis.* **2020**, *96*, 710–714. <https://doi.org/10.1016/j.ijid.2020.04.090>.
- (35) Tai, W.; He, L.; Zhang, X.; Pu, J.; Voronin, D.; Jiang, S.; Zhou, Y.; Du, L. Characterization of the Receptor-Binding Domain (RBD) of 2019 Novel Coronavirus: Implication for Development of RBD Protein as a Viral Attachment Inhibitor and Vaccine. *Cell. Mol. Immunol.* **2020**, *17* (6), 613–620. <https://doi.org/10.1038/s41423-020-0400-4>.

- (36) Ou, X.; Liu, Y.; Lei, X.; Li, P.; Mi, D.; Ren, L.; Guo, L.; Guo, R.; Chen, T.; Hu, J.; Xiang, Z.; Mu, Z.; Chen, X.; Chen, J.; Hu, K.; Jin, Q.; Wang, J.; Qian, Z. Characterization of Spike Glycoprotein of SARS-CoV-2 on Virus Entry and Its Immune Cross-Reactivity with SARS-CoV. *Nat. Commun.* **2020**, *11* (1), 1620. <https://doi.org/10.1038/s41467-020-15562-9>.
- (37) Huang, A. T.; Garcia-Carreras, B.; Hitchings, M. D. T.; Yang, B.; Katzelnick, L. C.; Rattigan, S. M.; Borgert, B. A.; Moreno, C. A.; Solomon, B. D.; Rodriguez-Barraquer, I.; Lessler, J.; Salje, H.; Burke, D.; Wesolowski, A.; Cummings, D. A. T. *A Systematic Review of Antibody Mediated Immunity to Coronaviruses: Antibody Kinetics, Correlates of Protection, and Association of Antibody Responses with Severity of Disease*; 2020; p 2020.04.14.20065771. <https://doi.org/10.1101/2020.04.14.20065771>.
- (38) Li, G.; Fan, Y.; Lai, Y.; Han, T.; Li, Z.; Zhou, P.; Pan, P.; Wang, W.; Hu, D.; Liu, X.; Zhang, Q.; Wu, J. Coronavirus Infections and Immune Responses. *J. Med. Virol.* **2020**, *92* (4), 424–432. <https://doi.org/10.1002/jmv.25685>.
- (39) Tay, M. Z.; Poh, C. M.; Rénia, L.; MacAry, P. A.; Ng, L. F. P. The Trinity of COVID-19: Immunity, Inflammation and Intervention. *Nat. Rev. Immunol.* **2020**, *20* (6), 363–374. <https://doi.org/10.1038/s41577-020-0311-8>.
- (40) Clay, C.; Donart, N.; Fomukong, N.; Knight, J. B.; Lei, W.; Price, L.; Hahn, F.; Van Westrienen, J.; Harrod, K. S. Primary Severe Acute Respiratory Syndrome Coronavirus Infection Limits Replication but Not Lung Inflammation upon Homologous Rechallenge. *J. Virol.* **2012**, *86* (8), 4234–4244. <https://doi.org/10.1128/JVI.06791-11>.
- (41) Marzi, A.; Gramberg, T.; Simmons, G.; Möller, P.; Rennekamp, A. J.; Krumbiegel, M.; Geier, M.; Eisemann, J.; Turza, N.; Saunier, B.; Steinkasserer, A.; Becker, S.; Bates, P.; Hofmann, H.; Pöhlmann, S. DC-SIGN and DC-SIGNR Interact with the Glycoprotein of Marburg Virus and the S Protein of Severe Acute Respiratory Syndrome Coronavirus. *J. Virol.* **2004**, *78* (21), 12090–12095. <https://doi.org/10.1128/JVI.78.21.12090-12095.2004>.
- (42) Yang, Z.-Y.; Huang, Y.; Ganesh, L.; Leung, K.; Kong, W.-P.; Schwartz, O.; Subbarao, K.; Nabel, G. J. PH-Dependent Entry of Severe Acute Respiratory Syndrome Coronavirus Is Mediated by the Spike Glycoprotein and Enhanced by Dendritic Cell Transfer through DC-SIGN. *J. Virol.* **2004**, *78* (11), 5642–5650. <https://doi.org/10.1128/JVI.78.11.5642-5650.2004>.

- (43) Vankadari, N.; Wilce, J. A. Emerging COVID-19 Coronavirus: Glycan Shield and Structure Prediction of Spike Glycoprotein and Its Interaction with Human CD26. *Emerg. Microbes Infect.* **2020**, *9* (1), 601–604. <https://doi.org/10.1080/22221751.2020.1739565>.
- (44) Wang, K.; Chen, W.; Zhou, Y.-S.; Lian, J.-Q.; Zhang, Z.; Du, P.; Gong, L.; Zhang, Y.; Cui, H.-Y.; Geng, J.-J.; Wang, B.; Sun, X.-X.; Wang, C.-F.; Yang, X.; Lin, P.; Deng, Y.-Q.; Wei, D.; Yang, X.-M.; Zhu, Y.-M.; Zhang, K.; Zheng, Z.-H.; Miao, J.-L.; Guo, T.; Shi, Y.; Zhang, J.; Fu, L.; Wang, Q.-Y.; Bian, H.; Zhu, P.; Chen, Z.-N. *SARS-CoV-2 Invades Host Cells via a Novel Route: CD147-Spike Protein*; 2020; p 2020.03.14.988345. <https://doi.org/10.1101/2020.03.14.988345>.
- (45) Huang, X.; Wei, F.; Hu, L.; Wen, L.; Chen, K. Epidemiology and Clinical Characteristics of COVID-19. *Arch. Iran. Med.* **2020**, *23* (4), 268–271. <https://doi.org/10.34172/aim.2020.09>.
- (46) Mehraeen, E.; Karimi, A.; Barzegary, A.; Vahedi, F.; Afsahi, A. M.; Dadras, O.; Moradmand-Badie, B.; Seyed Alinaghi, S. A.; Jahanfar, S. Predictors of Mortality in Patients with COVID-19—a Systematic Review. *Eur. J. Integr. Med.* **2020**, *40*, 101226. <https://doi.org/10.1016/j.eujim.2020.101226>.
- (47) Jutzeler, C. R.; Bourguignon, L.; Weis, C. V.; Tong, B.; Wong, C.; Rieck, B.; Pargger, H.; Tschudin-Sutter, S.; Egli, A.; Borgwardt, K.; Walter, M. Comorbidities, Clinical Signs and Symptoms, Laboratory Findings, Imaging Features, Treatment Strategies, and Outcomes in Adult and Pediatric Patients with COVID-19: A Systematic Review and Meta-Analysis. *Travel Med. Infect. Dis.* **2020**, *37*, 101825. <https://doi.org/10.1016/j.tmaid.2020.101825>.
- (48) Hemmer, C. J.; Geerdes-Fenge, H. F.; Reisinger, E. C. COVID-19: Epidemiologische und klinische Fakten. *Radiol.* **2020**, *60* (10), 893–898. <https://doi.org/10.1007/s00117-020-00741-y>.
- (49) Khan, E. COVID-19 in Children: Epidemiology, Presentation, Diagnosis and Management. *J. Pak. Med. Assoc.* **2020**, No. 0, 1. <https://doi.org/10.5455/JPMA.25>.
- (50) Dhakal, B. P.; Sweitzer, N. K.; Indik, J. H.; Acharya, D.; William, P. SARS-CoV-2 Infection and Cardiovascular Disease: COVID-19 Heart. *Heart Lung Circ.* **2020**, *29* (7), 973–987. <https://doi.org/10.1016/j.hlc.2020.05.101>.
- (51) Siddiqi, H. K.; Mehra, M. R. COVID-19 Illness in Native and Immunosuppressed

- States: A Clinical–Therapeutic Staging Proposal. *J. Heart Lung Transplant.* **2020**, *39* (5), 405–407. <https://doi.org/10.1016/j.healun.2020.03.012>.
- (52) Stawicki, S. P.; Jeanmonod, R.; Miller, A. C.; Paladino, L.; Gaieski, D. F.; Yaffee, A. Q.; Wulf, A. D.; Grover, J.; Papadimos, T. J.; Bloem, C.; Galwankar, S. C.; Chauhan, V.; Firstenberg, M. S.; Somma, S. D.; Jeanmonod, D.; Garg, S. M.; Tucci, V.; Anderson, H. L.; Fatimah, L.; Worlton, T. J.; Dubhashi, S. P.; Glaze, K. S.; Sinha, S.; Opara, I. N.; Yellapu, V.; Kelkar, D.; El-Menyar, A.; Krishnan, V.; Venkataramanaiah, S.; Leyfman, Y.; Thani, H. A. S. A.; Nanayakkara, P. W. B.; Nanda, S.; Cioè-Peña, E.; Sardesai, I.; Chandra, S.; Munasinghe, A.; Dutta, V.; Ponte, S. T. D.; Izurieta, R.; Asensio, J. A.; Garg, M. The 2019–2020 Novel Coronavirus (Severe Acute Respiratory Syndrome Coronavirus 2) Pandemic: A Joint American College of Academic International Medicine-World Academic Council of Emergency Medicine Multidisciplinary COVID-19 Working Group Consensus Paper. *J. Glob. Infect. Dis.* **2020**, *12* (2), 47. https://doi.org/10.4103/jgid.jgid_86_20.
- (53) Struyf, T.; Deeks, J. J.; Dinnes, J.; Takwoingi, Y.; Davenport, C.; Leeflang, M. M.; Spijker, R.; Hooft, L.; Emperador, D.; Dittrich, S.; Domen, J.; Horn, S. R. A.; Bruel, A. V. den; Group, C. C.-19 D. T. A. Signs and Symptoms to Determine If a Patient Presenting in Primary Care or Hospital Outpatient Settings Has COVID-19 Disease. *Cochrane Database Syst. Rev.* **2020**, No. 7. <https://doi.org/10.1002/14651858.CD013665>.
- (54) Dhama, K.; Khan, S.; Tiwari, R.; Sircar, S.; Bhat, S.; Malik, Y. S.; Singh, K. P.; Chaicumpa, W.; Bonilla-Aldana, D. K.; Rodriguez-Morales, A. J. Coronavirus Disease 2019–COVID-19. *Clin. Microbiol. Rev.* **33** (4), e00028-20. <https://doi.org/10.1128/CMR.00028-20>.
- (55) Larsen, J. R.; Martin, M. R.; Martin, J. D.; Kuhn, P.; Hicks, J. B. Modeling the Onset of Symptoms of COVID-19. *Front. Public Health* **2020**, *8*, 473. <https://doi.org/10.3389/fpubh.2020.00473>.
- (56) Response, E. W. G. for N. E.; Prevention, C. C. for D. C. and. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China. *Chin. J. Epidemiol.* **2020**, *41* (02), 145–151. <https://doi.org/10.3760/cma.j.issn.0254-6450.2020.02.003>.
- (57) Tang, Y.-W.; Schmitz, J. E.; Persing, D. H.; Stratton, C. W. Laboratory Diagnosis of COVID-19: Current Issues and Challenges. *J. Clin. Microbiol.* **58** (6),

- e00512-20. <https://doi.org/10.1128/JCM.00512-20>.
- (58) Oliveira, B. A.; Oliveira, L. C. de; Sabino, E. C.; Okay, T. S. SARS-CoV-2 and the COVID-19 Disease: A Mini Review on Diagnostic Methods. *Rev. Inst. Med. Trop. São Paulo* **2020**, *62*. <https://doi.org/10.1590/S1678-9946202062044>.
- (59) Udugama, B.; Kadhiresan, P.; Kozlowski, H. N.; Malekjahani, A.; Osborne, M.; Li, V. Y. C.; Chen, H.; Mubareka, S.; Gubbay, J. B.; Chan, W. C. W. Diagnosing COVID-19: The Disease and Tools for Detection. *ACS Nano* **2020**, *14* (4), 3822–3835. <https://doi.org/10.1021/acsnano.0c02624>.
- (60) Iyer, M.; Jayaramayya, K.; Subramaniam, M. D.; Lee, S. B.; Dayem, A. A.; Vellingiri, S.-G. C. & B. COVID-19: An Update on Diagnostic and Therapeutic Approaches. *BMB Rep.* **2020**, *53* (4), 191–205. <https://doi.org/10.5483/BMBRep.2020.53.4.080>.
- (61) Deeks, J. J.; Dinnes, J.; Takwoingi, Y.; Davenport, C.; Spijker, R.; Taylor-Phillips, S.; Adriano, A.; Beese, S.; Dretzke, J.; Ruffano, L. F. di; Harris, I. M.; Price, M. J.; Dittrich, S.; Emperador, D.; Hooft, L.; Leeftang, M. M.; Bruel, A. V. den; Group, C. C.-19 D. T. A. Antibody Tests for Identification of Current and Past Infection with SARS-CoV-2. *Cochrane Database Syst. Rev.* **2020**, No. 6. <https://doi.org/10.1002/14651858.CD013652>.
- (62) Alou, L. *Rev Esp Quimioter* 2020; 33(6): 466-484 – SEQ.
- (63) Pascarella, G.; Strumia, A.; Piliago, C.; Bruno, F.; Buono, R. D.; Costa, F.; Scarlata, S.; Agrò, F. E. COVID-19 Diagnosis and Management: A Comprehensive Review. *J. Intern. Med.* **2020**, *288* (2), 192–206. <https://doi.org/10.1111/joim.13091>.
- (64) Liu, C.; Zhou, Q.; Li, Y.; Garner, L. V.; Watkins, S. P.; Carter, L. J.; Smoot, J.; Gregg, A. C.; Daniels, A. D.; Jervey, S.; Albaiu, D. Research and Development on Therapeutic Agents and Vaccines for COVID-19 and Related Human Coronavirus Diseases. *ACS Cent. Sci.* **2020**, *6* (3), 315–331. <https://doi.org/10.1021/acscentsci.0c00272>.
- (65) Muralidharan, N.; Sakthivel, R.; Velmurugan, D.; Gromiha, M. M. Computational Studies of Drug Repurposing and Synergism of Lopinavir, Oseltamivir and Ritonavir Binding with SARS-CoV-2 Protease against COVID-19. *J. Biomol. Struct. Dyn.* **2021**, *39* (7), 2673–2678. <https://doi.org/10.1080/07391102.2020.1752802>.
- (66) Cong, Y.; Hart, B. J.; Gross, R.; Zhou, H.; Frieman, M.; Bollinger, L.; Wada, J.;

- Hensley, L. E.; Jahrling, P. B.; Dyal, J.; Holbrook, M. R. MERS-CoV Pathogenesis and Antiviral Efficacy of Licensed Drugs in Human Monocyte-Derived Antigen-Presenting Cells. *PLOS ONE* **2018**, *13* (3), e0194868. <https://doi.org/10.1371/journal.pone.0194868>.
- (67) De Clercq, E. Potential Antivirals and Antiviral Strategies against SARS Coronavirus Infections. *Expert Rev. Anti Infect. Ther.* **2006**, *4* (2), 291–302. <https://doi.org/10.1586/14787210.4.2.291>.
- (68) Sheahan, T. P.; Sims, A. C.; Leist, S. R.; Schäfer, A.; Won, J.; Brown, A. J.; Montgomery, S. A.; Hogg, A.; Babusis, D.; Clarke, M. O.; Spahn, J. E.; Bauer, L.; Sellers, S.; Porter, D.; Feng, J. Y.; Cihlar, T.; Jordan, R.; Denison, M. R.; Baric, R. S. Comparative Therapeutic Efficacy of Remdesivir and Combination Lopinavir, Ritonavir, and Interferon Beta against MERS-CoV. *Nat. Commun.* **2020**, *11* (1), 222. <https://doi.org/10.1038/s41467-019-13940-6>.
- (69) Kadam, R. U.; Wilson, I. A. Structural Basis of Influenza Virus Fusion Inhibition by the Antiviral Drug Arbidol. *Proc. Natl. Acad. Sci.* **2017**, *114* (2), 206–214. <https://doi.org/10.1073/pnas.1617020114>.
- (70) Hoffmann, M.; Kleine-Weber, H.; Schroeder, S.; Krüger, N.; Herrler, T.; Erichsen, S.; Schiergens, T. S.; Herrler, G.; Wu, N.-H.; Nitsche, A.; Müller, M. A.; Drosten, C.; Pöhlmann, S. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **2020**, *181* (2), 271-280.e8. <https://doi.org/10.1016/j.cell.2020.02.052>.
- (71) Richardson, P.; Griffin, I.; Tucker, C.; Smith, D.; Oechsle, O.; Phelan, A.; Rawling, M.; Savory, E.; Stebbing, J. Baricitinib as Potential Treatment for 2019-NCoV Acute Respiratory Disease. *The Lancet* **2020**, *395* (10223), e30–e31. [https://doi.org/10.1016/S0140-6736\(20\)30304-4](https://doi.org/10.1016/S0140-6736(20)30304-4).
- (72) Zeldin, R. K.; Petruschke, R. A. Pharmacological and Therapeutic Properties of Ritonavir-Boosted Protease Inhibitor Therapy in HIV-Infected Patients. *J. Antimicrob. Chemother.* **2004**, *53* (1), 4–9. <https://doi.org/10.1093/jac/dkh029>.
- (73) Guo, D. Old Weapon for New Enemy: Drug Repurposing for Treatment of Newly Emerging Viral Diseases. *Viol. Sin.* **2020**, *35* (3), 253–255. <https://doi.org/10.1007/s12250-020-00204-7>.
- (74) Mifsud, E. J.; Hayden, F. G.; Hurt, A. C. Antivirals Targeting the Polymerase Complex of Influenza Viruses. *Antiviral Res.* **2019**, *169*, 104545. <https://doi.org/10.1016/j.antiviral.2019.104545>.

- (75) Wang, M.; Cao, R.; Zhang, L.; Yang, X.; Liu, J.; Xu, M.; Shi, Z.; Hu, Z.; Zhong, W.; Xiao, G. Remdesivir and Chloroquine Effectively Inhibit the Recently Emerged Novel Coronavirus (2019-NCoV) in Vitro. *Cell Res.* **2020**, *30* (3), 269–271. <https://doi.org/10.1038/s41422-020-0282-0>.
- (76) Li, Y.; Xie, Z.; Lin, W.; Cai, W.; Wen, C.; Guan, Y.; Mo, X.; Wang, J.; Wang, Y.; Peng, P.; Chen, X.; Hong, W.; Xiao, G.; Liu, J.; Zhang, L.; Hu, F.; Li, F.; Zhang, F.; Deng, X.; Li, L. Efficacy and Safety of Lopinavir/Ritonavir or Arbidol in Adult Patients with Mild/Moderate COVID-19: An Exploratory Randomized Controlled Trial. *Med* **2020**, *1* (1), 105-113.e4. <https://doi.org/10.1016/j.medj.2020.04.001>.
- (77) Possible Treatment and Strategies for COVID-19: Review and Assessment. *European Review*, 2020.
- (78) Li, Y.; Liu, S.; Zhang, S.; Ju, Q.; Zhang, S.; Yang, Y.; Wang, H. Current Treatment Approaches for COVID-19 and the Clinical Value of Transfusion-Related Technologies. *Transfus. Apher. Sci.* **2020**, *59* (5). <https://doi.org/10.1016/j.transci.2020.102839>.
- (79) Hung, I. F.; To, K. K.; Lee, C.-K.; Lee, K.-L.; Chan, K.; Yan, W.-W.; Liu, R.; Watt, C.-L.; Chan, W.-M.; Lai, K.-Y.; Koo, C.-K.; Buckley, T.; Chow, F.-L.; Wong, K.-K.; Chan, H.-S.; Ching, C.-K.; Tang, B. S.; Lau, C. C.; Li, I. W.; Liu, S.-H.; Chan, K.-H.; Lin, C.-K.; Yuen, K.-Y. Convalescent Plasma Treatment Reduced Mortality in Patients With Severe Pandemic Influenza A (H1N1) 2009 Virus Infection. *Clin. Infect. Dis.* **2011**, *52* (4), 447–456. <https://doi.org/10.1093/cid/ciq106>.
- (80) Kraft, C. S.; Hewlett, A. L.; Koepsell, S.; Winkler, A. M.; Kratochvil, C. J.; Larson, L.; Varkey, J. B.; Mehta, A. K.; Lyon, G. M., III; Friedman-Moraco, R. J.; Marconi, V. C.; Hill, C. E.; Sullivan, J. N.; Johnson, D. W.; Lisco, S. J.; Mulligan, M. J.; Uyeki, T. M.; McElroy, A. K.; Sealy, T.; Campbell, S.; Spiropoulou, C.; Ströher, U.; Crozier, I.; Sacra, R.; Connor, M. J., Jr; Sueblinvong, V.; Franch, H. A.; Smith, P. W.; Ribner, B. S.; for the Nebraska Biocontainment Unit and the Emory Serious Communicable Diseases Unit; Smith, P.; Hewlett, A.; Schwedhelm, S.; Boulter, K.; Beam, E.; Gibbs, S.; Lowe, J.; Kratochvil, C.; Sullivan, J.; Johnson, D.; Lisco, S.; Piquette, C.; Bailey, K.; Auxier, J.; Boer, B.; Hanson, T.; Kaseman, J.; Khan, M. S.; Rhee, J. H.; Wells, A.; Florescu, D.; Kalil, A.; Rupp, M.; Becker, V.; Boeckman, B.; Elder, E.; Fitch, A.; Flood, E.; Fremly, M.; Frigge, R.; Hanzlik, K. A.; Jensen, L.; Kraus, N.; Molacek, D.; Nevins, J.;

Parker, A.; Peters, J.; Rand, C.; Roesler, K.; Ryalls, K.; Shradar, M.; Strain, A.; Sunderman, T.; Sundermeier, J.; Swanhorst, J.; Vasa, A.; Bellinghausen, J.; Freihaut, F.; Denny, S.; Mainelli, L.; Pinkney, D.; Ray, D.; Jevne, J.; Knight, K.; McCroy, D.; Nadeau, R.; Nightser, A.; Iwen, P.; Sambol, T.; Herrera, V.; Morgan, D.; Trotter, S.; Kerby, A.; Peters, S.; Southern, T.; Murphy, C.; Morris, R.; Nuss, S.; Franco, T.; Ogden, C.; Lazure, J.; Straub, D.; Koepsell, S.; Hinrichs, S.; Plumb, T.; Florescu, M.; Becker, J.; Souchek, J.; Beck, J.; Larson, L.; Heires, P.; Gordon, B.; Paulsen, G.; Wolford, B.; Petersen, J.; Marion, N.; Hayes, K.; Tyner, K.; Wilson, T.; Baltes, P.; Nowatzke, J.; Nguyen, J.; Turner, P.; Boonstra, B.; Jelden, K.; Portrey, R.; Stringfield, D.; Scofield, B.; Svanda, G.; Horihan, J.; Dahl, C.; Bruno, M.; Malm, K.; Litz, R.; Fehring, J.; Paladino, K.; Opp, T.; Bell, S.; Adams, A.; Allen, M. B.; Bachman, R.; Bornstein, W.; Cantrell, D.; Cosper, P.; Feistritzer, N.; Fox, J.; Gartland, B.; Goodman, J.; Grant, S.; Howard-Crow, D.; Horowitz, I.; Pugh, D.; Ritenour, C.; Ash, T.; Barnes, C.; Calhoun, J.; Chapman, L.; Daye, T.; Durr, H.; Evans, S.; Gentry, J.; Ginnane, J.; Grant, S.; Haynes, C.; Hill, C.; Hillis, D.; Johnson, C.; Loomis, J.; Mamora, J.; Mitchell, L.; Morgan, J.; Osakwe, N.; Owen, J.; Piazza, S.; Shirley, K.; Siddens, J.; Silas, C.; Slabach, J.; Tirador, E.; Todd, D.; Vanairsdale, S.; Brammer, N.; Buchanan, J.; Burd, E.; Cardella, J.; Eaves, B.; Evans, C.; Hostetler, K.; Jenkins, K.; Lindsey, M.; Magee, J.; Powers, R.; Ritchie, J.; Ryan, E.; Bonds, S.; Emamifar, A.; Kuban, T.; Pack, J.; Rogers, S.; Golston, G.; Kaufman, S.; Olinger, P.; Olinger, S.; Rengarajan, K.; Thomaston, S.; Beck, E.; Desroches, P.; Hall, C.; Walker, C.; Bryant, C.; Hackman, B.; Howard, R.; Jones, M.; Broughton, J.; Frisle, B.; Jackson, R.; Lewis, J.; Brown-Haithco, R.; Gartin, M. L.; Gerald-Washington, E.; James-Jones, R.; Miller, D.; Stark, D.; McGee, G.; Jones, P.; Scott-Harris, L.; Cain, J.; Davis, R.; Johnson, T.; Pickett, T.; Shaw, A.; Truesdale, T.; Blum, J.; Hill, L.; Sevransky, J.; Zivot, J.; Walker, S.; Klopman, M.; Matkins, R.; Meehan, C.; Meehan, P.; Schwock, K.; Schuck, J.; Stack, K.; Wolf, F.; Bray, B.; Isakov, A.; Shartar, S.; Miles, W.; Jamison, A.; Arevalo, J.; Stallings, G.; Christenbury, J.; Dollard, V.; De Gennaro, M.; Korschun, H.; Seideman, N.; Ziegler, T.; Griffith, D. P.; Dave, N.; Wall, S.; Hall, M.; McGhee, D.; Clarke, T.; Vaught, R.; Peterson-Pileri, K.; for the Nebraska Biocontainment Unit and the Emory Serious Communicable Diseases Unit. The Use of TKM-100802 and Convalescent Plasma in 2 Patients With Ebola Virus Disease in the United

- States. *Clin. Infect. Dis.* **2015**, *61* (4), 496–502.
<https://doi.org/10.1093/cid/civ334>.
- (81) Ko, J.-H.; Seok, H.; Cho, S. Y.; Eun Ha, Y.; Baek, J. Y.; Kim, S. H.; Kim, Y. J.; Park, J. K.; Chung, C. R.; Kang, E.-S.; Cho, D.; Müller, M. A.; Drosten, C.; Kang, C.-I.; Chung, D. R.; Song, J.-H.; Peck, K. R. Challenges of Convalescent Plasma Infusion Therapy in Middle East Respiratory Coronavirus Infection: A Single Centre Experience. *Antivir. Ther.* **2018**, *23* (7), 617–622.
<https://doi.org/10.3851/IMP3243>.
- (82) Song, P.; Li, W.; Xie, J.; Hou, Y.; You, C. Cytokine Storm Induced by SARS-CoV-2. *Clin. Chim. Acta* **2020**, *509*, 280–287.
<https://doi.org/10.1016/j.cca.2020.06.017>.
- (83) Modjarrad, K.; Roberts, C. C.; Mills, K. T.; Castellano, A. R.; Paolino, K.; Muthumani, K.; Reuschel, E. L.; Robb, M. L.; Racine, T.; Oh, M.; Lamarre, C.; Zaidi, F. I.; Boyer, J.; Kudchodkar, S. B.; Jeong, M.; Darden, J. M.; Park, Y. K.; Scott, P. T.; Remigio, C.; Parikh, A. P.; Wise, M. C.; Patel, A.; Duperret, E. K.; Kim, K. Y.; Choi, H.; White, S.; Bagarazzi, M.; May, J. M.; Kane, D.; Lee, H.; Kobinger, G.; Michael, N. L.; Weiner, D. B.; Thomas, S. J.; Maslow, J. N. Safety and Immunogenicity of an Anti-Middle East Respiratory Syndrome Coronavirus DNA Vaccine: A Phase 1, Open-Label, Single-Arm, Dose-Escalation Trial. *Lancet Infect. Dis.* **2019**, *19* (9), 1013–1022.
[https://doi.org/10.1016/S1473-3099\(19\)30266-X](https://doi.org/10.1016/S1473-3099(19)30266-X).
- (84) Download CAS COVID-19 Antiviral Candidate Compounds Dataset
<https://www.cas.org/covid-19-antiviral-compounds-dataset> (accessed 2021 -08 -13).
- (85) Index of [/pub/databases/chembl/ChEMBLdb/releases/chembl_25/](https://pub/databases/chembl/ChEMBLdb/releases/chembl_25/)
https://ftp.ebi.ac.uk/pub/databases/chembl/ChEMBLdb/releases/chembl_25/
 (accessed 2021 -08 -13).
- (86) Sorokina, M.; Steinbeck, C. Review on Natural Products Databases: Where to Find Data in 2020. *J. Cheminformatics* **2020**, *12* (1), 20.
<https://doi.org/10.1186/s13321-020-00424-9>.
- (87) FooDB <https://foodb.ca/> (accessed 2021 -08 -13).
- (88) REAL Database - Enamine
<https://enamine.net/compound-collections/real-compounds/real-database>
 (accessed 2021 -08 -13).

- (89) Wassermann, A. M.; Lounkine, E.; Hoepfner, D.; Le Goff, G.; King, F. J.; Studer, C.; Peltier, J. M.; Grippo, M. L.; Prindle, V.; Tao, J.; Schuffenhauer, A.; Wallace, I. M.; Chen, S.; Krastel, P.; Cobos-Correa, A.; Parker, C. N.; Davies, J. W.; Glick, M. Dark Chemical Matter as a Promising Starting Point for Drug Lead Discovery. *Nat. Chem. Biol.* **2015**, *11* (12), 958–966. <https://doi.org/10.1038/nchembio.1936>.
- (90) Díaz-Eufracio, B. I.; Palomino-Hernández, O.; Houghten, R. A.; Medina-Franco, J. L. Exploring the Chemical Space of Peptides for Drug Discovery: A Focus on Linear and Cyclic Penta-Peptides. *Mol. Divers.* **2018**, *22* (2), 259–267. <https://doi.org/10.1007/s11030-018-9812-9>.
- (91) Chávez-Hernández, A. L.; Sánchez-Cruz, N.; Medina-Franco, J. L. A Fragment Library of Natural Products and Its Comparative Chemoinformatic Characterization. *Mol. Inform.* **2020**, *39* (11), 2000050. <https://doi.org/10.1002/minf.202000050>.
- (92) RDKit <http://www.rdkit.org/> (accessed 2021 -08 -13).
- (93) MolVS: Molecule Validation and Standardization — MolVS 0.1.1 documentation <https://molvs.readthedocs.io/en/latest/> (accessed 2021 -08 -13).
- (94) Pillaiyar, T.; Meenakshisundaram, S.; Manickam, M. Recent Discovery and Development of Inhibitors Targeting Coronaviruses. *Drug Discov. Today* **2020**, *25* (4), 668–688. <https://doi.org/10.1016/j.drudis.2020.01.015>.
- (95) Johansen, L. M.; Brannan, J. M.; Delos, S. E.; Shoemaker, C. J.; Stossel, A.; Lear, C.; Hoffstrom, B. G.; DeWald, L. E.; Schornberg, K. L.; Scully, C.; Lehár, J.; Hensley, L. E.; White, J. M.; Olinger, G. G. FDA-Approved Selective Estrogen Receptor Modulators Inhibit Ebola Virus Infection. *Sci. Transl. Med.* **2013**, *5* (190), 190ra79-190ra79. <https://doi.org/10.1126/scitranslmed.3005471>.
- (96) Gordon, C. J.; Tchesnokov, E. P.; Feng, J. Y.; Porter, D. P.; Götte, M. The Antiviral Compound Remdesivir Potently Inhibits RNA-Dependent RNA Polymerase from Middle East Respiratory Syndrome Coronavirus. *J. Biol. Chem.* **2020**, *295* (15), 4773–4779. <https://doi.org/10.1074/jbc.AC120.013056>.
- (97) Dyal, J.; Coleman, C. M.; Hart, B. J.; Venkataraman, T.; Holbrook, M. R.; Kindrachuk, J.; Johnson, R. F.; Olinger, G. G.; Jahrling, P. B.; Laidlaw, M.; Johansen, L. M.; Lear-Rooney, C. M.; Glass, P. J.; Hensley, L. E.; Frieman, M. B. Repurposing of Clinically Developed Drugs for Treatment of Middle East Respiratory Syndrome Coronavirus Infection. *Antimicrob. Agents Chemother.*

- 2014**, 58 (8), 4885–4893. <https://doi.org/10.1128/AAC.03036-14>.
- (98) Caly, L.; Druce, J. D.; Catton, M. G.; Jans, D. A.; Wagstaff, K. M. The FDA-Approved Drug Ivermectin Inhibits the Replication of SARS-CoV-2 in Vitro. *Antiviral Res.* **2020**, 178, 104787. <https://doi.org/10.1016/j.antiviral.2020.104787>.
- (99) Chen, F.; Chan, K. H.; Jiang, Y.; Kao, R. Y. T.; Lu, H. T.; Fan, K. W.; Cheng, V. C. C.; Tsui, W. H. W.; Hung, I. F. N.; Lee, T. S. W.; Guan, Y.; Peiris, J. S. M.; Yuen, K. Y. In Vitro Susceptibility of 10 Clinical Isolates of SARS Coronavirus to Selected Antiviral Compounds. *J. Clin. Virol.* **2004**, 31 (1), 69–75. <https://doi.org/10.1016/j.jcv.2004.03.003>.
- (100) Bajusz, D.; Rácz, A.; Héberger, K. Why Is Tanimoto Index an Appropriate Choice for Fingerprint-Based Similarity Calculations? *J. Cheminformatics* **2015**, 7 (1), 20. <https://doi.org/10.1186/s13321-015-0069-3>.
- (101) Sterling, T.; Irwin, J. J. ZINC 15 – Ligand Discovery for Everyone. *J. Chem. Inf. Model.* **2015**, 55 (11), 2324–2337. <https://doi.org/10.1021/acs.jcim.5b00559>.
- (102) Irwin, J. J.; Tang, K. G.; Young, J.; Dandarchuluun, C.; Wong, B. R.; Khurelbaatar, M.; Moroz, Y. S.; Mayfield, J.; Sayle, R. A. ZINC20—A Free Ultralarge-Scale Chemical Database for Ligand Discovery. *J. Chem. Inf. Model.* **2020**, 60 (12), 6065–6073. <https://doi.org/10.1021/acs.jcim.0c00675>.
- (103) Daina, A.; Michielin, O.; Zoete, V. SwissADME: A Free Web Tool to Evaluate Pharmacokinetics, Drug-Likeness and Medicinal Chemistry Friendliness of Small Molecules. *Sci. Rep.* **2017**, 7 (1), 42717. <https://doi.org/10.1038/srep42717>.
- (104) Pires, D. E. V.; Blundell, T. L.; Ascher, D. B. PkCSM: Predicting Small-Molecule Pharmacokinetic and Toxicity Properties Using Graph-Based Signatures. *J. Med. Chem.* **2015**, 58 (9), 4066–4072. <https://doi.org/10.1021/acs.jmedchem.5b00104>.
- (105) Gawriljuk, V. O.; Zin, P. P. K.; Puhl, A. C.; Zorn, K. M.; Foil, D. H.; Lane, T. R.; Hurst, B.; Tavella, T. A.; Costa, F. T. M.; Lakshmanane, P.; Bernatchez, J.; Godoy, A. S.; Oliva, G.; Siqueira-Neto, J. L.; Madrid, P. B.; Ekins, S. Machine Learning Models Identify Inhibitors of SARS-CoV-2. *J. Chem. Inf. Model.* **2021**. <https://doi.org/10.1021/acs.jcim.1c00683>.
- (106) Gawriljuk, V. O.; Zin, P. P. K.; Foil, D. H.; Bernatchez, J.; Beck, S.; Beutler, N.; Ricketts, J.; Yang, L.; Rogers, T.; Puhl, A. C.; Zorn, K. M.; Lane, T. R.; Godoy, A.

- S.; Oliva, G.; Siqueira-Neto, J. L.; Madrid, P. B.; Ekins, S. *Machine Learning Models Identify Inhibitors of SARS-CoV-2*; 2020; p 2020.06.16.154765. <https://doi.org/10.1101/2020.06.16.154765>.
- (107) Bobrowski, T.; Alves, V.; Melo-Filho, C. C.; Korn, D.; Auerbach, S. S.; Schmitt, C.; Muratov, E.; Tropsha, A. Computational Models Identify Several FDA Approved or Experimental Drugs as Putative Agents Against SARS-CoV-2. **2020**. <https://doi.org/10.26434/chemrxiv.12153594.v1>.
- (108) Molecular Operating Environment (MOE) | MOEsaic | PSILO <https://www.chemcomp.com/Products.htm> (accessed 2021 -08 -13).
- (109) Zhang, N.; Zhao, H. Enriching Screening Libraries with Bioactive Fragment Space. *Bioorg. Med. Chem. Lett.* **2016**, 26 (15), 3594–3597. <https://doi.org/10.1016/j.bmcl.2016.06.013>.
- (110) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2010**, 31 (2), 455–461. <https://doi.org/10.1002/jcc.21334>.
- (111) Yang, K. S.; Ma, X. R.; Ma, Y.; Alugubelli, Y. R.; Scott, D. A.; Vatansever, E. C.; Drelich, A. K.; Sankaran, B.; Geng, Z. Z.; Blankenship, L. R.; Ward, H. E.; Sheng, Y. J.; Hsu, J. C.; Kratch, K. C.; Zhao, B.; Hayatshahi, H. S.; Liu, J.; Li, P.; Fierke, C. A.; Tseng, C.-T. K.; Xu, S.; Liu, W. R. *A Speedy Route to Multiple Highly Potent SARS-CoV-2 Main Protease Inhibitors*; 2020; p 2020.07.28.223784. <https://doi.org/10.1101/2020.07.28.223784>.
- (112) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **2009**, 30 (16), 2785–2791. <https://doi.org/10.1002/jcc.21256>.
- (113) Sharun, K.; Tiwari, R.; Dhama, K. Protease Inhibitor GC376 for COVID-19: Lessons Learned from Feline Infectious Peritonitis. *Ann. Med. Surg.* **2021**, 61, 122–125. <https://doi.org/10.1016/j.amsu.2020.12.030>.
- (114) SARS-CoV-2 Mpro Protease Assay Service | Reaction Biology https://www.reactionbiology.com/datasheet/sars-cov-2_mpro_protea_malvern (accessed 2021 -08 -13).
- (115) Guo, J.; Xu, C.; Li, X.; Chen, S. A Simple, Rapid and Sensitive FRET Assay for Botulinum Neurotoxin Serotype B Detection. *PLOS ONE* **2014**, 9 (12),

- e114124. <https://doi.org/10.1371/journal.pone.0114124>.
- (116) Bemis, G. W.; Murcko, M. A. The Properties of Known Drugs. 1. Molecular Frameworks. *J. Med. Chem.* **1996**, *39* (15), 2887–2893. <https://doi.org/10.1021/jm9602928>.
- (117) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50* (5), 742–754. <https://doi.org/10.1021/ci100050t>.
- (118) Durant, J. L.; Leland, B. A.; Henry, D. R.; Nourse, J. G. Reoptimization of MDL Keys for Use in Drug Discovery. *J. Chem. Inf. Comput. Sci.* **2002**, *42* (6), 1273–1280. <https://doi.org/10.1021/ci010132r>.
- (119) Medina-Franco, J. L.; Martínez-Mayorga, K.; Bender, A.; Marín, R. M.; Giulianotti, M. A.; Pinilla, C.; Houghten, R. A. Characterization of Activity Landscapes Using 2D and 3D Similarity Methods: Consensus Activity Cliffs. *J. Chem. Inf. Model.* **2009**, *49* (2), 477–491. <https://doi.org/10.1021/ci800379q>.
- (120) Veber, D. F.; Johnson, S. R.; Cheng, H.-Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. *J. Med. Chem.* **2002**, *45* (12), 2615–2623. <https://doi.org/10.1021/jm020017n>.
- (121) Terstappen, G. C.; Reggiani, A. In Silico Research in Drug Discovery. *Trends Pharmacol. Sci.* **2001**, *22* (1), 23–26. [https://doi.org/10.1016/S0165-6147\(00\)01584-4](https://doi.org/10.1016/S0165-6147(00)01584-4).
- (122) Eze, F. U.; Okoro, U. C.; Ugwu, D. I.; Okafor, S. N. Biological Activity Evaluation of Some New Benzenesulphonamide Derivatives. *Front. Chem.* **2019**, *7*, 634. <https://doi.org/10.3389/fchem.2019.00634>.
- (123) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Original Article: S0169-409X(96)00423-1*. The Article Was Originally Published in *Advanced Drug Delivery Reviews* 23 (1997) 3–25.1. *Adv. Drug Deliv. Rev.* **2001**, *46* (1), 3–26. [https://doi.org/10.1016/S0169-409X\(00\)00129-0](https://doi.org/10.1016/S0169-409X(00)00129-0).
- (124) Shultz, M. D. Two Decades under the Influence of the Rule of Five and the Changing Properties of Approved Oral Drugs. *J. Med. Chem.* **2019**, *62* (4), 1701–1714. <https://doi.org/10.1021/acs.jmedchem.8b00686>.
- (125) Ottaviani, G.; Gosling, D. J.; Patissier, C.; Rodde, S.; Zhou, L.; Faller, B. What Is Modulating Solubility in Simulated Intestinal Fluids? *Eur. J. Pharm. Sci.* **2010**,

- 41 (3), 452–457. <https://doi.org/10.1016/j.ejps.2010.07.012>.
- (126) Savjani, K. T.; Gajjar, A. K.; Savjani, J. K. Drug Solubility: Importance and Enhancement Techniques. *ISRN Pharm.* **2012**, 2012, e195727. <https://doi.org/10.5402/2012/195727>.
- (127) Sanders, M. P. A.; Barbosa, A. J. M.; Zarzycka, B.; Nicolaes, G. A. F.; Klomp, J. P. G.; de Vlieg, J.; Del Rio, A. Comparative Analysis of Pharmacophore Screening Tools. *J. Chem. Inf. Model.* **2012**, 52 (6), 1607–1620. <https://doi.org/10.1021/ci2005274>.
- (128) Taylor, R.; Cole, J. C.; Cosgrove, D. A.; Gardiner, E. J.; Gillet, V. J.; Korb, O. Development and Validation of an Improved Algorithm for Overlaying Flexible Molecules. *J. Comput. Aided Mol. Des.* **2012**, 26 (4), 451–472. <https://doi.org/10.1007/s10822-012-9573-y>.
- (129) Kaserer, T.; Schuster, D.; Rollinger, J. M. Chemoinformatics in Natural Product Research. In *Applied Chemoinformatics*; John Wiley & Sons, Ltd, 2018; pp 207–236. <https://doi.org/10.1002/9783527806539.ch6c>.
- (130) Ginex, T.; Vazquez, J.; Gilbert, E.; Herrero, E.; Luque, F. J. Lipophilicity in Drug Design: An Overview of Lipophilicity Descriptors in 3D-QSAR Studies. *Future Med. Chem.* **2019**, 11 (10), 1177–1193. <https://doi.org/10.4155/fmc-2018-0435>.
- (131) Tsopeles, F.; Giaginis, C.; Tsantili-Kakoulidou, A. Lipophilicity and Biomimetic Properties to Support Drug Discovery. *Expert Opin. Drug Discov.* **2017**, 12 (9), 885–896. <https://doi.org/10.1080/17460441.2017.1344210>.
- (132) Yang, C.; Rathman, J. F.; Tarkhov, A.; Sacher, O.; Kleinoeder, T.; Liu, J.; Magdziarz, T.; Mostraq, A.; Marusczyk, J.; Mehta, D.; Schwab, C.; Bienfait, B. Chemoinformatics in Modern Regulatory Science. In *Applied Chemoinformatics*; John Wiley & Sons, Ltd, 2018; pp 439–470. <https://doi.org/10.1002/9783527806539.ch8>.
- (133) Cheng, T.; Zhao, Y.; Li, X.; Lin, F.; Xu, Y.; Zhang, X.; Li, Y.; Wang, R.; Lai, L. Computation of Octanol–Water Partition Coefficients by Guiding an Additive Model with Knowledge. *J. Chem. Inf. Model.* **2007**, 47 (6), 2140–2148. <https://doi.org/10.1021/ci700257y>.
- (134) Wildman, S. A.; Crippen, G. M. Prediction of Physicochemical Parameters by Atomic Contributions. *J. Chem. Inf. Comput. Sci.* **1999**, 39 (5), 868–873. <https://doi.org/10.1021/ci990307l>.

- (135) Moriguchi, I.; Hirono, S.; Liu, Q.; Nakagome, I.; Matsushita, Y. Simple Method of Calculating Octanol/Water Partition Coefficient. *Chem. Pharm. Bull. (Tokyo)* **1992**, *40* (1), 127–130. <https://doi.org/10.1248/cpb.40.127>.
- (136) Daina, A.; Michielin, O.; Zoete, V. ILOGP: A Simple, Robust, and Efficient Description of n-Octanol/Water Partition Coefficient for Drug Design Using the GB/SA Approach. *J. Chem. Inf. Model.* **2014**, *54* (12), 3284–3301. <https://doi.org/10.1021/ci500467k>.
- (137) Kiela, P. R.; Ghishan, F. K. Physiology of Intestinal Absorption and Secretion. *Best Pract. Res. Clin. Gastroenterol.* **2016**, *30* (2), 145–159. <https://doi.org/10.1016/j.bpg.2016.02.007>.
- (138) Yan, A.; Wang, Z.; Cai, Z. Prediction of Human Intestinal Absorption by GA Feature Selection and Support Vector Machine Regression. *Int. J. Mol. Sci.* **2008**, *9* (10), 1961–1976. <https://doi.org/10.3390/ijms9101961>.
- (139) Wang, N.-N.; Huang, C.; Dong, J.; Yao, Z.-J.; Zhu, M.-F.; Deng, Z.-K.; Lv, B.; Lu, A.-P.; Chen, A. F.; Cao, D.-S. Predicting Human Intestinal Absorption with Modified Random Forest Approach: A Comprehensive Evaluation of Molecular Representation, Unbalanced Data, and Applicability Domain Issues. *RSC Adv.* **2017**, *7* (31), 19007–19018. <https://doi.org/10.1039/C6RA28442F>.
- (140) Radchenko, E. V.; Dyabina, A. S.; Palyulin, V. A.; Zefirov, N. S. Prediction of Human Intestinal Absorption of Drug Compounds. *Russ. Chem. Bull.* **2016**, *65* (2), 576–580. <https://doi.org/10.1007/s11172-016-1340-0>.
- (141) Hou, T.; Wang, J.; Zhang, W.; Xu, X. ADME Evaluation in Drug Discovery. 7. Prediction of Oral Absorption by Correlation and Classification. *J. Chem. Inf. Model.* **2007**, *47* (1), 208–218. <https://doi.org/10.1021/ci600343x>.
- (142) Rock, B. M.; Foti, R. S. Pharmacokinetic and Drug Metabolism Properties of Novel Therapeutic Modalities. *Drug Metab. Dispos.* **2019**, *47* (10), 1097–1099. <https://doi.org/10.1124/dmd.119.088708>.
- (143) Cheng, F.; Li, W.; Zhou, Y.; Shen, J.; Wu, Z.; Liu, G.; Lee, P. W.; Tang, Y. AdmetSAR: A Comprehensive Source and Free Tool for Assessment of Chemical ADMET Properties. *J. Chem. Inf. Model.* **2012**, *52* (11), 3099–3105. <https://doi.org/10.1021/ci300367a>.
- (144) Tian, S.; Wang, J.; Li, Y.; Li, D.; Xu, L.; Hou, T. The Application of in Silico Drug-Likeness Predictions in Pharmaceutical Research. *Adv. Drug Deliv. Rev.* **2015**, *86*, 2–10. <https://doi.org/10.1016/j.addr.2015.01.009>.

- (145) Deeksha, S.; Anju, S.; Mohammed, H. S.; Rajnish, K. Blood Brain Barrier Permeability Prediction Using Machine Learning Techniques: An Update. *Curr. Pharm. Biotechnol.* **2019**, *20* (14), 1163–1171.
- (146) Gupta, M.; Lee, H. J.; Barden, C. J.; Weaver, D. F. The Blood–Brain Barrier (BBB) Score. *J. Med. Chem.* **2019**, *62* (21), 9824–9836. <https://doi.org/10.1021/acs.jmedchem.9b01220>.
- (147) Suenderhauf, C.; Hammann, F.; Huwyler, J. Computational Prediction of Blood-Brain Barrier Permeability Using Decision Tree Induction. *Molecules* **2012**, *17* (9), 10429–10445. <https://doi.org/10.3390/molecules170910429>.
- (148) Lin, J. H.; Yamazaki, M. Role of P-Glycoprotein in Pharmacokinetics. *Clin. Pharmacokinet.* **2003**, *42* (1), 59–98. <https://doi.org/10.2165/00003088-200342010-00003>.
- (149) Gottesman, M. M.; Pastan, I. Biochemistry of Multidrug Resistance Mediated by the Multidrug Transporter. *Annu. Rev. Biochem.* **1993**, *62* (1), 385–427. <https://doi.org/10.1146/annurev.bi.62.070193.002125>.
- (150) Hung, L.-W.; Wang, I. X.; Nikaido, K.; Liu, P.-Q.; Ames, G. F.-L.; Kim, S.-H. Crystal Structure of the ATP-Binding Subunit of an ABC Transporter. *Nature* **1998**, *396* (6712), 703–707. <https://doi.org/10.1038/25393>.
- (151) Ambudkar, S. V.; Dey, S.; Hrycyna, C. A.; Ramachandra, M.; Pastan, I.; Gottesman, M. M. Biochemical, Cellular, and Pharmacological Aspects of the Multidrug Transporter. *Annu. Rev. Pharmacol. Toxicol.* **1999**, *39* (1), 361–398. <https://doi.org/10.1146/annurev.pharmtox.39.1.361>.
- (152) Wu, Q.; Bounaud, P.-Y.; Kuduk, S. D.; Yang, C.-P. H.; Ojima, I.; Horwitz, S. B.; Orr, G. A. Identification of the Domains of Photoincorporation of the 3'- and 7-Benzophenone Analogues of Taxol in the Carboxyl-Terminal Half of Murine Mdr1b P-Glycoprotein. *Biochemistry* **1998**, *37* (32), 11272–11279. <https://doi.org/10.1021/bi980598b>.
- (153) Taguchi, Y.; Kino, K.; Morishima, M.; Komano, T.; Kane, S. E.; Ueda, K. Alteration of Substrate Specificity by Mutations at the His61 Position in Predicted Transmembrane Domain 1 of Human MDR1/P-Glycoprotein. *Biochemistry* **1997**, *36* (29), 8883–8889. <https://doi.org/10.1021/bi970553v>.
- (154) Taguchi, Y.; Morishima, M.; Komano, T.; Ueda, K. Amino Acid Substitutions in the First Transmembrane Domain (TM1) of P-Glycoprotein That Alter Substrate Specificity. *FEBS Lett.* **1997**, *413* (1), 142–146.

- [https://doi.org/10.1016/S0014-5793\(97\)00899-5](https://doi.org/10.1016/S0014-5793(97)00899-5).
- (155) Broccatelli, F.; Carosati, E.; Neri, A.; Frosini, M.; Goracci, L.; Oprea, T. I.; Cruciani, G. A Novel Approach for Predicting P-Glycoprotein (ABCB1) Inhibition Using Molecular Interaction Fields. *J. Med. Chem.* **2011**, *54* (6), 1740–1751. <https://doi.org/10.1021/jm101421d>.
- (156) Tornio, A.; Backman, J. T. Chapter One - Cytochrome P450 in Pharmacogenetics: An Update. In *Advances in Pharmacology*; Brøsen, K., Damkier, P., Eds.; Pharmacogenetics; Academic Press, 2018; Vol. 83, pp 3–32. <https://doi.org/10.1016/bs.apha.2018.04.007>.
- (157) Palrasu, M.; Siddavaram, N. Cytochrome P450 Structure, Function and Clinical Significance: A Review. *Curr. Drug Targets* **2017**, *19* (1), 38–54.
- (158) Kapelyukh, Y.; Henderson, C. J.; Scheer, N.; Rode, A.; Wolf, C. R. Defining the Contribution of CYP1A1 and CYP1A2 to Drug Metabolism Using Humanized CYP1A1/1A2 and Cyp1a1/Cyp1a2 Knockout Mice. *Drug Metab. Dispos.* **2019**, *47* (8), 907–918. <https://doi.org/10.1124/dmd.119.087718>.
- (159) Zhou, B.; Song, Z.; Qian, M.; Li, L.; Gong, J.; Zou, S. Functional Polymorphisms in the CYP2C19 Gene Contribute to Digestive System Cancer Risk: Evidence from 11,042 Subjects. *PLOS ONE* **2013**, *8* (7), e66865. <https://doi.org/10.1371/journal.pone.0066865>.
- (160) Wang, K.; Gao, Q.; Zhang, T.; Rao, J.; Ding, L.; Qiu, F. Inhibition of CYP2C9 by Natural Products: Insight into the Potential Risk of Herb-Drug Interactions. *Drug Metab. Rev.* **2020**, *52* (2), 235–257. <https://doi.org/10.1080/03602532.2020.1758714>.
- (161) Ferguson, S. S.; LeCluyse, E. L.; Negishi, M.; Goldstein, J. A. Regulation of Human CYP2C9 by the Constitutive Androstane Receptor: Discovery of a New Distal Binding Site. *Mol. Pharmacol.* **2002**, *62* (3), 737–746. <https://doi.org/10.1124/mol.62.3.737>.
- (162) Daly, A. K.; Rettie, A. E.; Fowler, D. M.; Miners, J. O. Pharmacogenomics of CYP2C9: Functional and Clinical Considerations. *J. Pers. Med.* **2018**, *8* (1), 1. <https://doi.org/10.3390/jpm8010001>.
- (163) Kneller, L. A.; Zubiaur, P.; Koller, D.; Abad-Santos, F.; Hempel, G. Influence of CYP2D6 Phenotypes on the Pharmacokinetics of Aripiprazole and Dehydro-Aripiprazole Using a Physiologically Based Pharmacokinetic Approach. *Clin. Pharmacokinet.* **2021**.

<https://doi.org/10.1007/s40262-021-01041-x>.

- (164) Werk, A. N.; Cascorbi, I. Functional Gene Variants of CYP3A4. *Clin. Pharmacol. Ther.* **2014**, *96* (3), 340–348. <https://doi.org/10.1038/clpt.2014.129>.
- (165) Yap, C. W.; Li, Z. R.; Chen, Y. Z. Quantitative Structure–Pharmacokinetic Relationships for Drug Clearance by Using Statistical Learning Methods. *J. Mol. Graph. Model.* **2006**, *24* (5), 383–395. <https://doi.org/10.1016/j.jmgm.2005.10.004>.
- (166) Pratama, M. R. F.; Poerwono, H.; Siswodiharjo, S. ADMET Properties of Novel 5-O-Benzoylpinostrobin Derivatives. *J. Basic Clin. Physiol. Pharmacol.* **2019**, *30* (6). <https://doi.org/10.1515/jbcpp-2019-0251>.
- (167) Berellini, G.; Waters, N. J.; Lombardo, F. In Silico Prediction of Total Human Plasma Clearance. *J. Chem. Inf. Model.* **2012**, *52* (8), 2069–2078. <https://doi.org/10.1021/ci300155y>.
- (168) Zeiger, E. The Test That Changed the World: The Ames Test and the Regulation of Chemicals. *Mutat. Res. Toxicol. Environ. Mutagen.* **2019**, *841*, 43–48. <https://doi.org/10.1016/j.mrgentox.2019.05.007>.
- (169) Vian, M.; Raitano, G.; Roncaglioni, A.; Benfenati, E. In Silico Model for Mutagenicity (Ames Test), Taking into Account Metabolism. *Mutagenesis* **2019**, *34* (1), 41–48. <https://doi.org/10.1093/mutage/gey045>.
- (170) Patlewicz, G.; Jeliaskova, N.; Safford, R. J.; Worth, A. P.; Aleksiev, B. An Evaluation of the Implementation of the Cramer Classification Scheme in the Toxtree Software. *SAR QSAR Environ. Res.* **2008**, *19* (5–6), 495–524. <https://doi.org/10.1080/10629360802083871>.
- (171) Fourches, D.; Barnes, J. C.; Day, N. C.; Bradley, P.; Reed, J. Z.; Tropsha, A. Cheminformatics Analysis of Assertions Mined from Literature That Describe Drug-Induced Liver Injury in Different Species. *Chem. Res. Toxicol.* **2010**, *23* (1), 171–183. <https://doi.org/10.1021/tx900326k>.
- (172) Miteva, M. A.; Lee, W. H.; Montes, M. O.; Villoutreix, B. O. Fast Structure-Based Virtual Ligand Screening Combining FRED, DOCK, and Surflex. *J. Med. Chem.* **2005**, *48* (19), 6012–6022. <https://doi.org/10.1021/jm050262h>.
- (173) Ma, C.; Hu, Y.; Townsend, J. A.; Lagarias, P. I.; Marty, M. T.; Kolocouris, A.; Wang, J. Ebselen, Disulfiram, Carmofur, PX-12, Tideglusib, and Shikonin Are Nonspecific Promiscuous SARS-CoV-2 Main Protease Inhibitors. *ACS*

Pharmacol. Transl. Sci. **2020**, 3 (6), 1265–1277.
<https://doi.org/10.1021/acsptsci.0c00130>.

9. Apéndice

Contenido

Contenido		Páginas
Figuras S1-S65	Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs. cada uno de los 65 <i>queries</i> .	102-134
Tablas S1-S7	Parámetros estadísticos de las propiedades ADME/Tox calculadas con los servidores <i>SwissADME</i> y <i>pkCSM pharmacokinetics</i> .	135-138
Tabla S8	Nomenclatura SMILES de los 18 <i>hits</i> computacionales.	139
Figuras S66-S83	Diagramas de posibles interacciones entre los hits computacionales y la Mpro del SARS-Cov-2 en 2D.	140-156
Resultados del proyecto	Cartel y resumen como parte de los resultados presentados del proyecto.	157-158
Publicaciones científicas	<i>Chapter 2 - Computational Drug Design Methods - Current and Future Perspectives</i>	159-184
	<i>Expanding the Structural Diversity of DNA Methyltransferase Inhibitors</i>	185-201

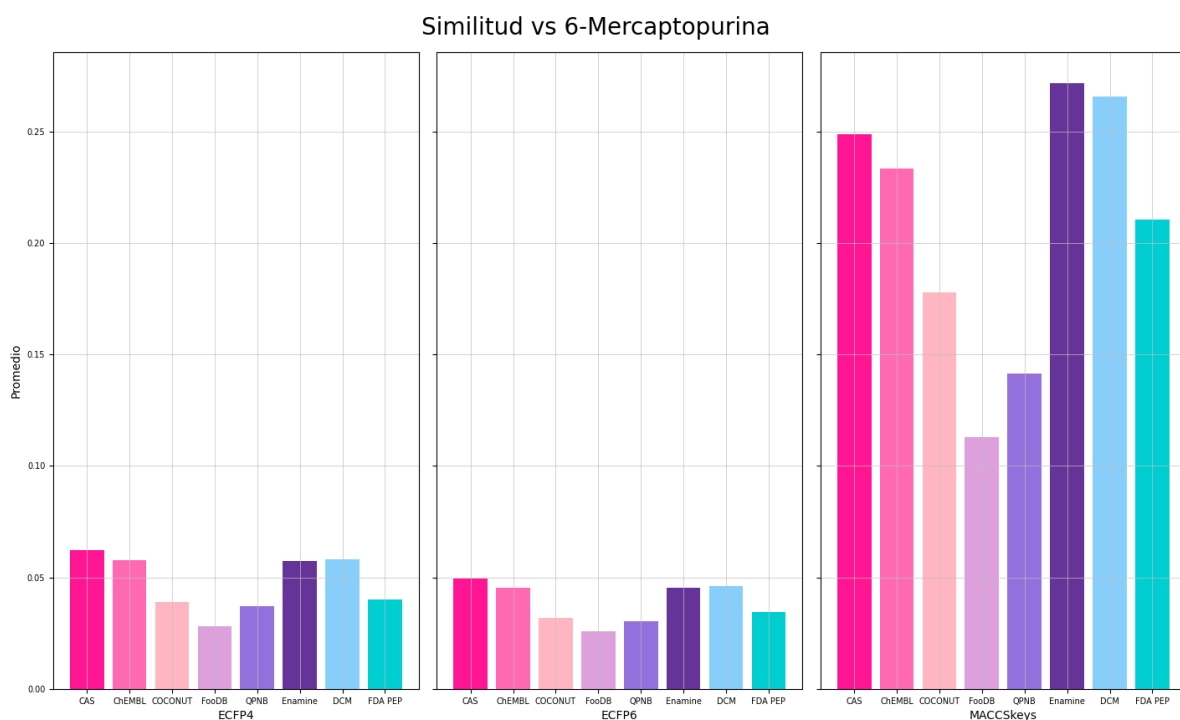


Figura S1. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs 6-mercaptopurina.

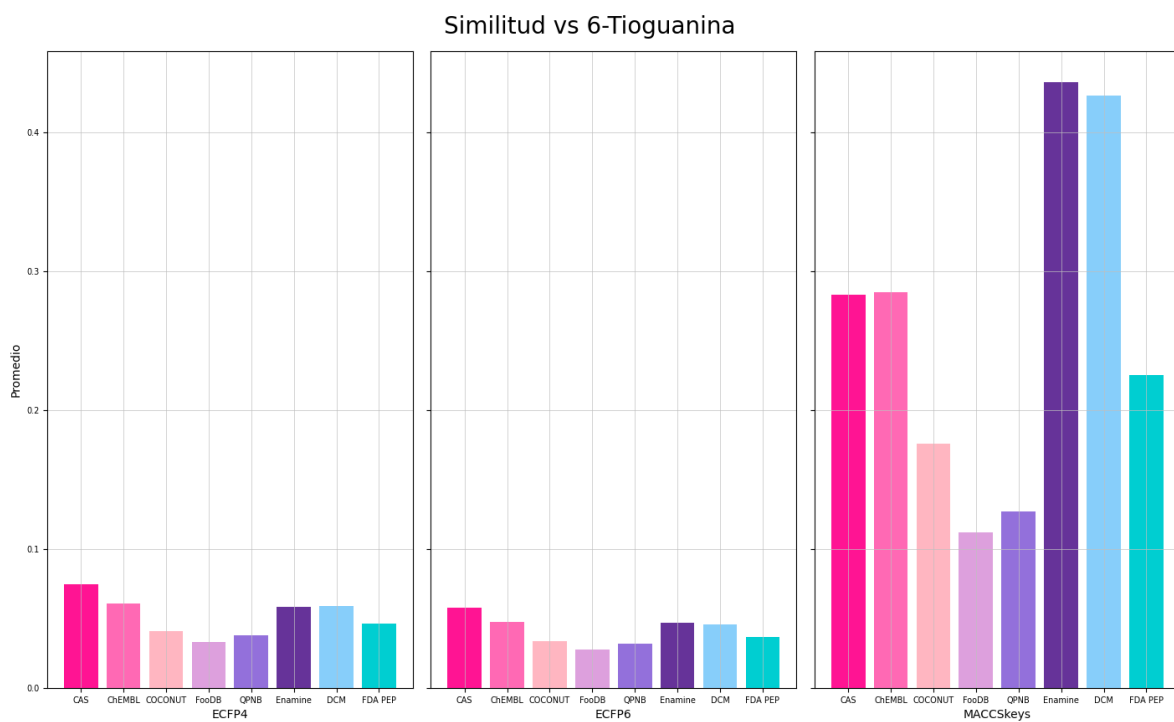


Figura S2. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs 6-tioguanina.

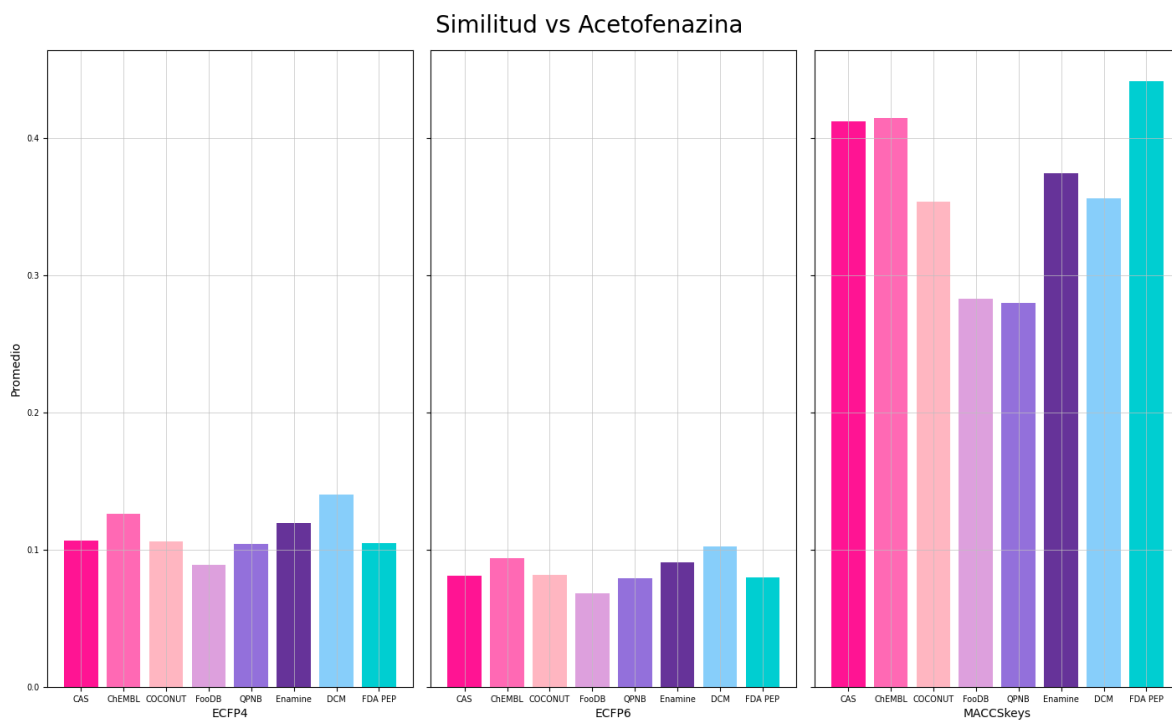


Figura S3. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs acetofenazina.

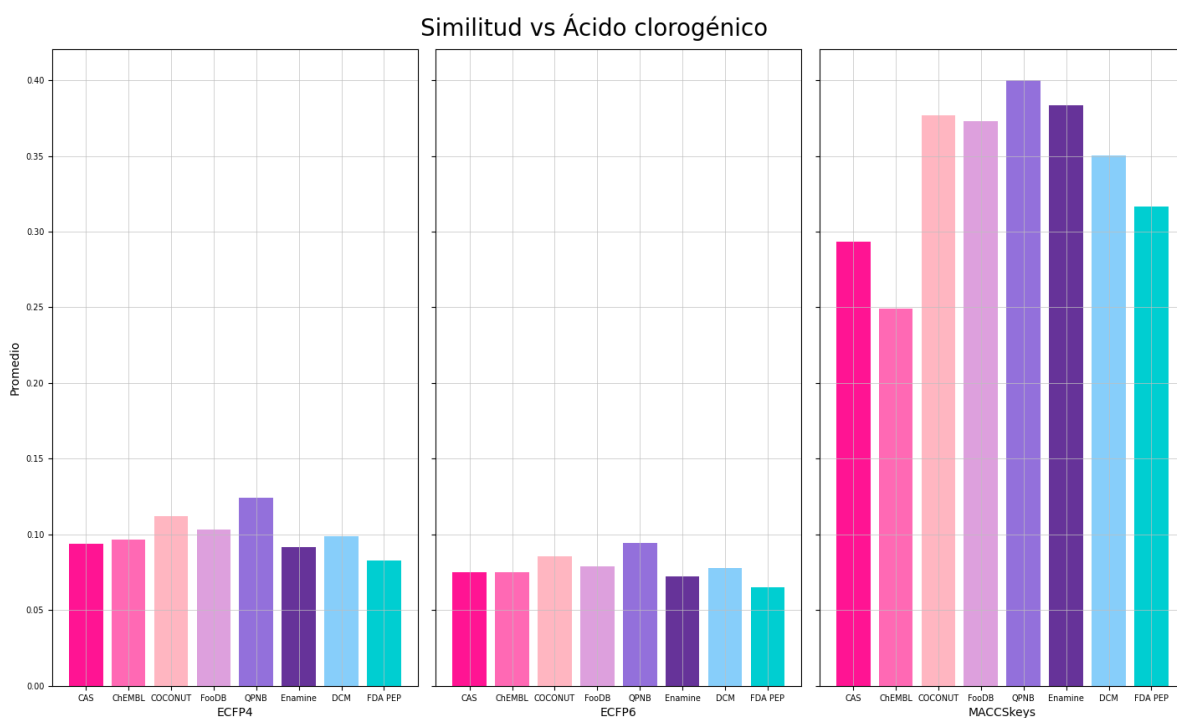


Figura S4. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs ácido clorogénico.

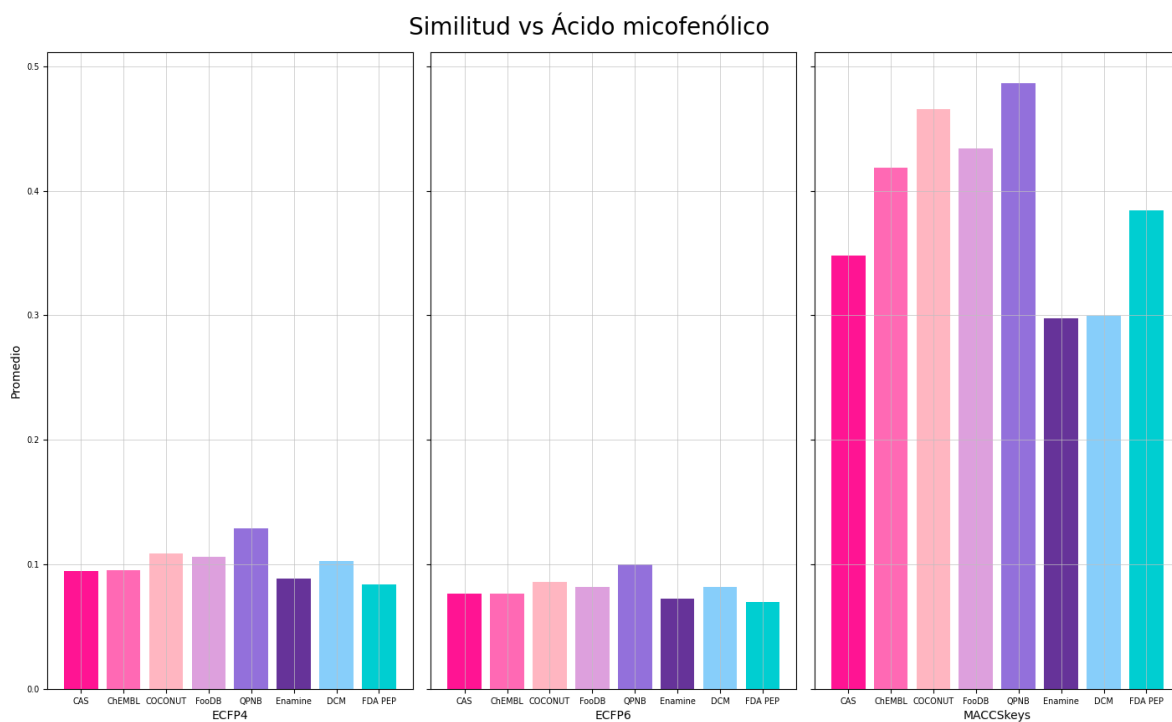


Figura S5. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs ácido micofenólico.

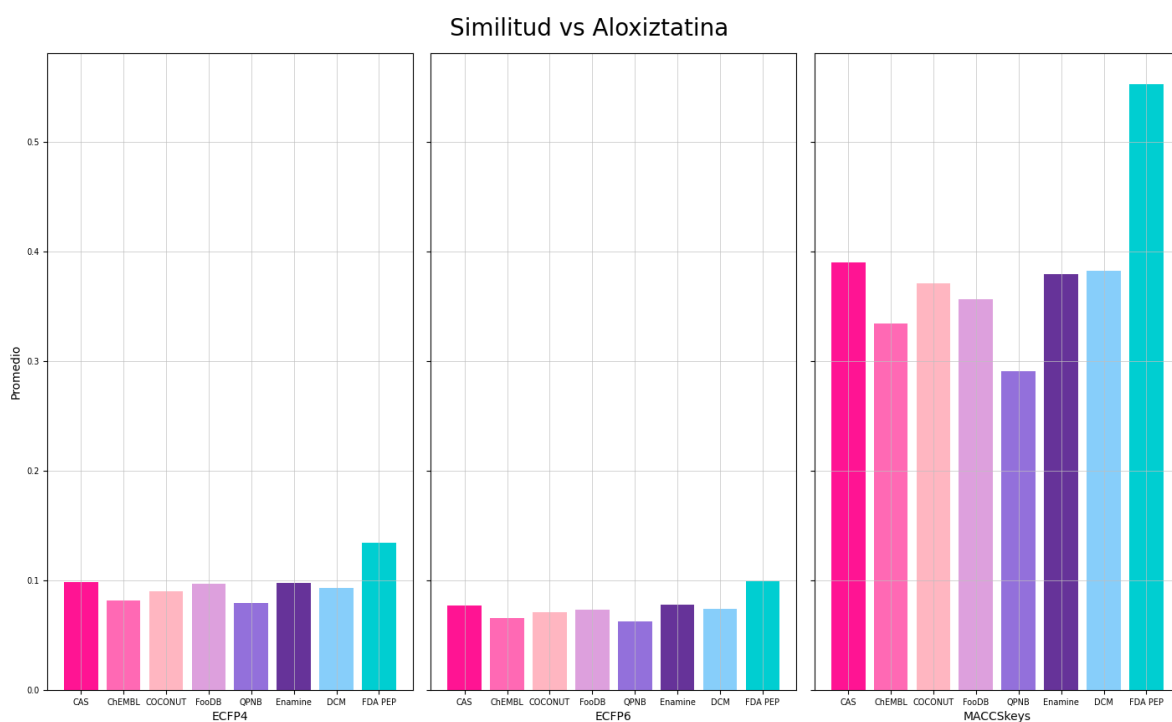


Figura S6. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs aloxitatina.

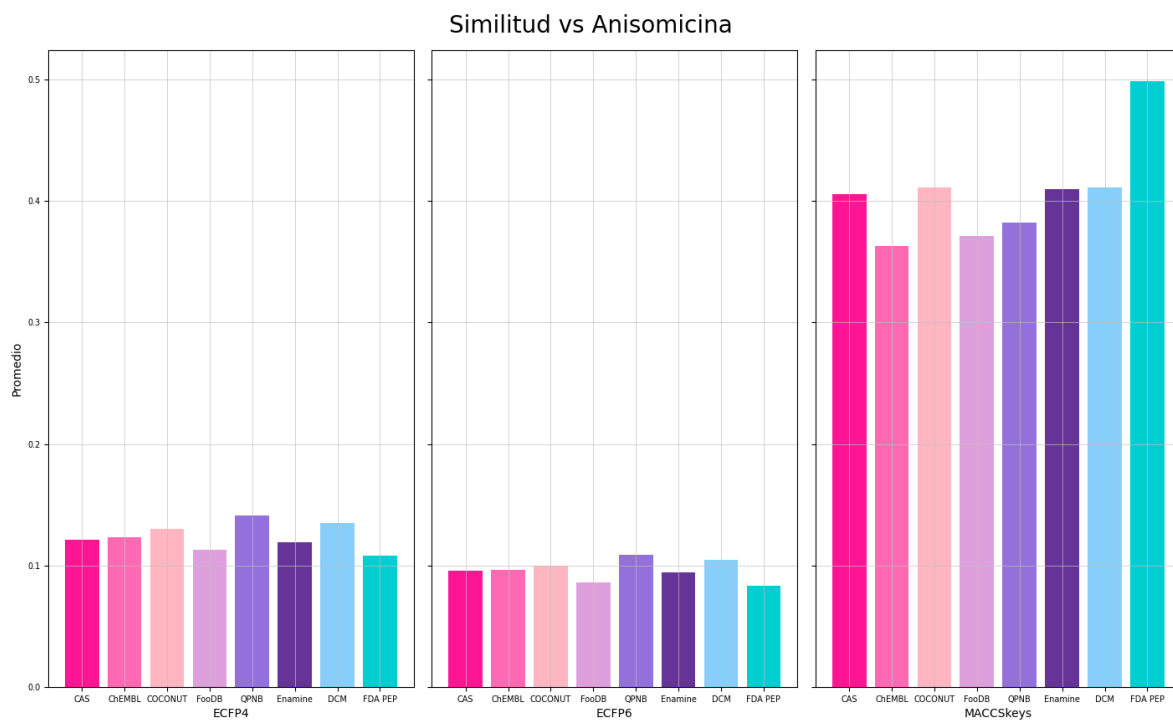


Figura S7. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs anisomicina.

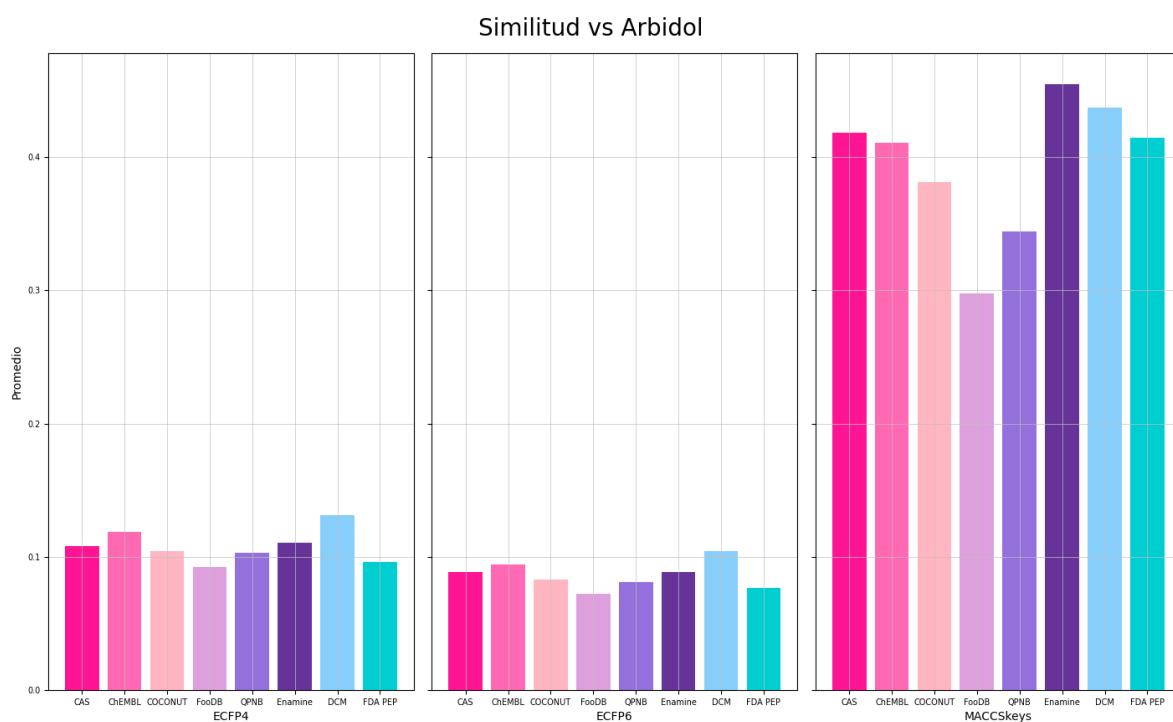


Figura S8. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs arbidol.

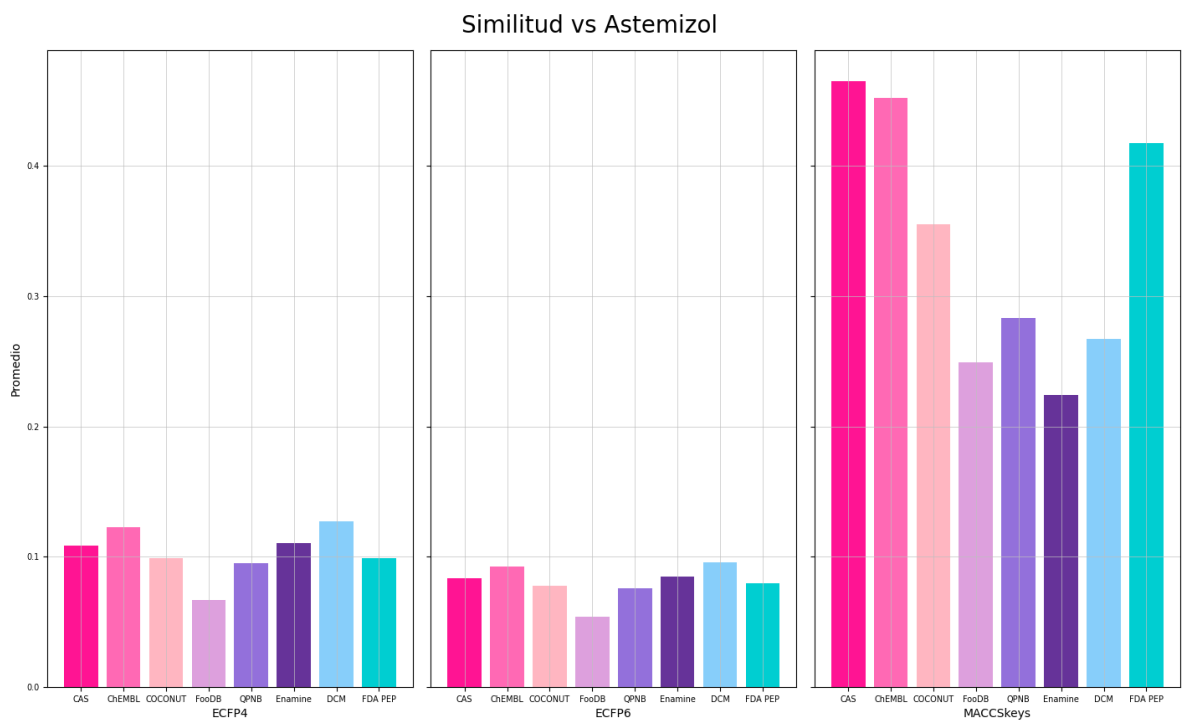


Figura S9. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs astemizol.

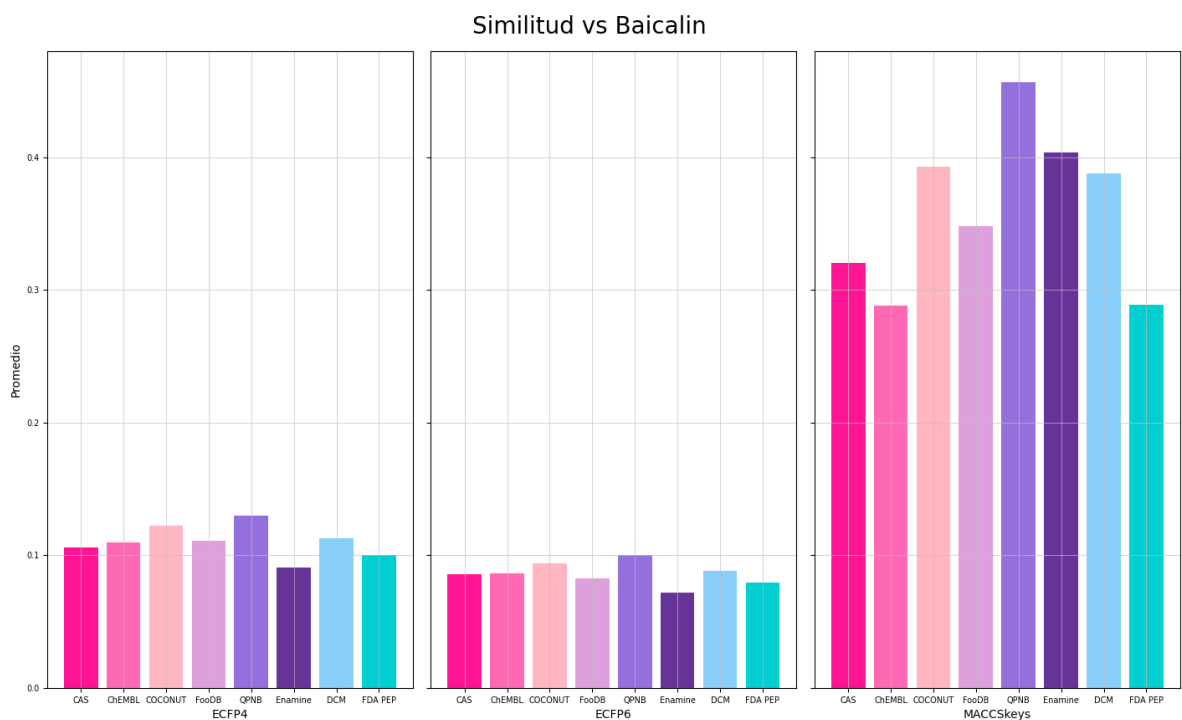


Figura S10. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs baicalin

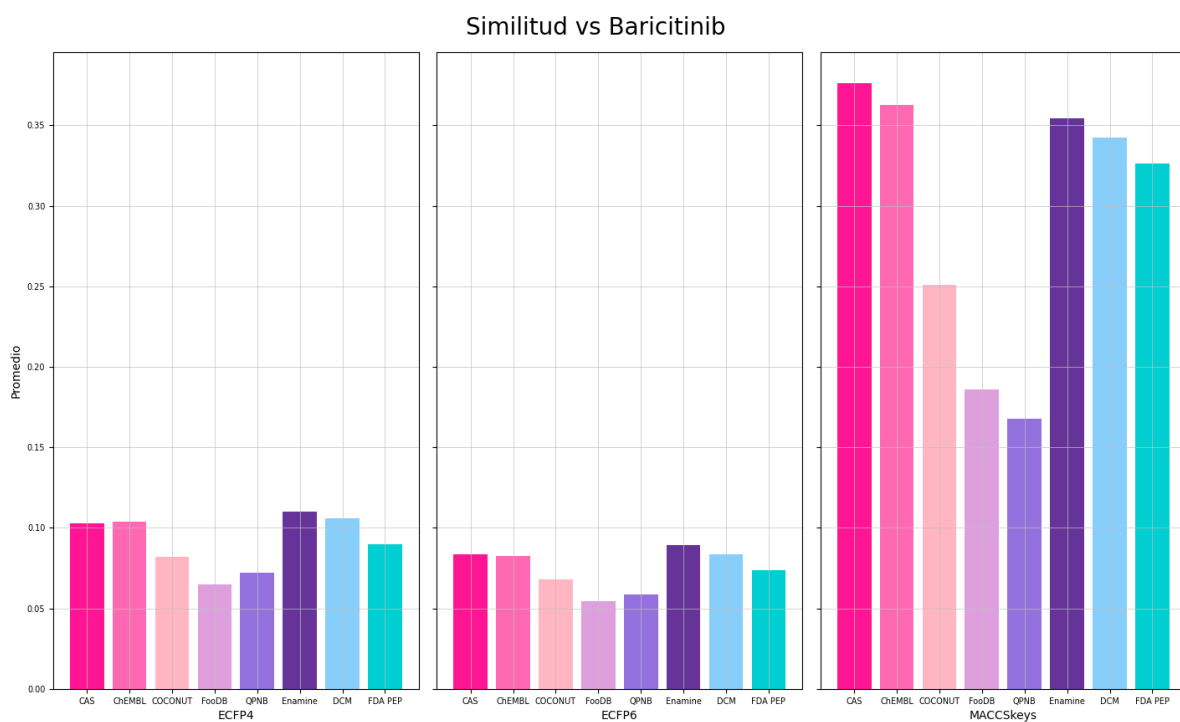


Figura S11. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs baricitinib.

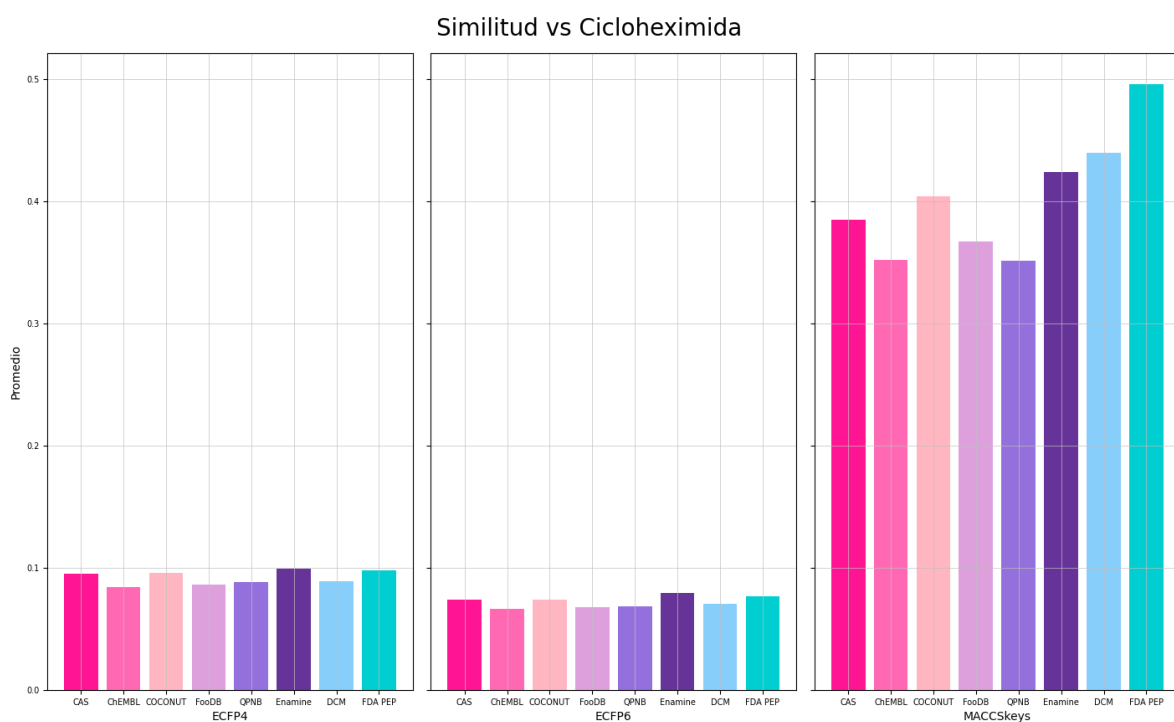


Figura S12. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs cicloheximida.

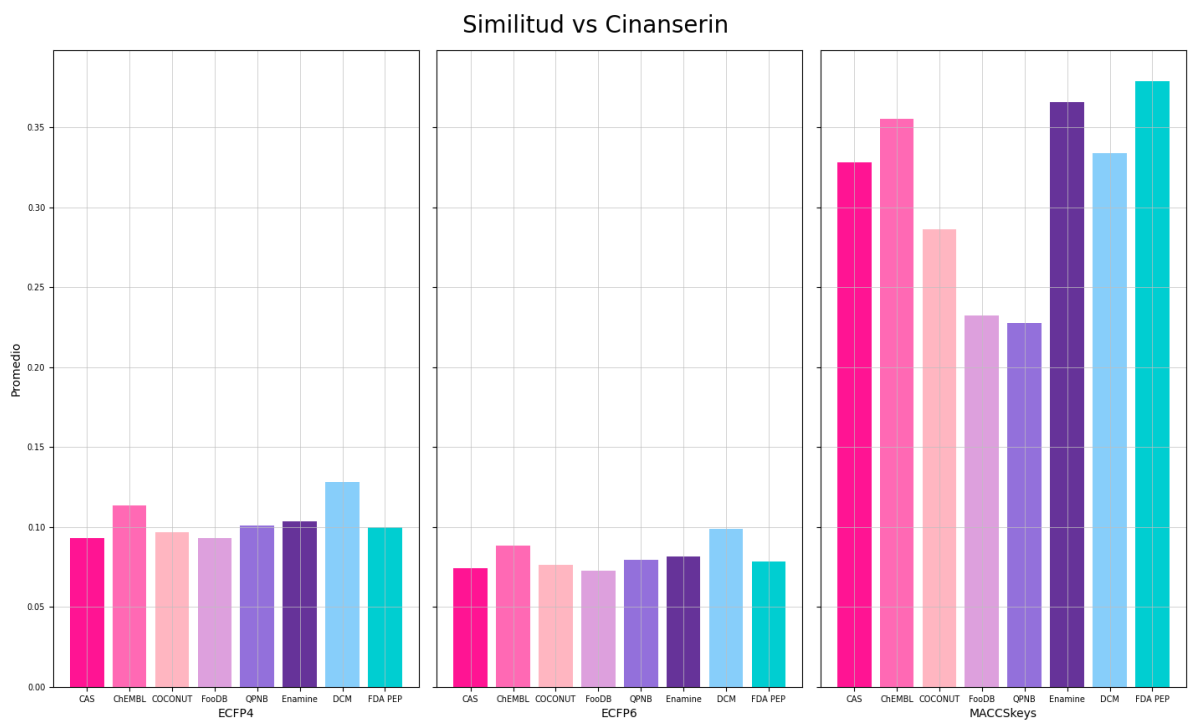


Figura S13. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs cinanserin.

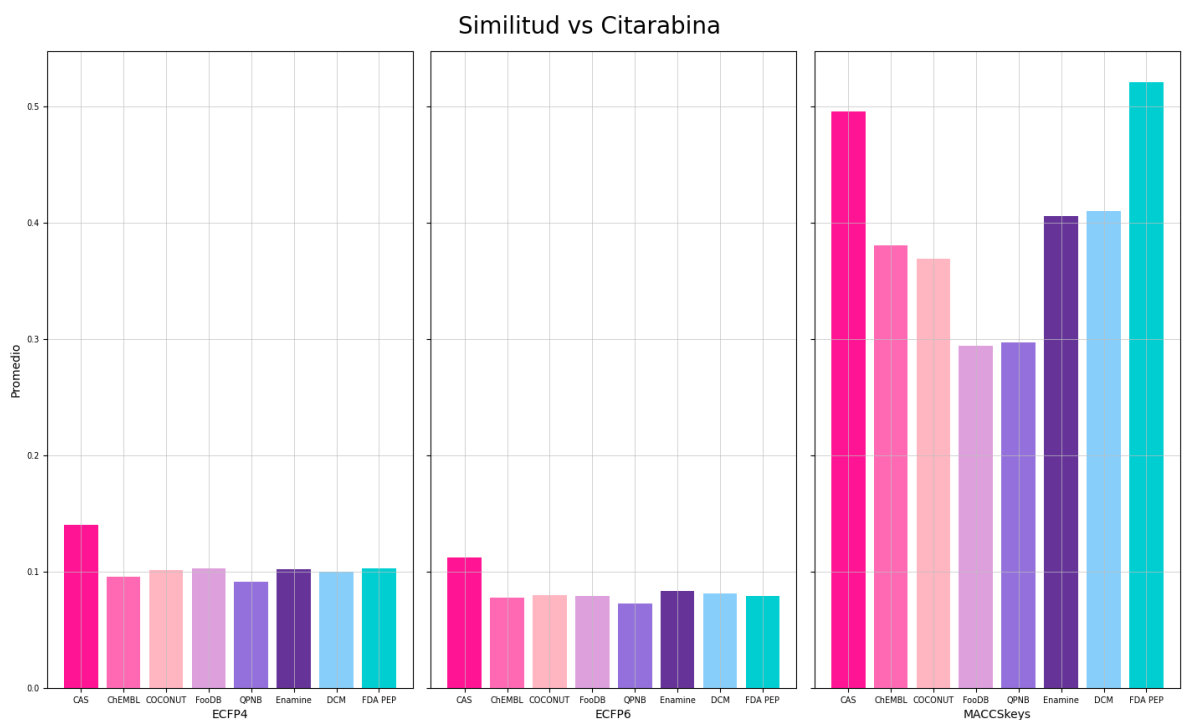


Figura S14. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs citarabina.

Similitud vs Citrato de tamoxifeno

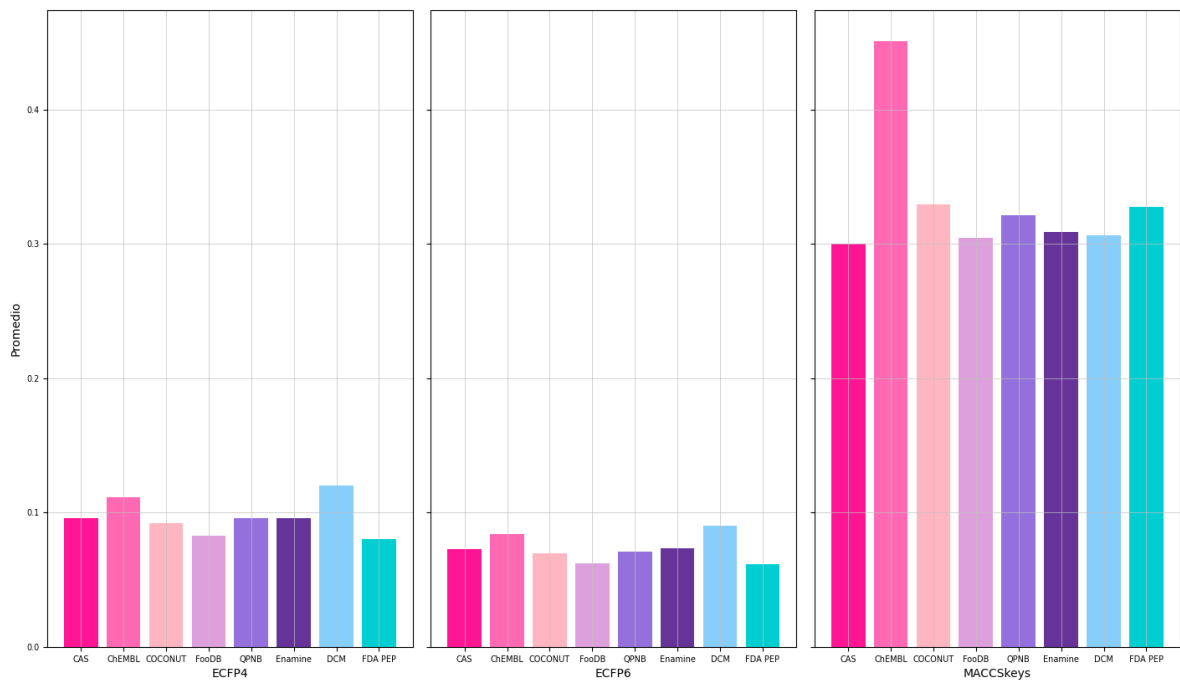


Figura S15. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs citrato de tamoxifeno.

Similitud vs Citrato de toremifeno

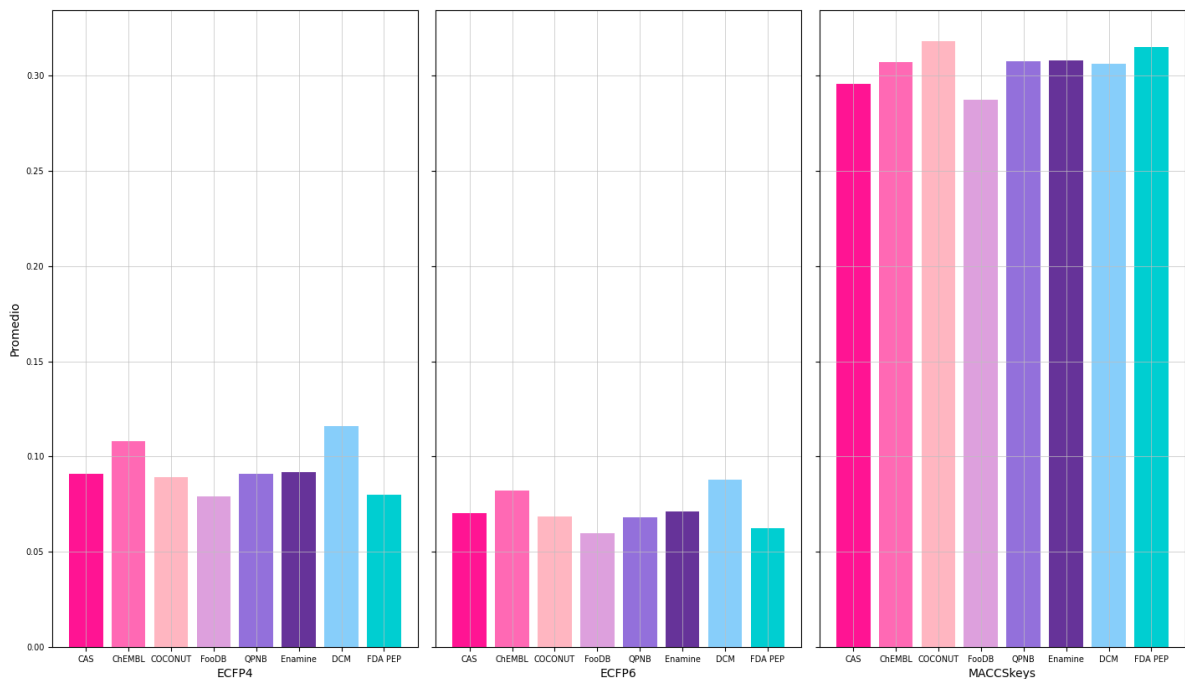


Figura S16. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs citrato de toremifeno.

Similitud vs Clorhidrato de amodiaquina

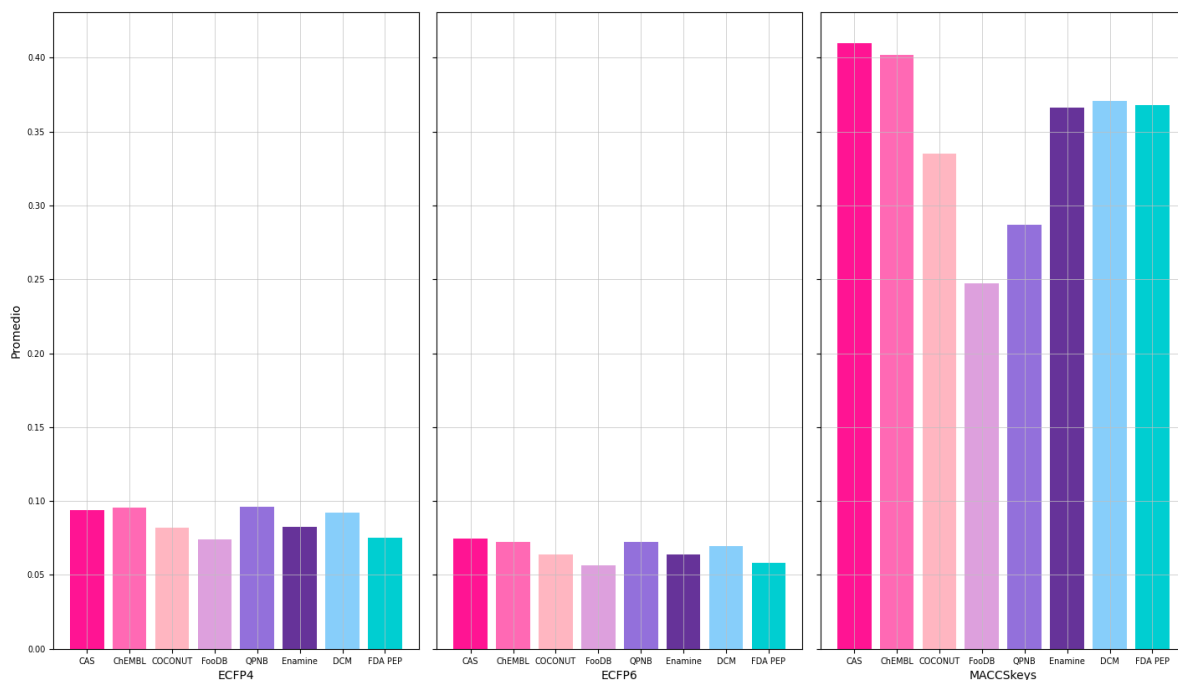


Figura S17. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs clorhidrato de amodiaquina.

Similitud vs Clorhidrato de clomipramina

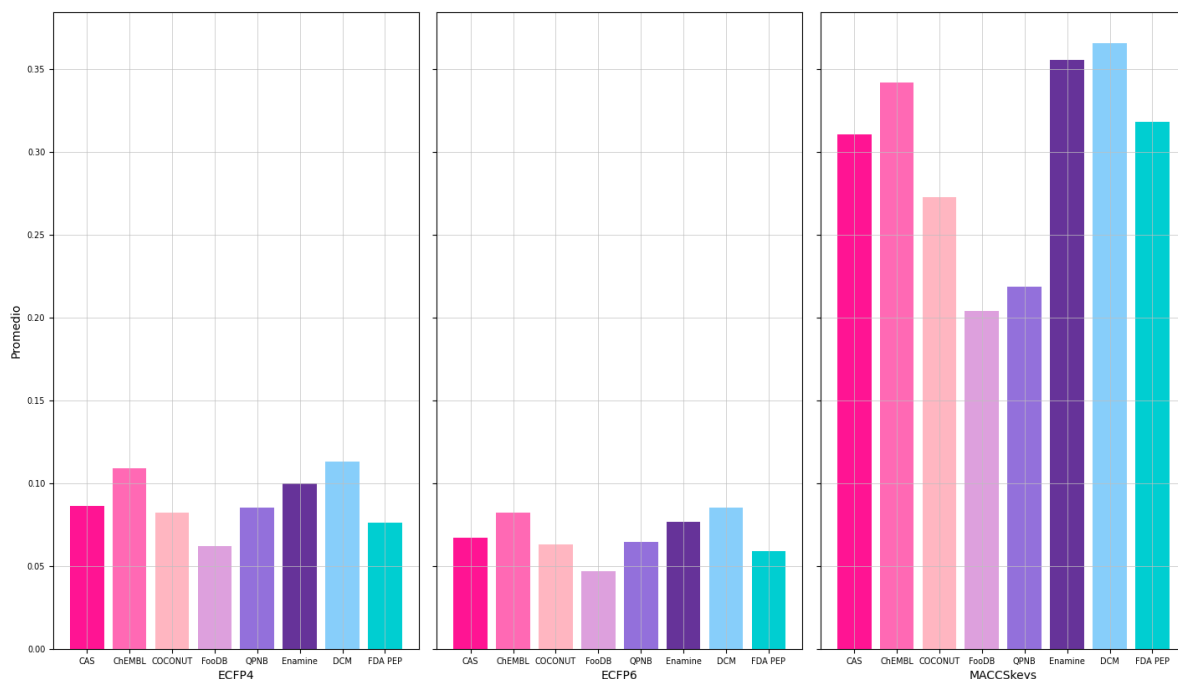


Figura S18. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs clorhidrato de clomipramina.

Similitud vs Clorhidrato de clorfenoxamina

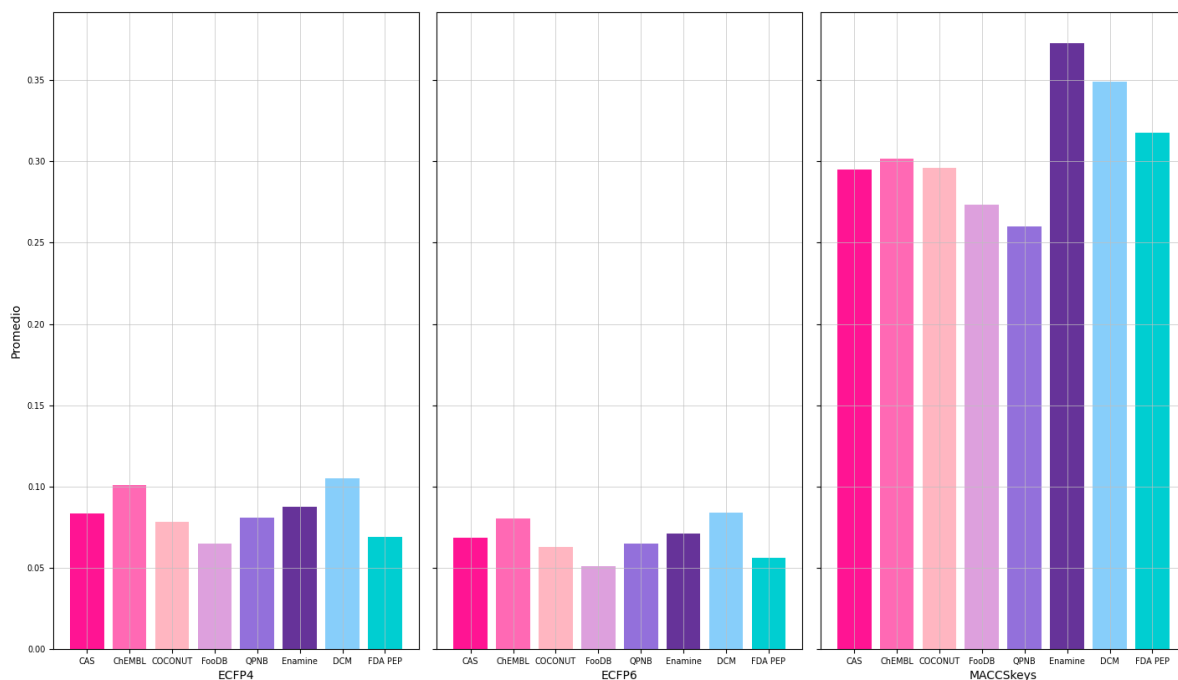


Figura S19. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs clorhidrato de clorfenoxamina.

Similitud vs Clorhidrato de clorpromazina

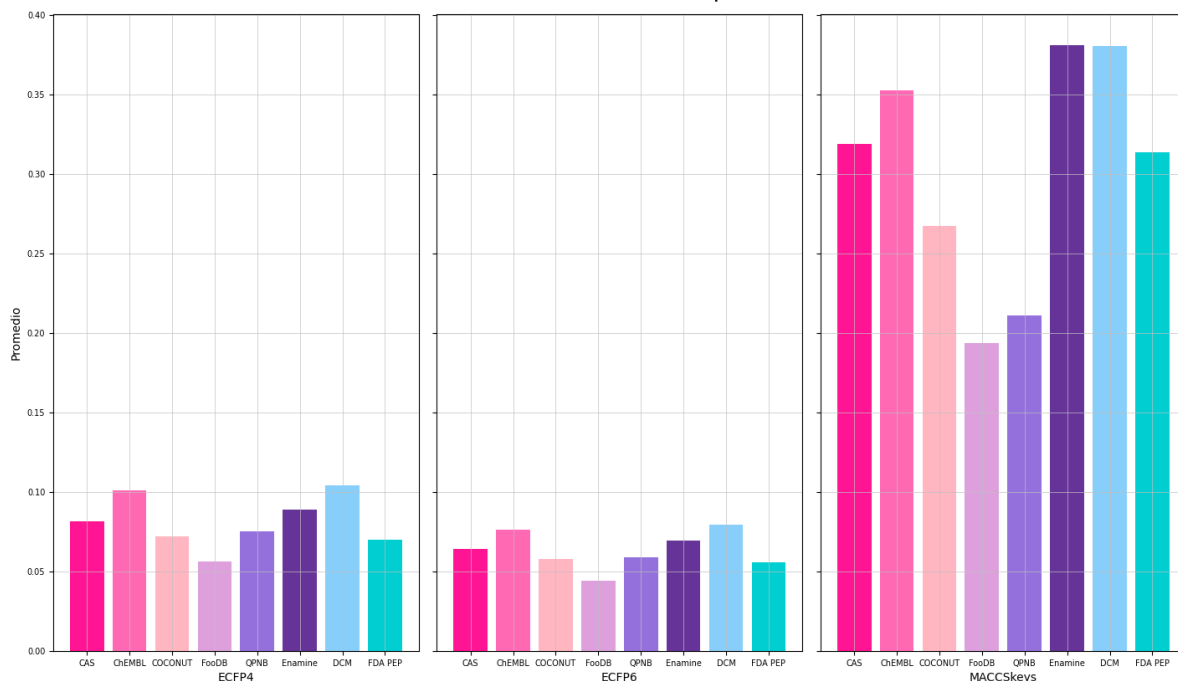


Figura S20. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs clorhidrato de clorpromazina.

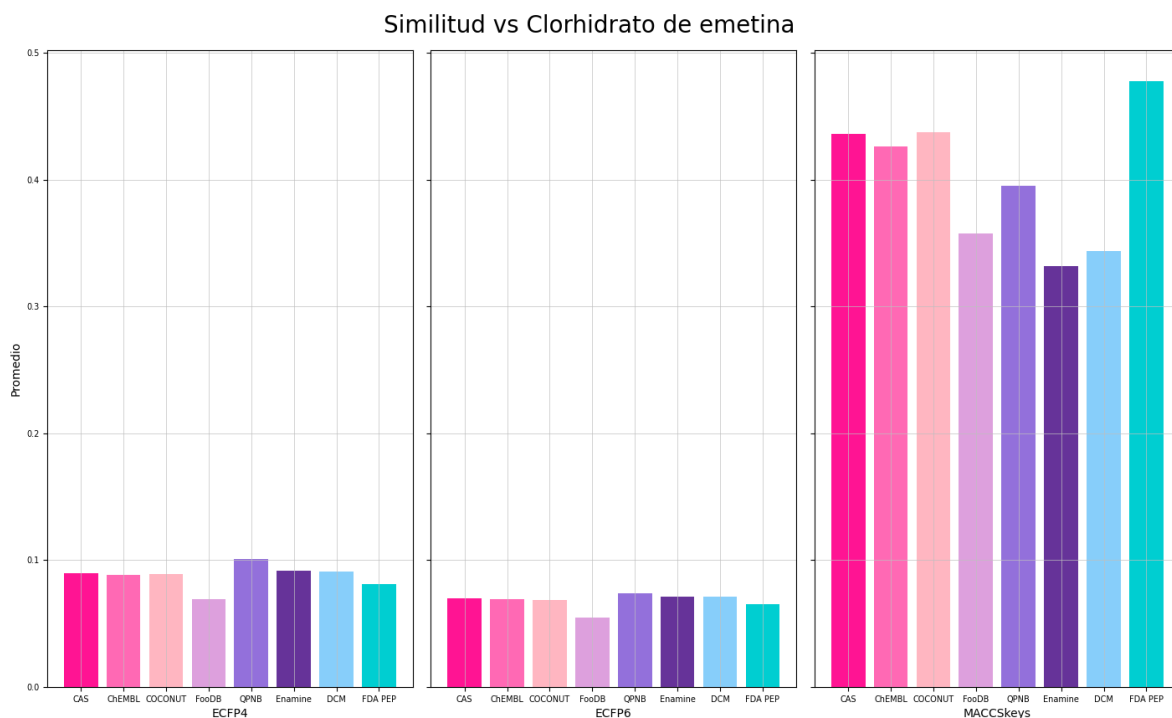


Figura S21. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs clorhidrato de emetina.

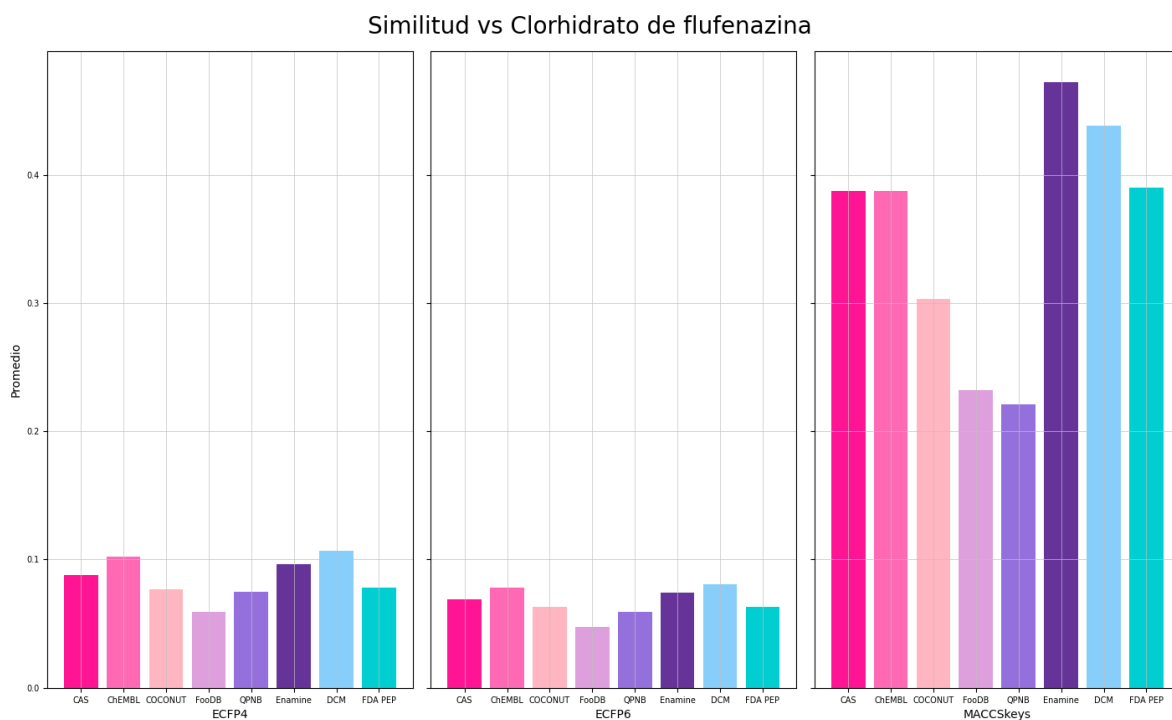


Figura S22. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs clorhidrato de flufenazina.

Similitud vs Clorhidrato de gemcitabina

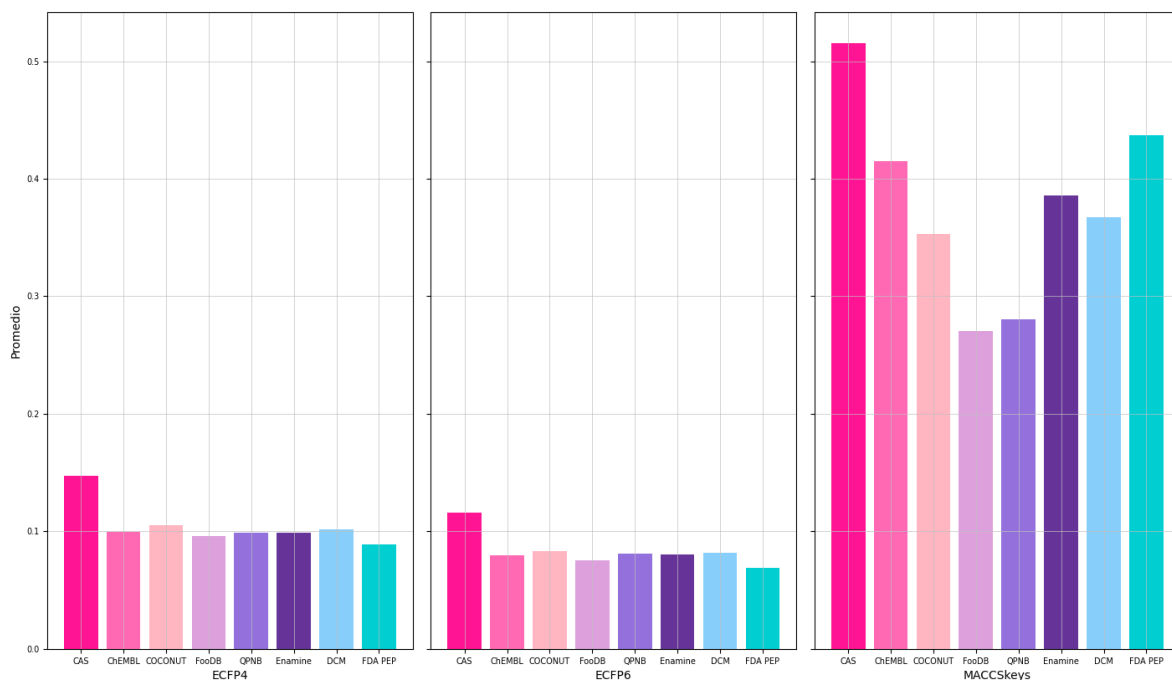


Figura S23. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs clorhidrato de gemcitabina.

Similitud vs Clorhidrato de prometazina

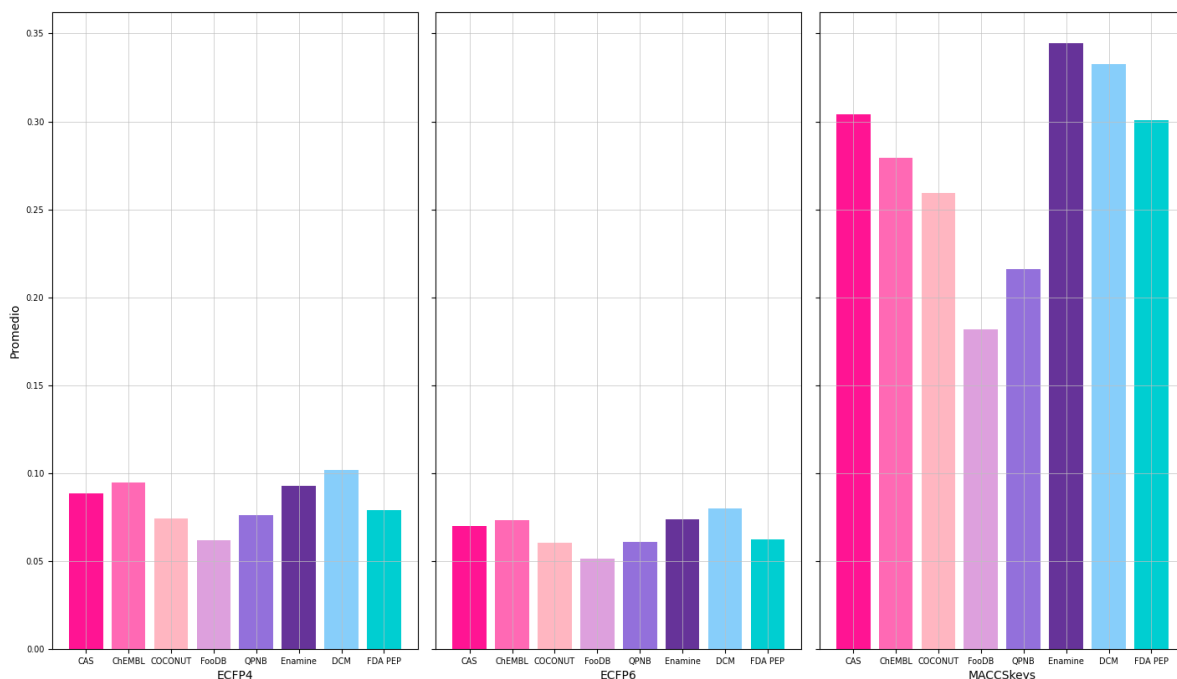


Figura S24. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs clorhidrato de prometazina.

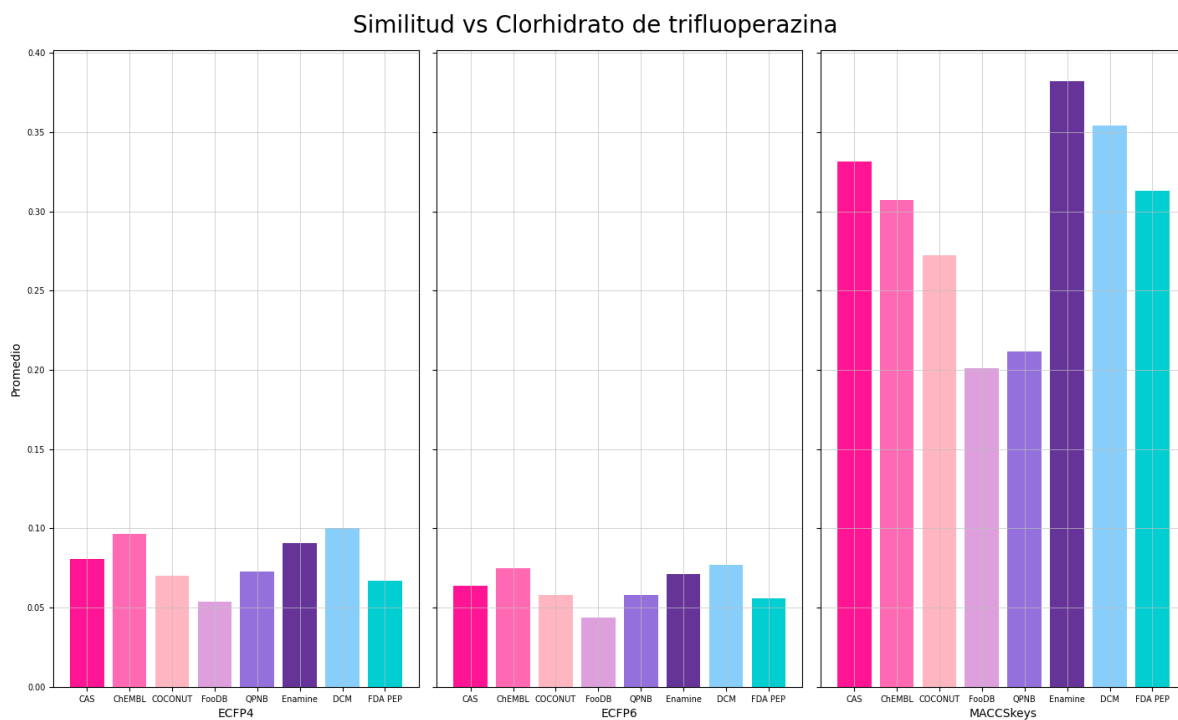


Figura S25. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs clorhidrato de trifluoperazina.

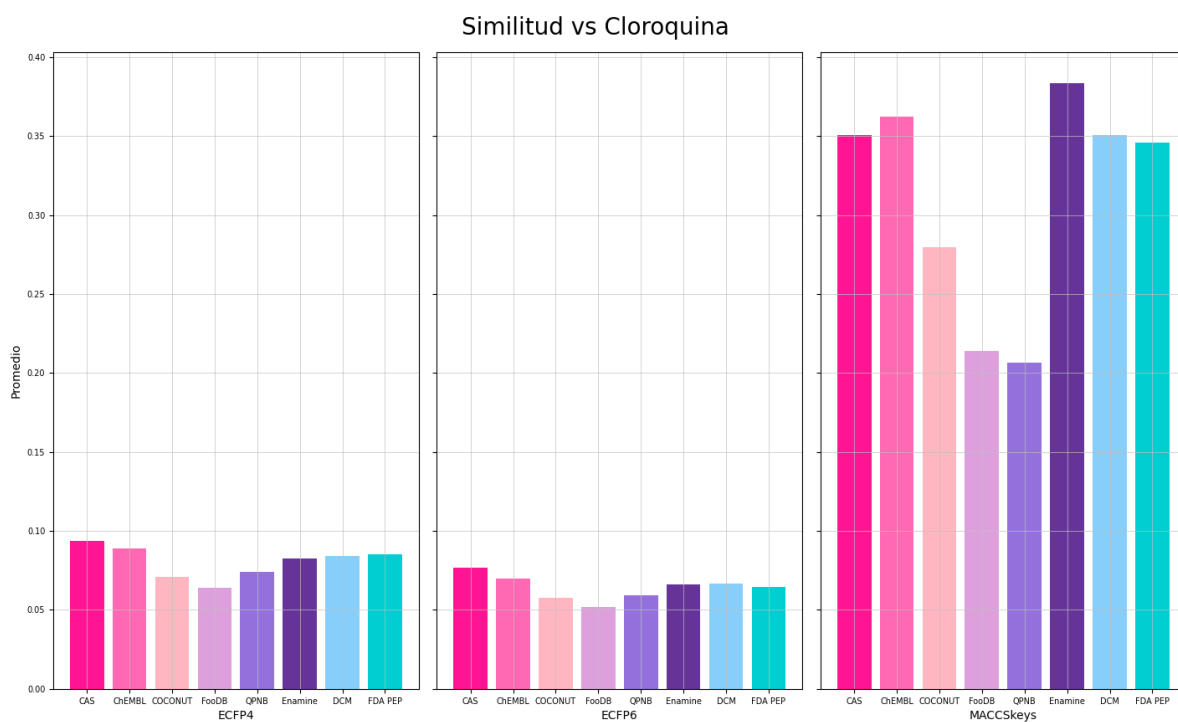


Figura S26. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs cloroquina.

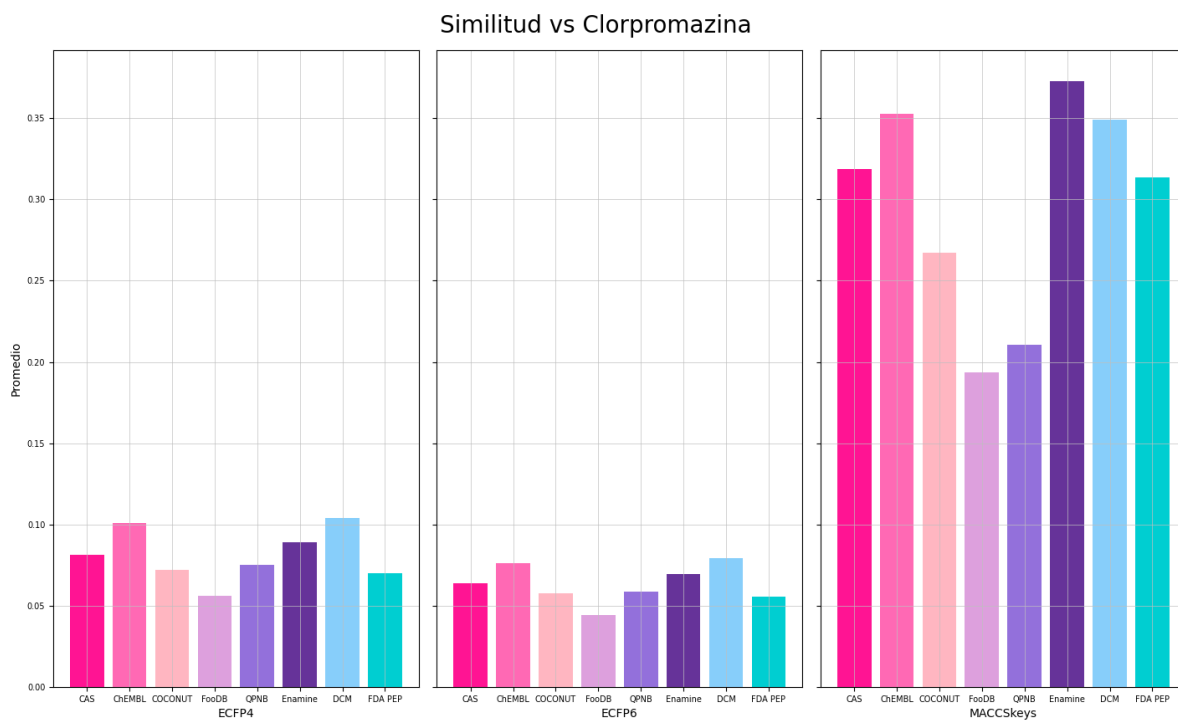


Figura S27. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs clorpromazina.

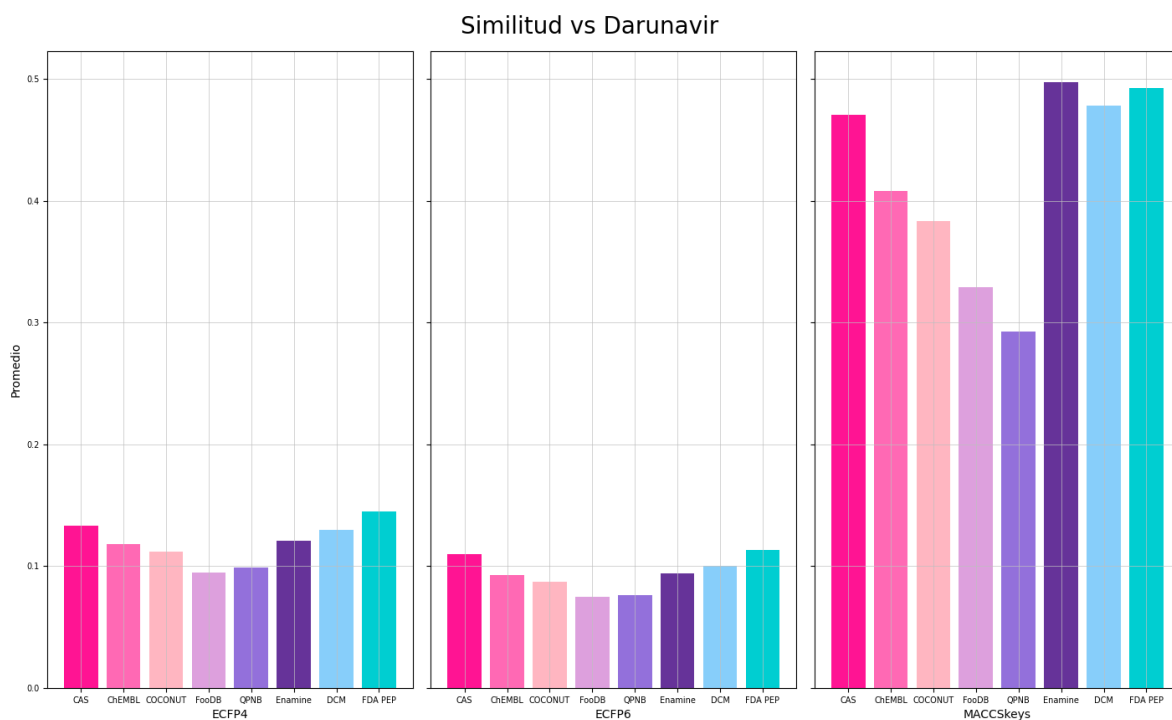


Figura S28. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs darunavir.

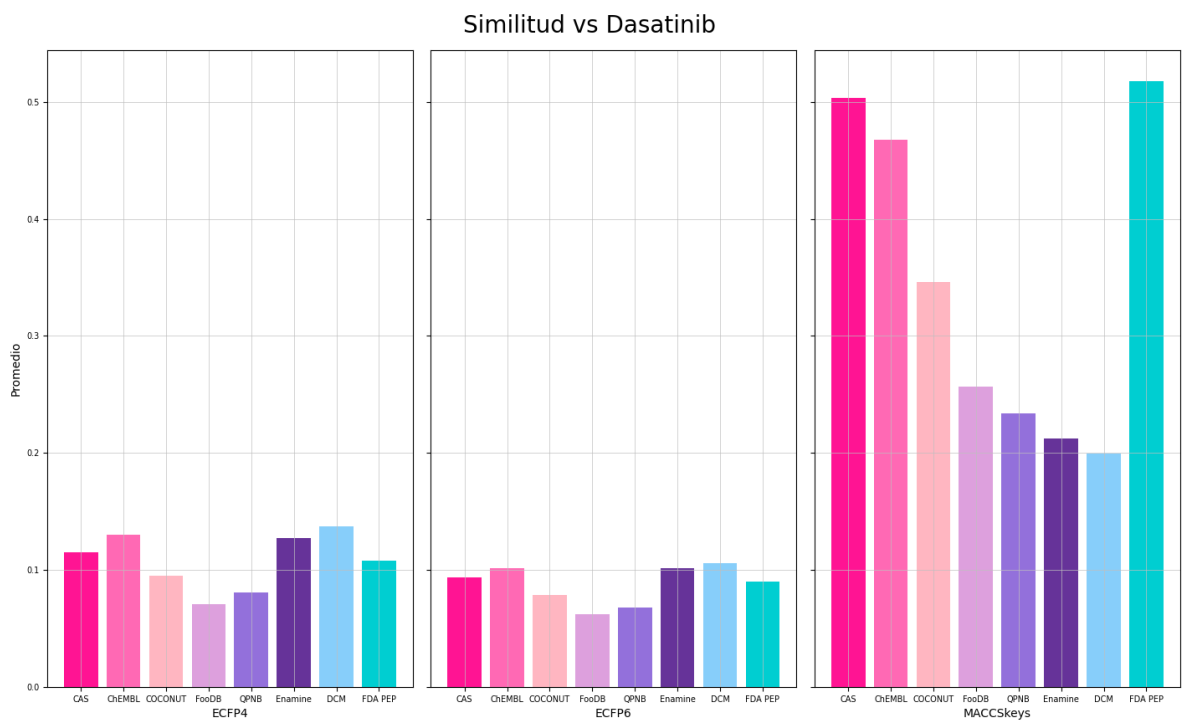


Figura S29. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs dasatinib.

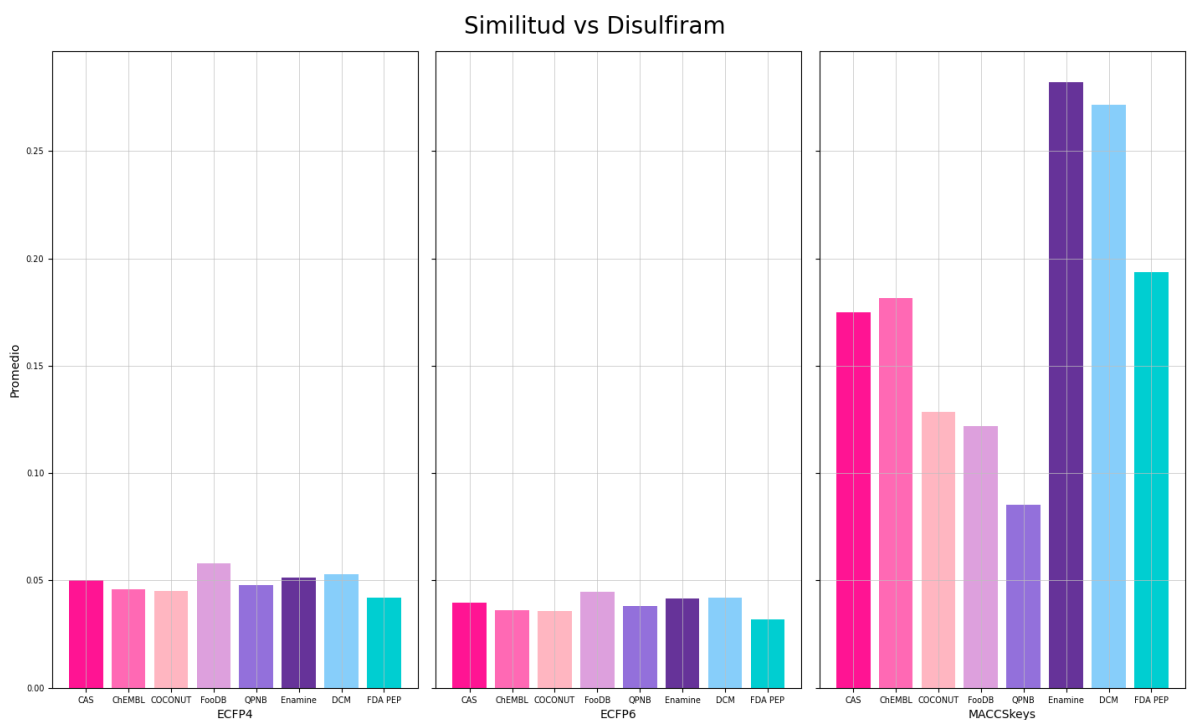


Figura S30. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs disulfiram.

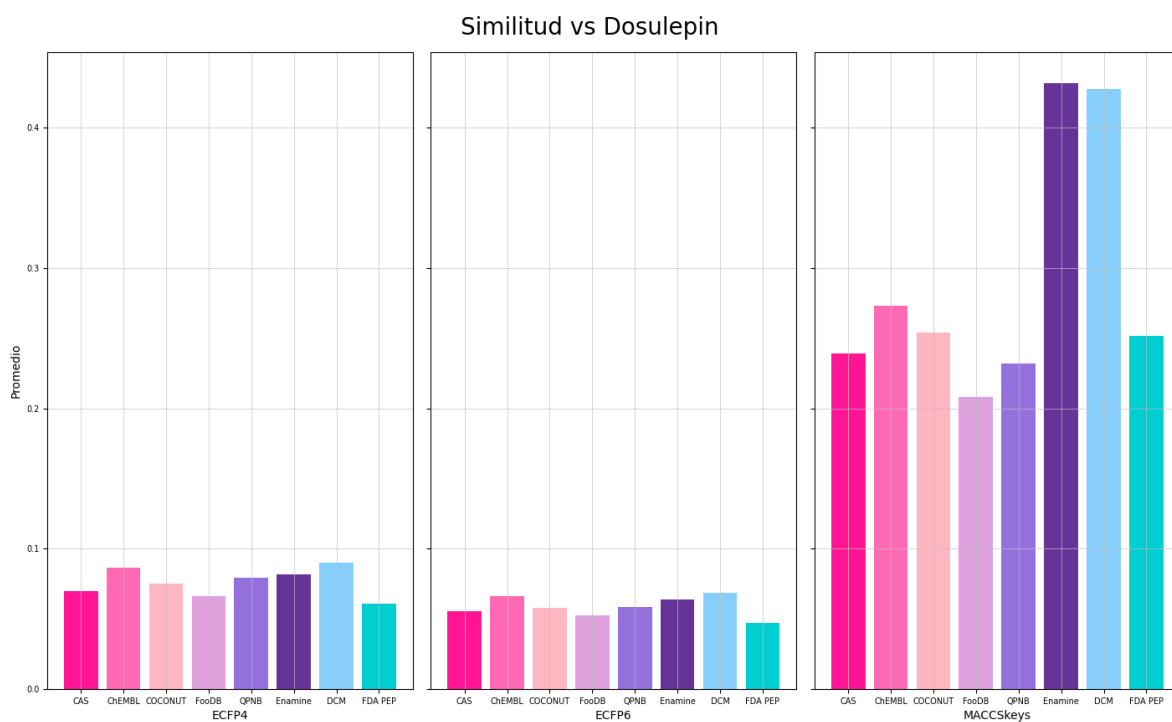


Figura S31. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs dosulepin.

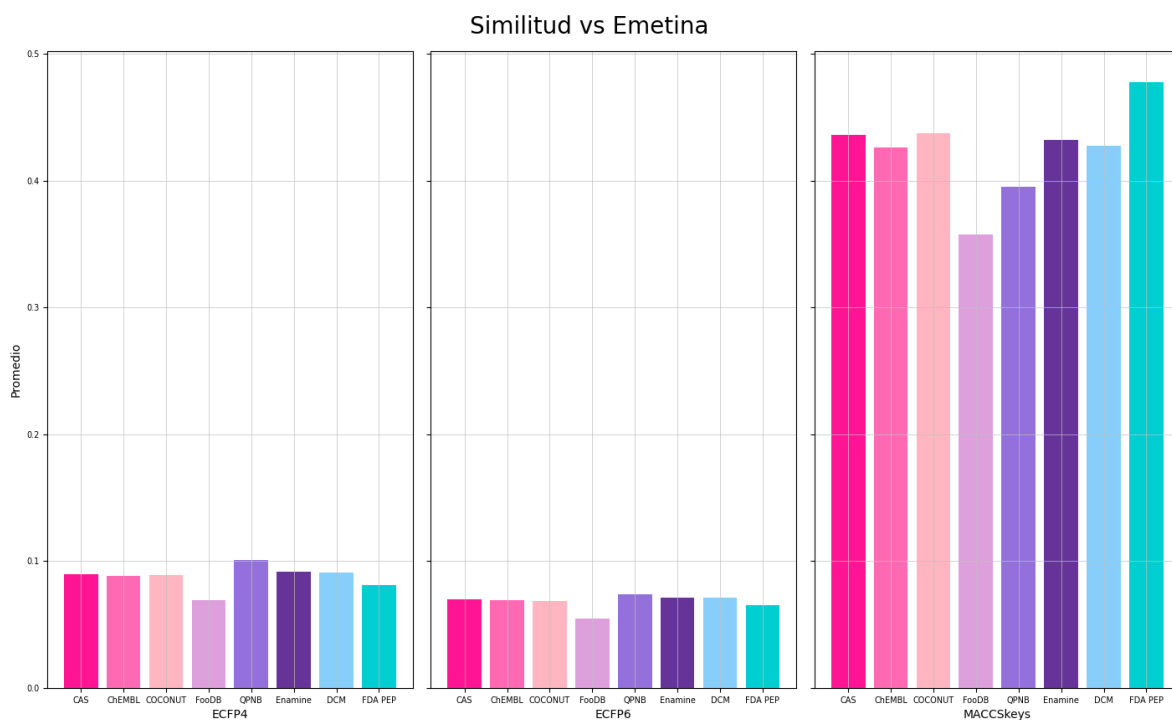


Figura S32. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs emetina.

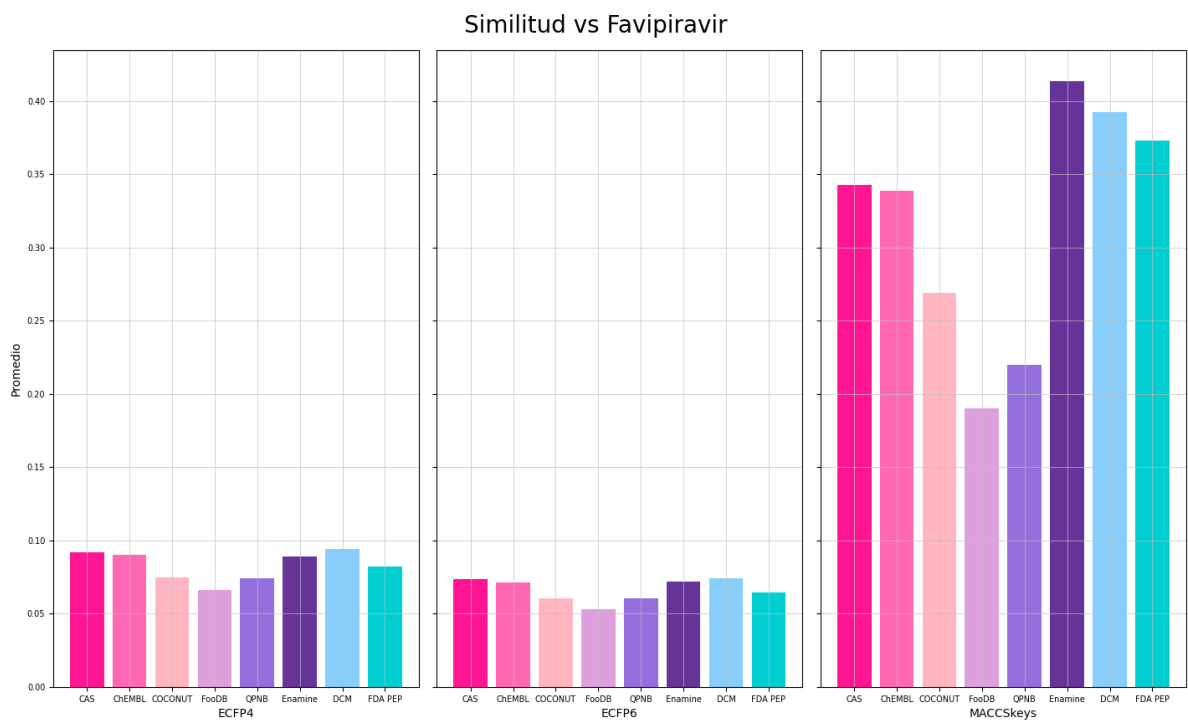


Figura S33. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs favipiravir.

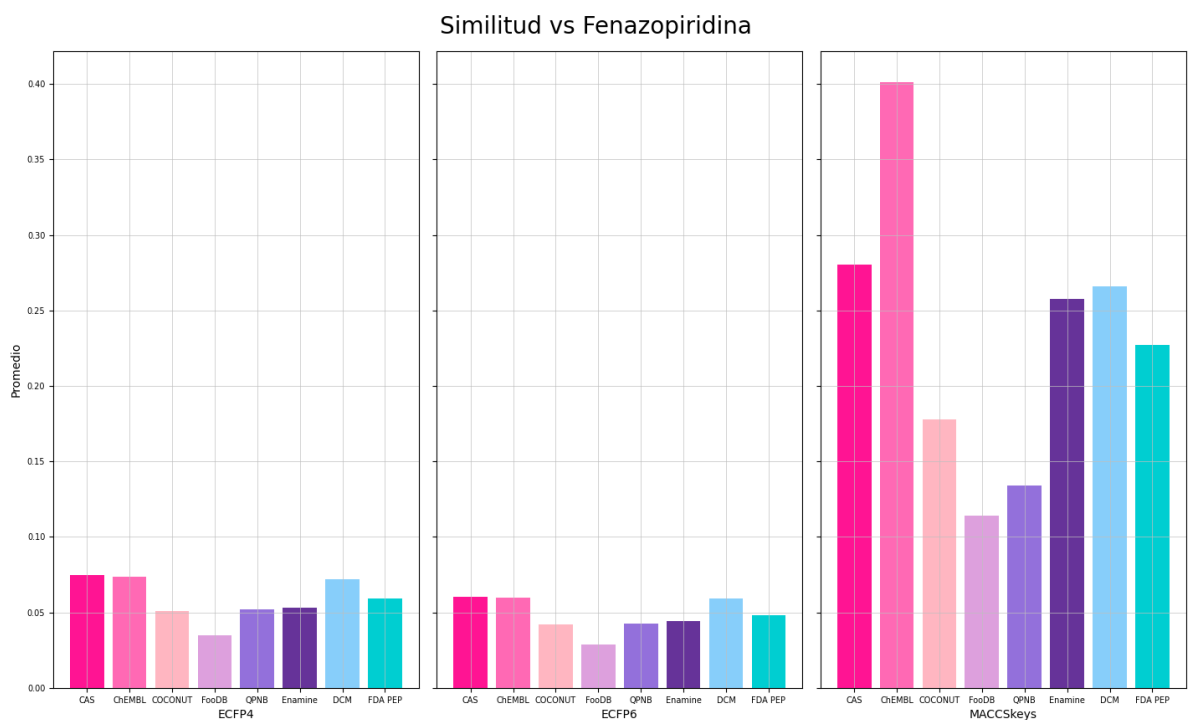


Figura S34. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs fenazopiridina.

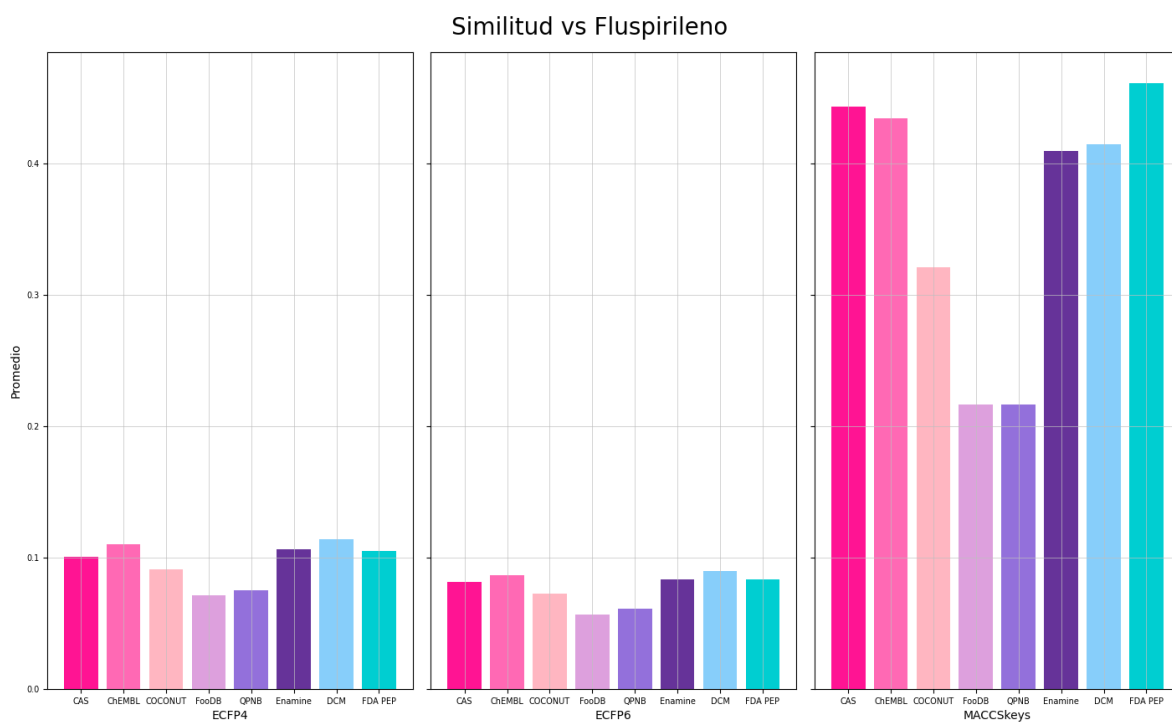


Figura S35. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs fluspirileno.

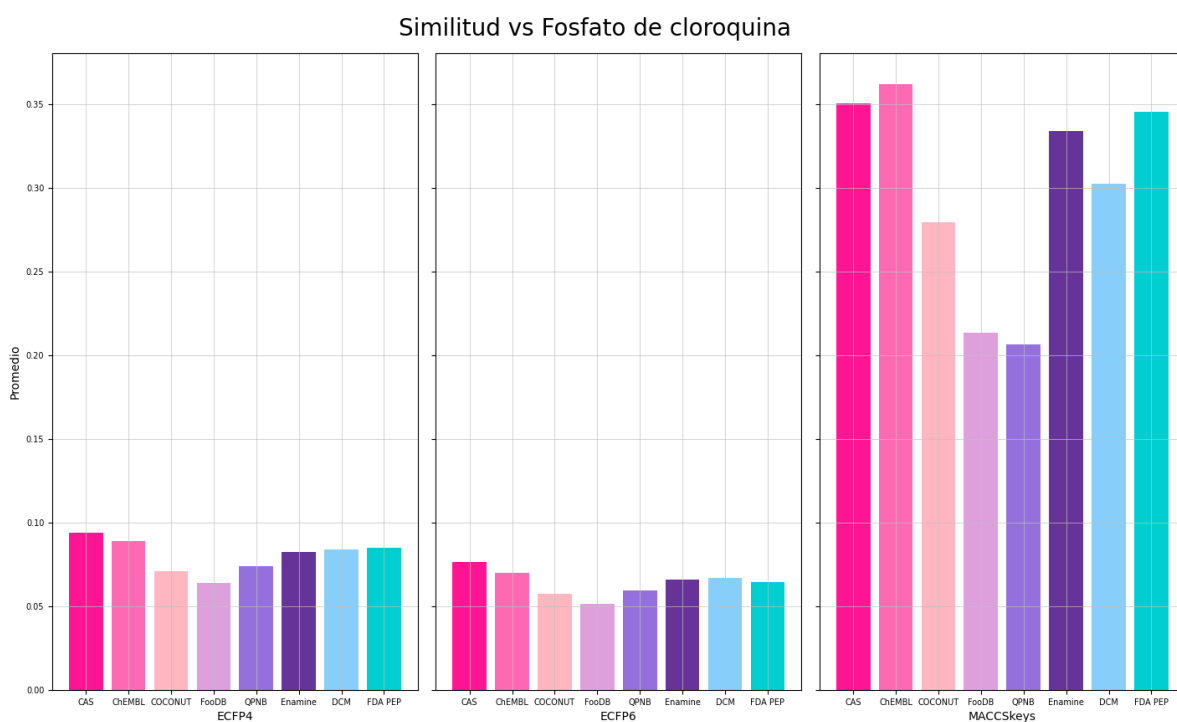


Figura S36. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs fosfato de cloroquina.

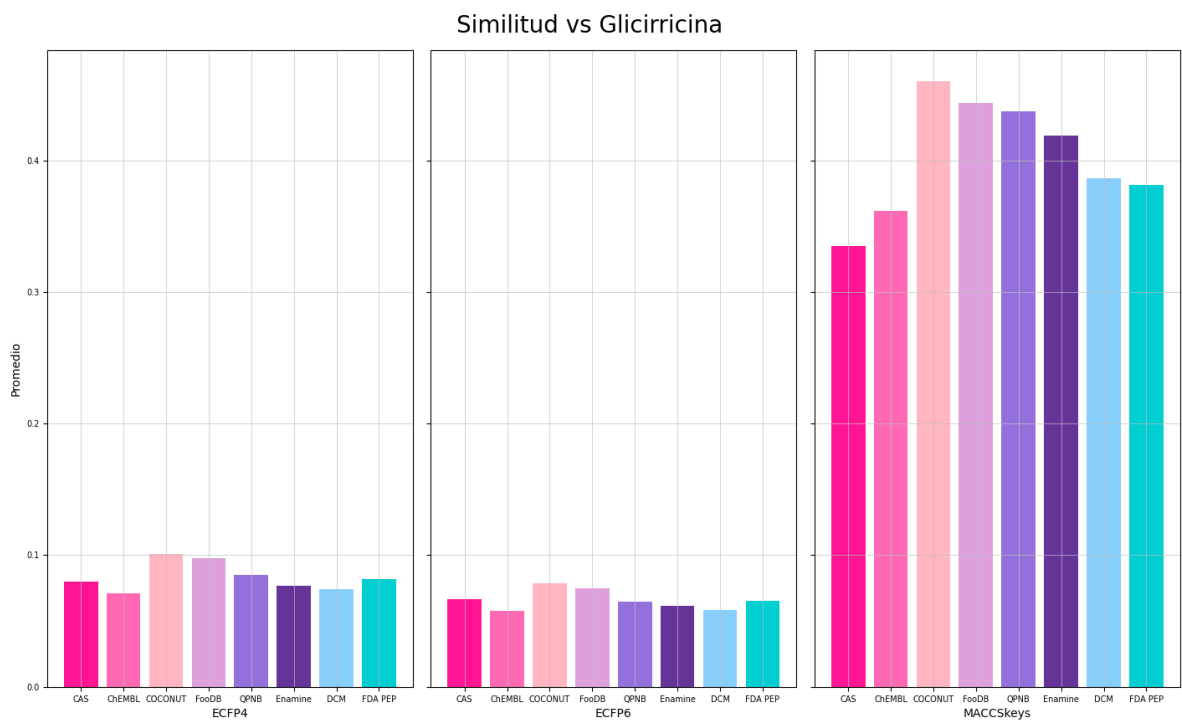


Figura S37. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs glicirricina.

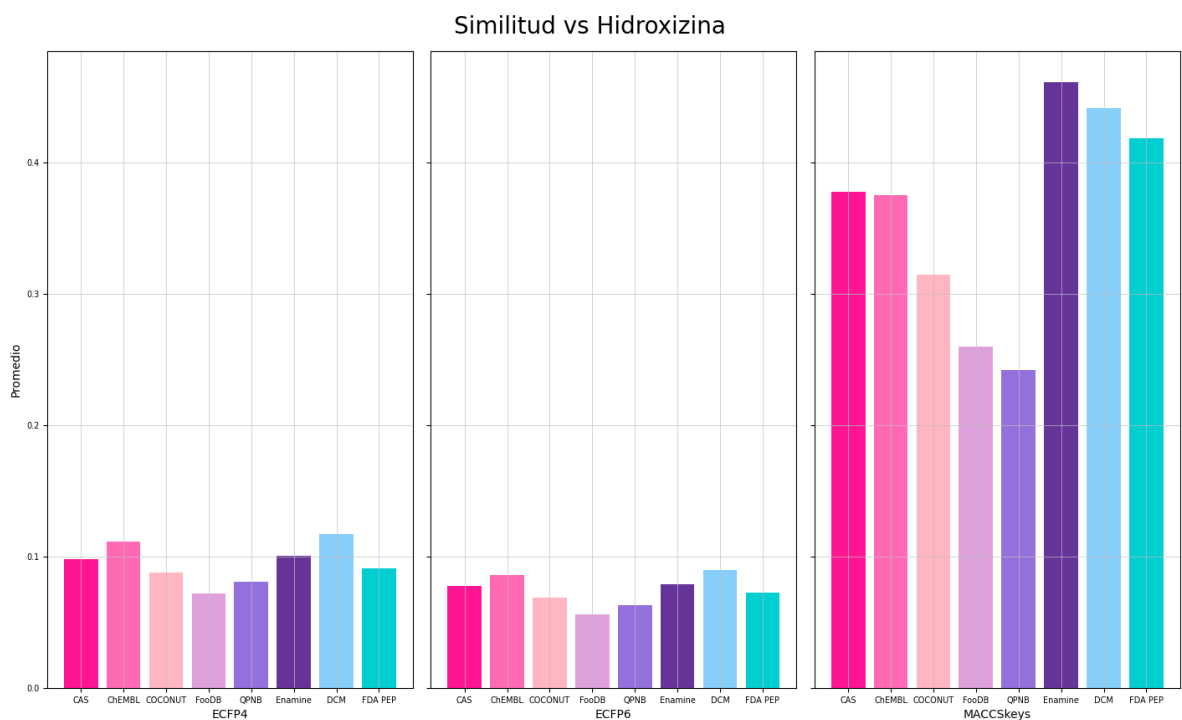


Figura S38. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs hidroxizina.

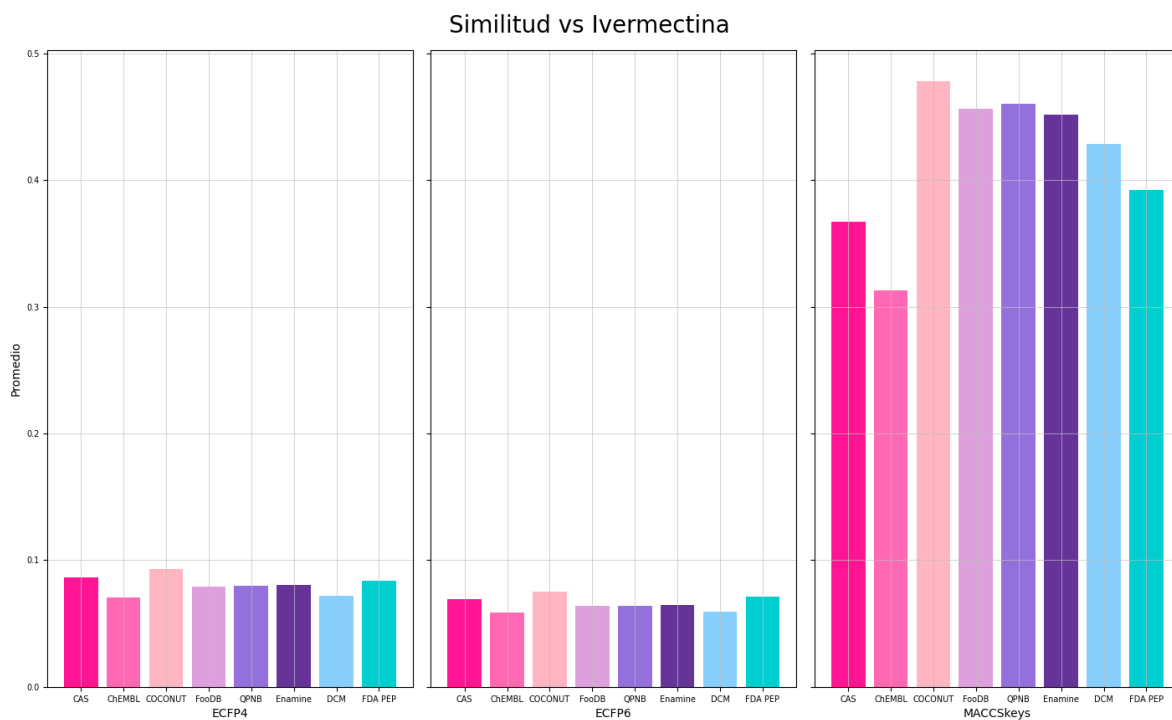


Figura S39. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs ivermectina.

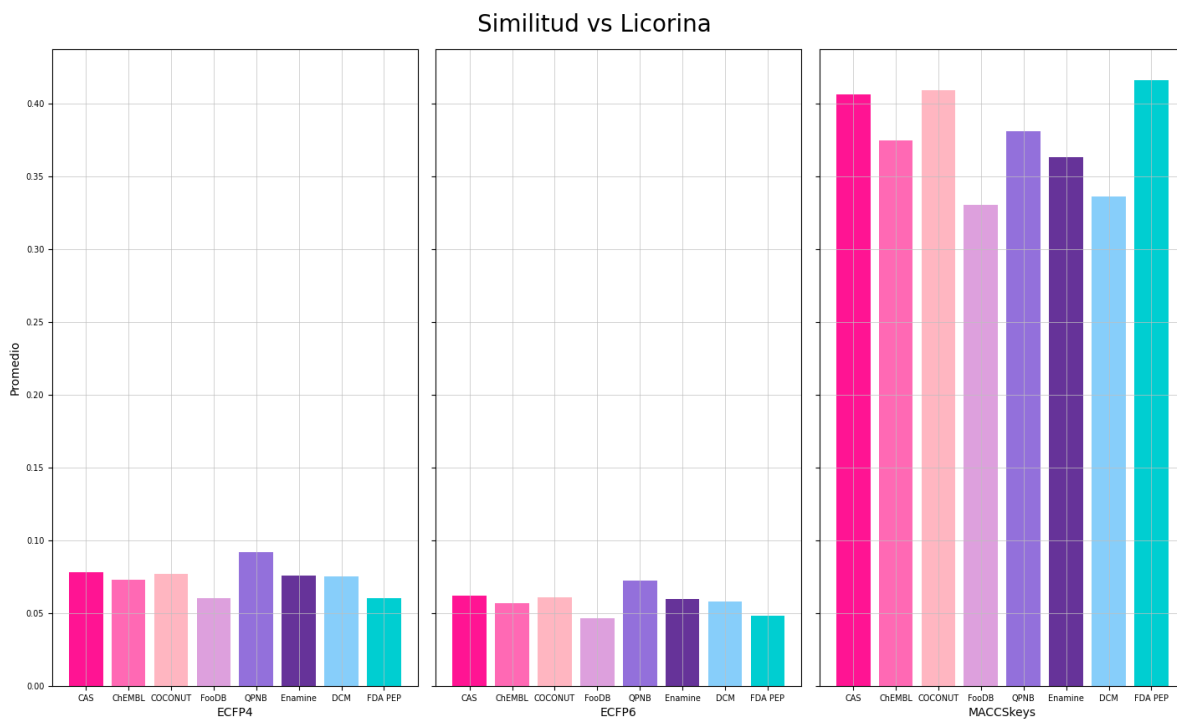


Figura S40. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs licorina.

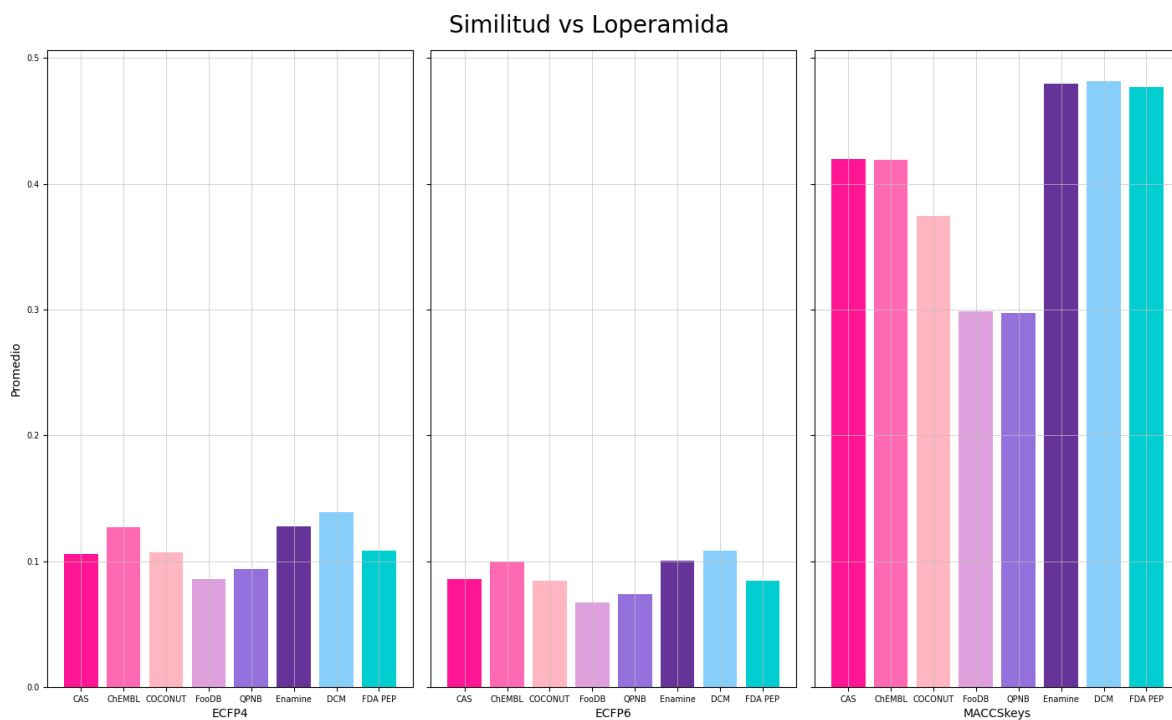


Figura S41. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs loperamida.

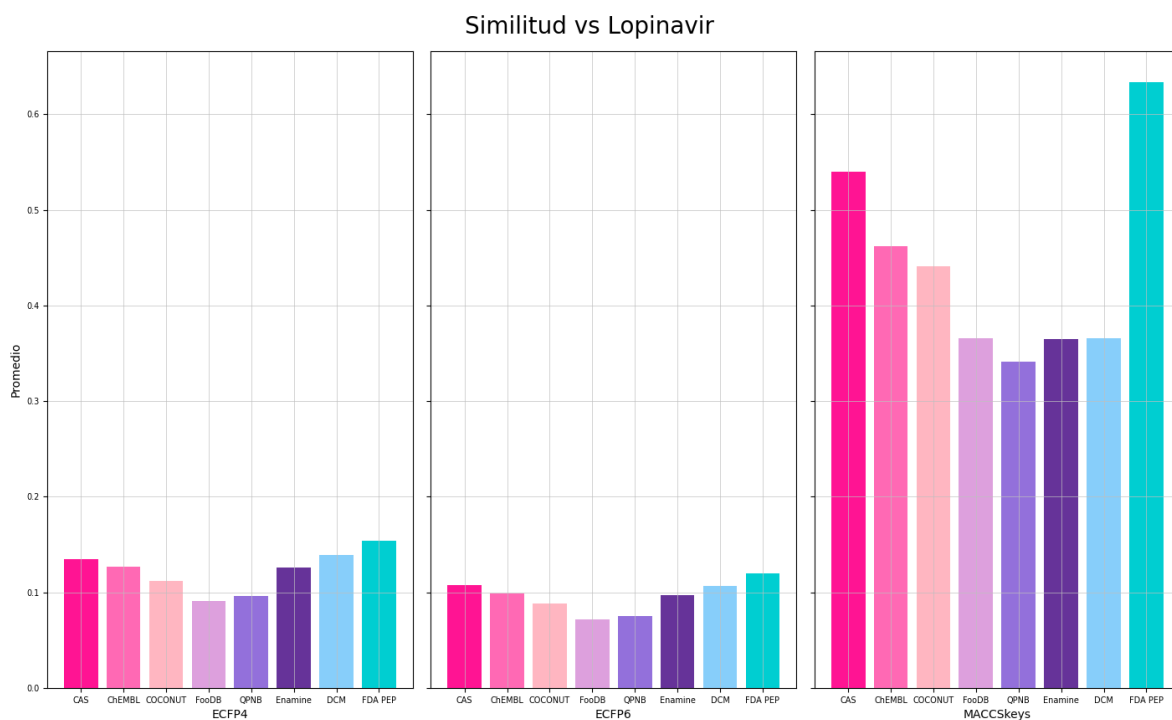


Figura S42. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs lopinavir.

Similitud vs Maleato de tietilperazina

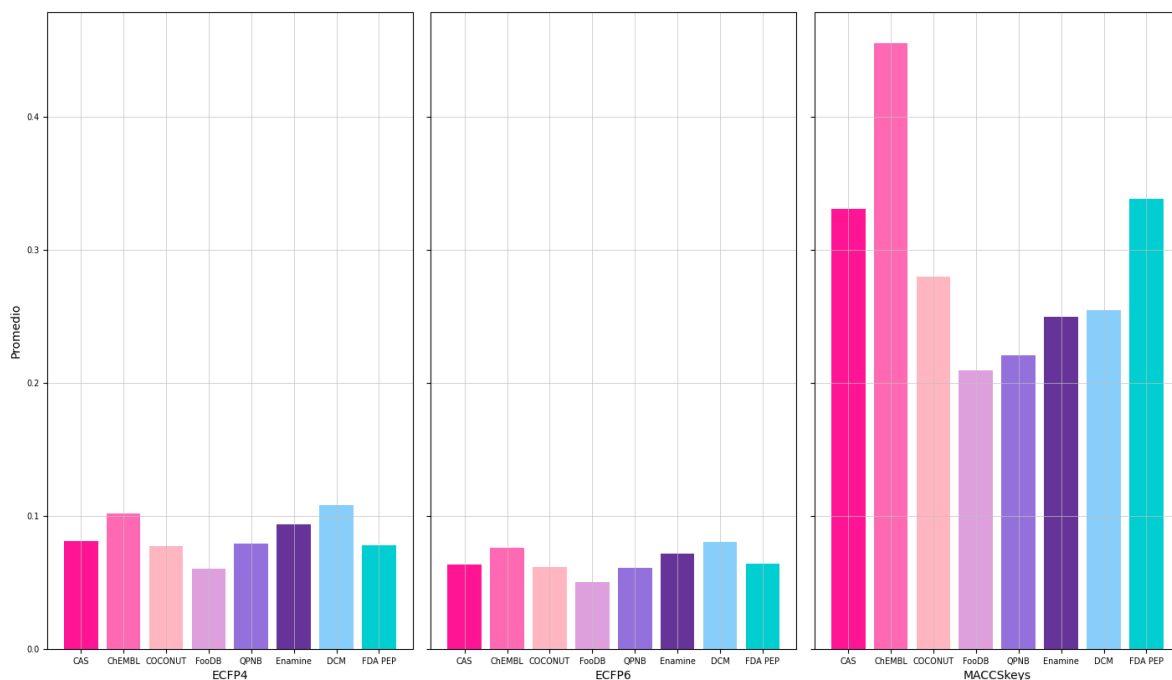


Figura S43. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs maleato de tietilperazina.

Similitud vs Mefloquina

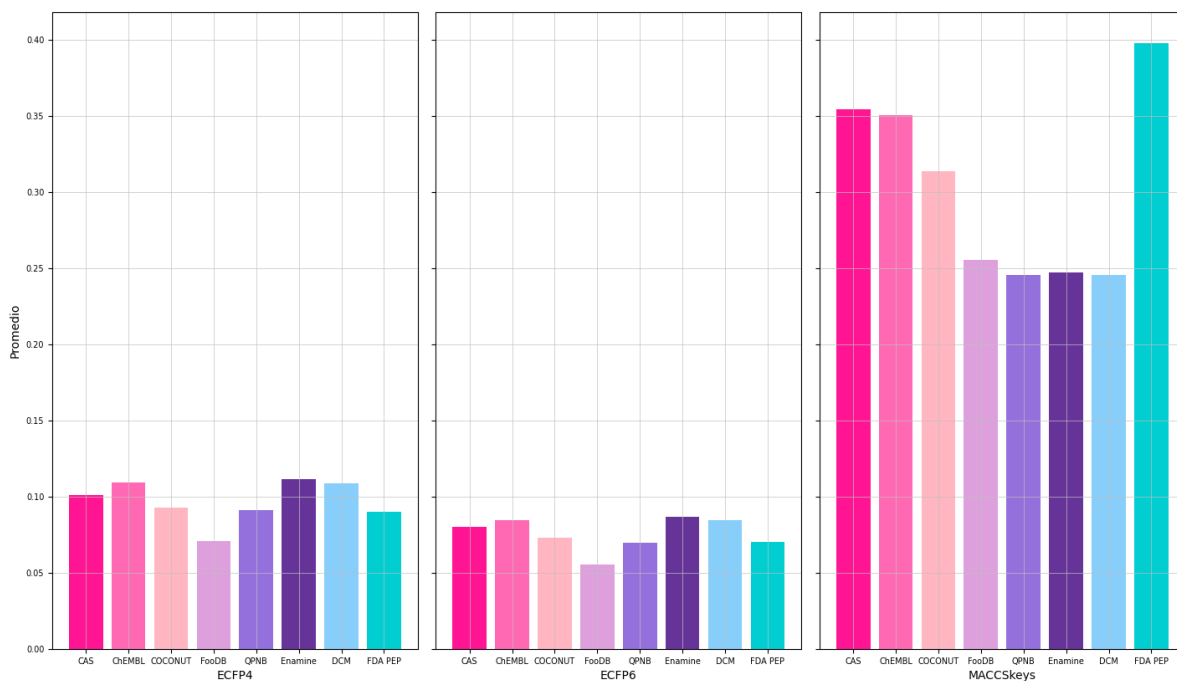


Figura S44. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs mefloquina.

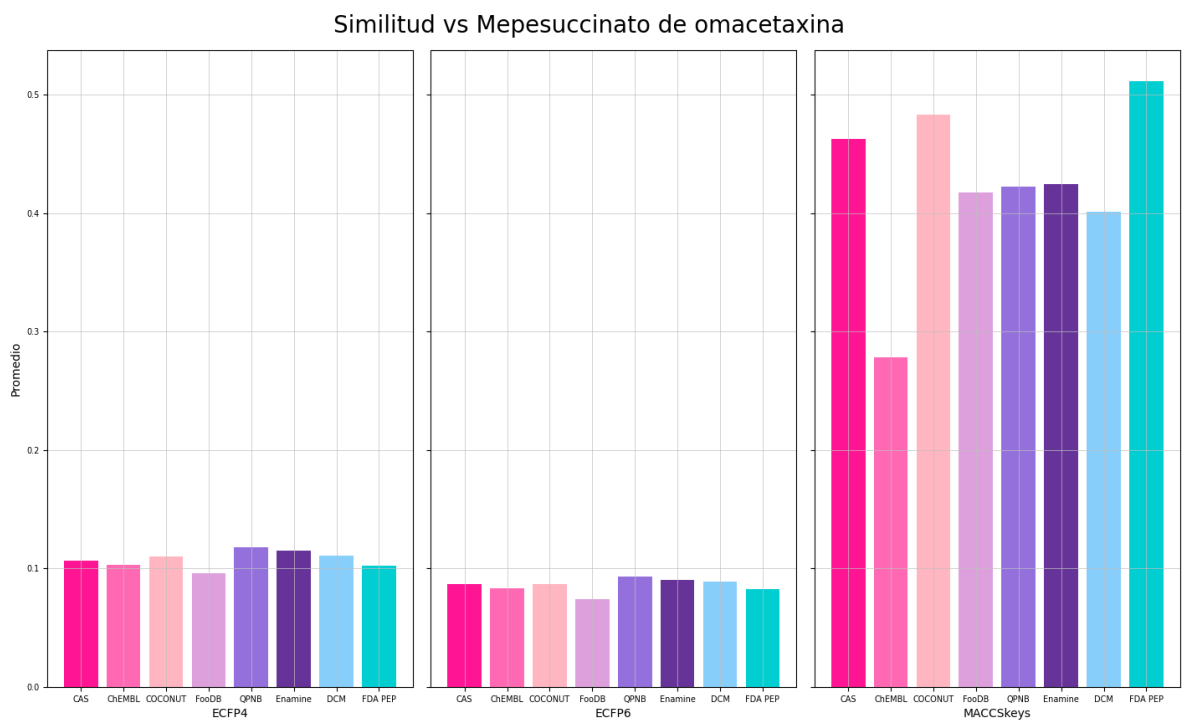


Figura S45. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs mepesuccinato de omacetaxina.

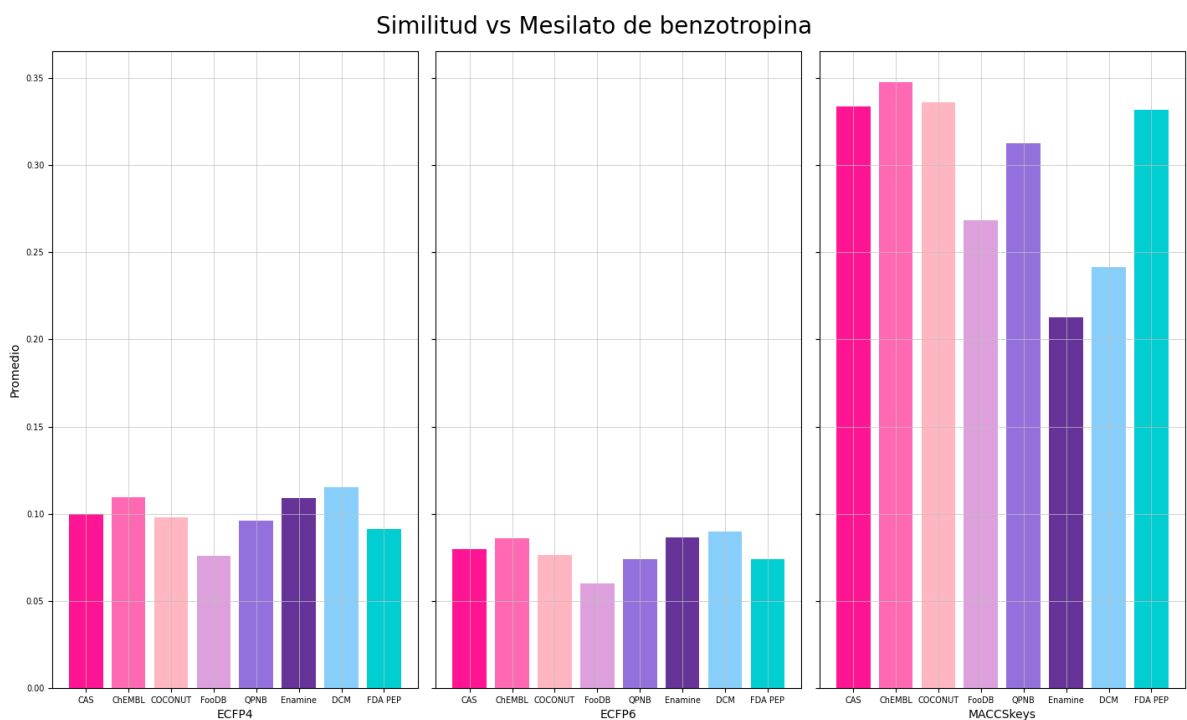


Figura S46. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs mesilato de benzotropina.

Similitud vs Mesilato de imatinib

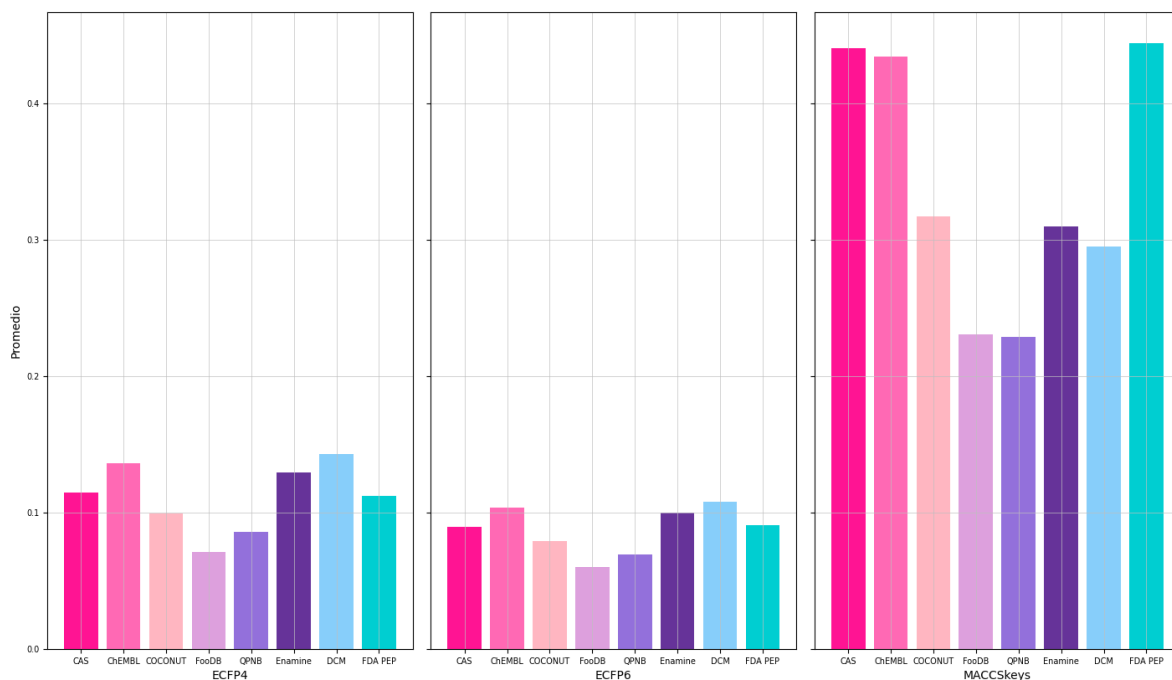


Figura S47. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs mesilato de imatinib.

Similitud vs Metotrimeprazina

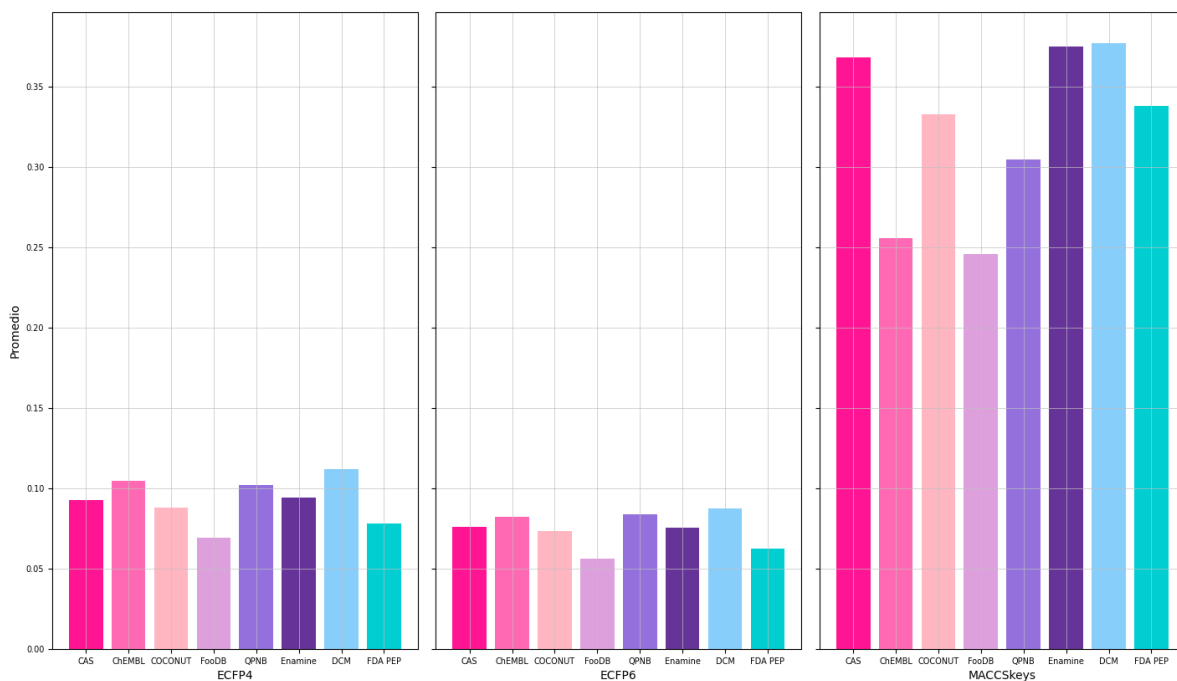


Figura S48. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs metotrimeprazina.

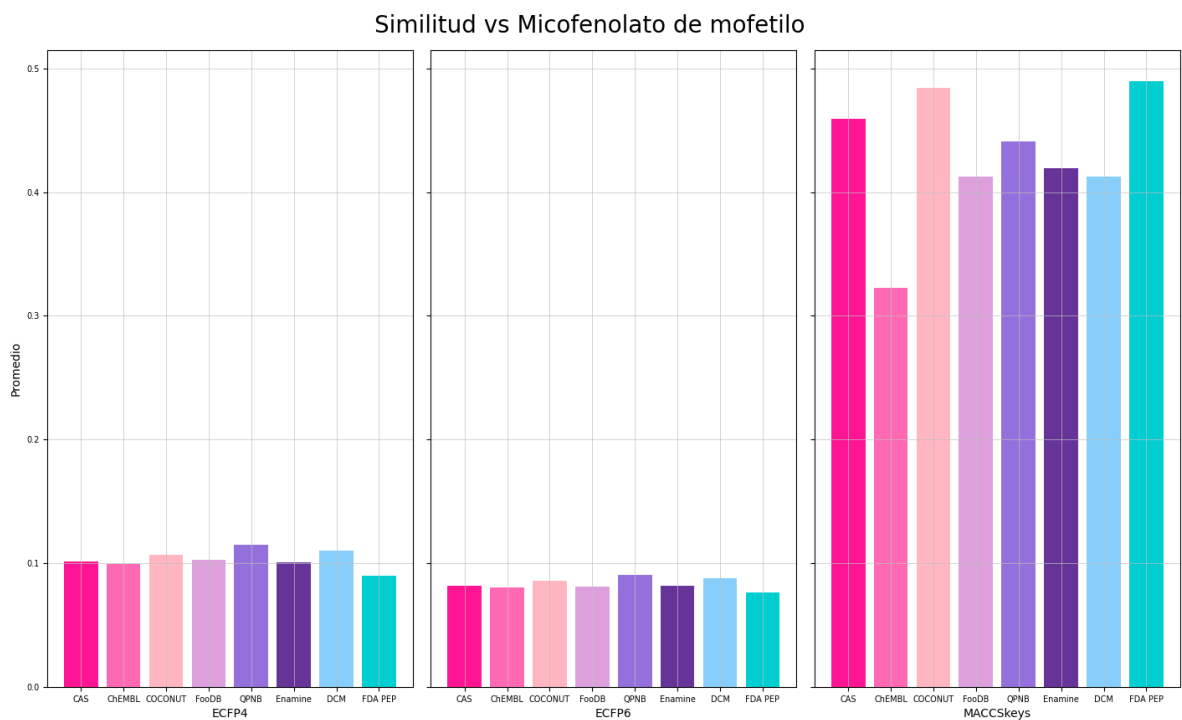


Figura S49. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs micofenolato de mofetilo.

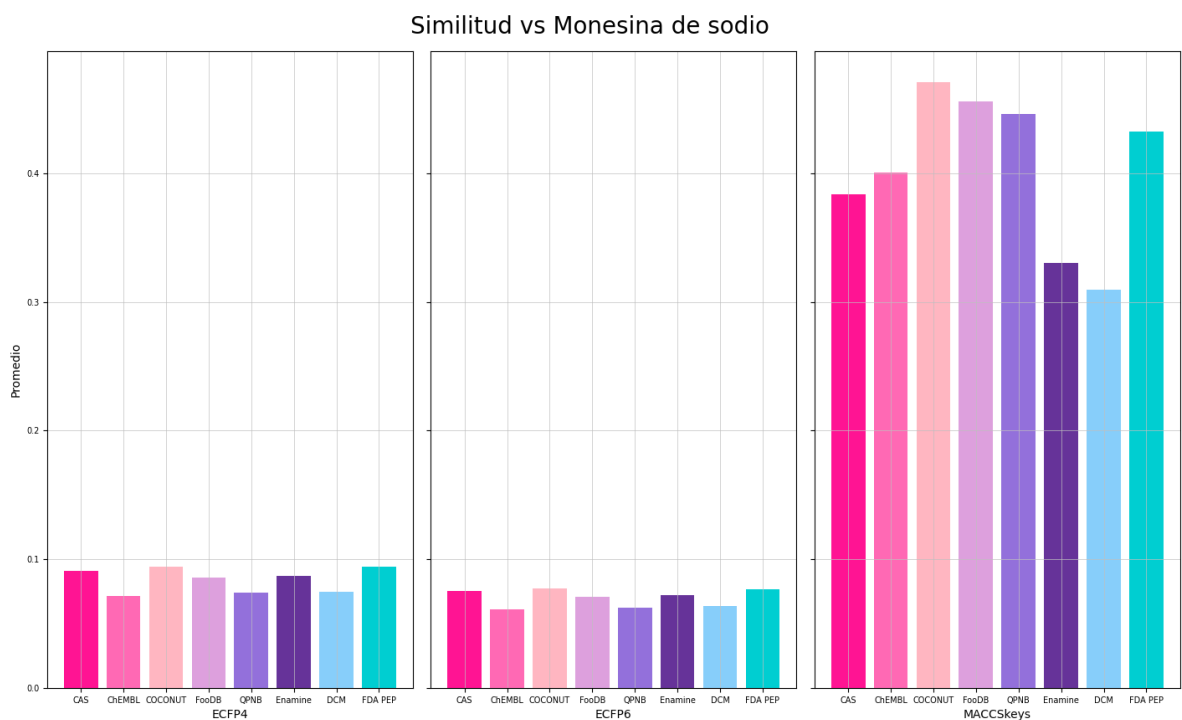


Figura S50. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs monesina de sodio.

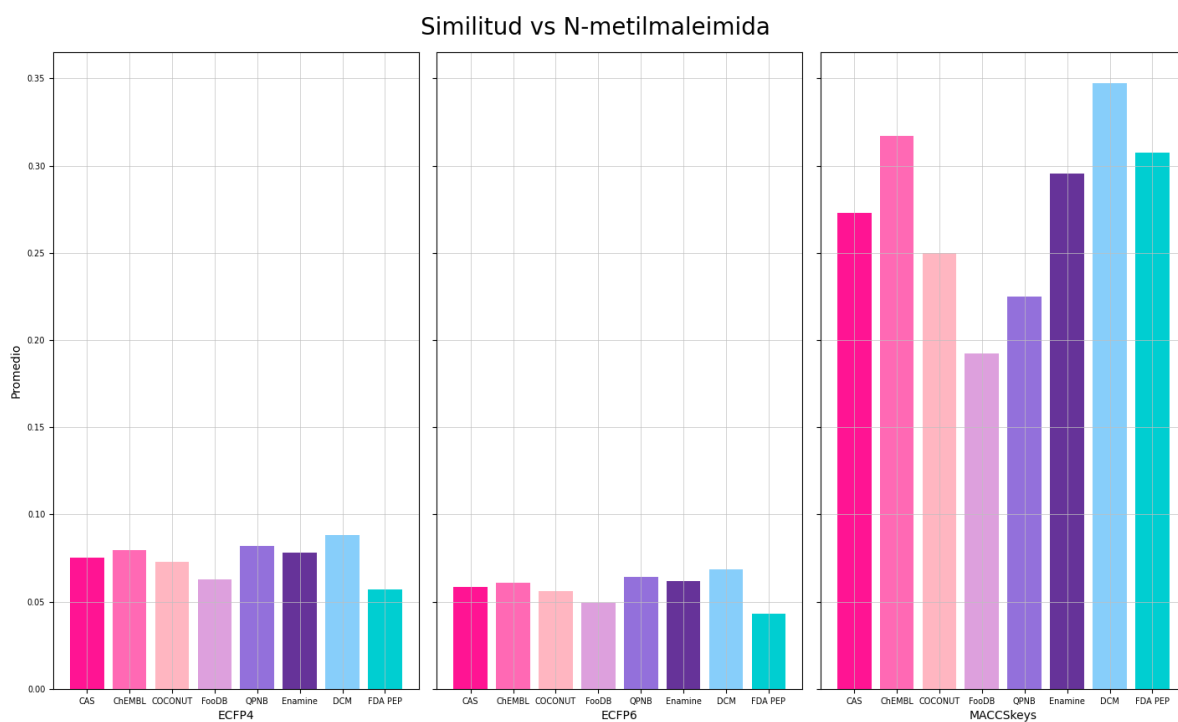


Figura S51. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs N-metilmaleimida.

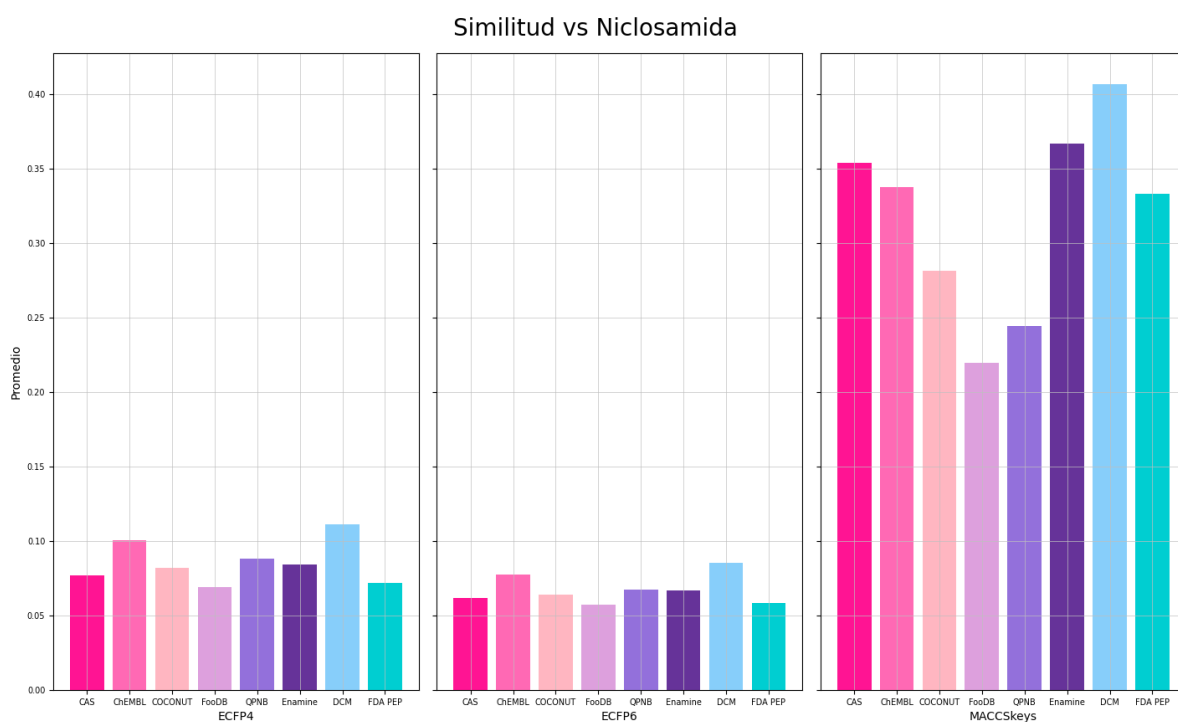


Figura S52. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs niclosamida.

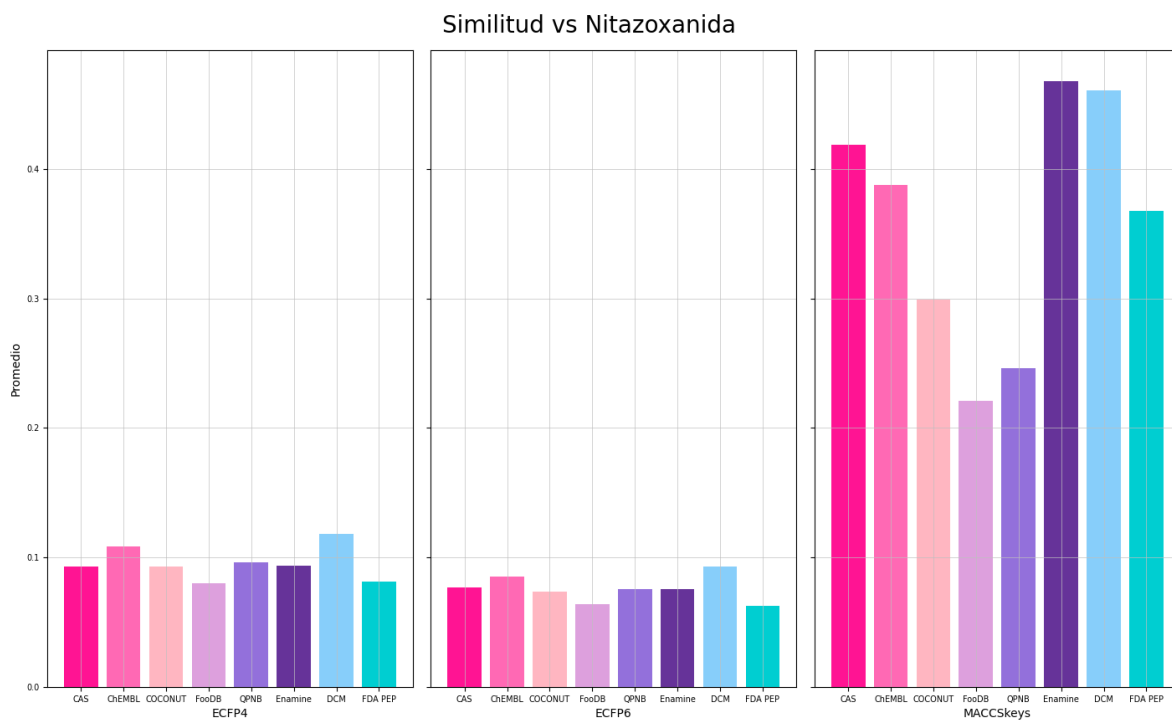


Figura S53. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs nitazoxanida.

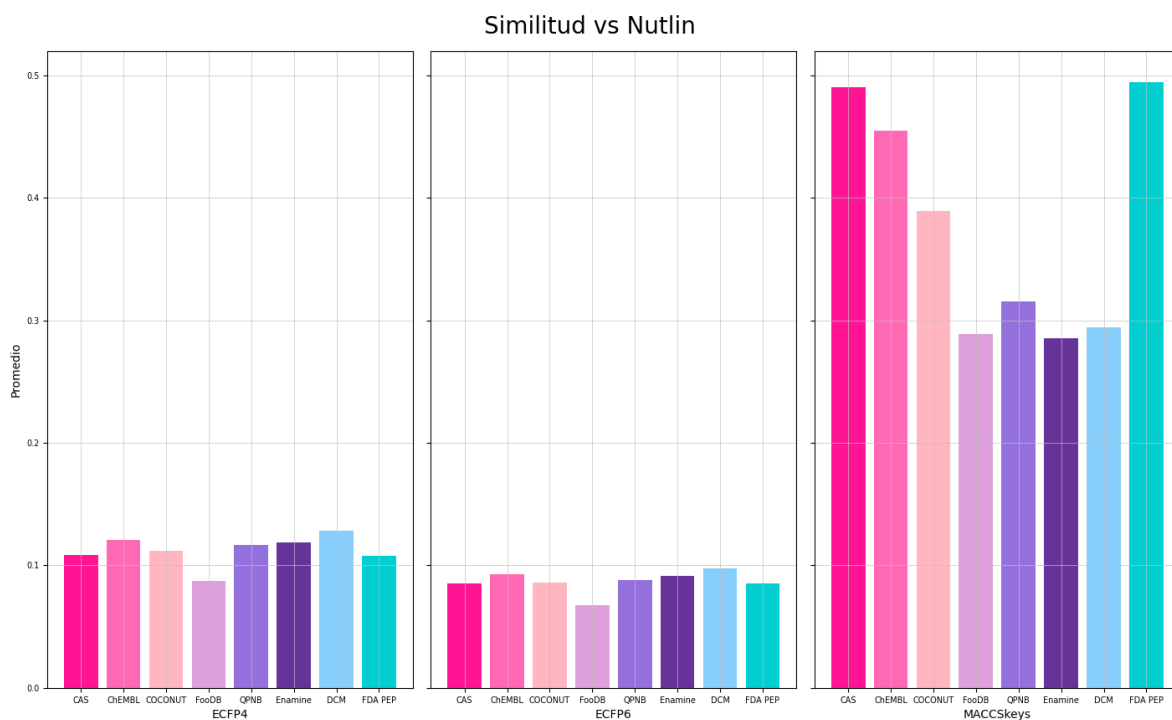


Figura S54. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs nutlin.

Similitud vs Pamoato de pirvinio

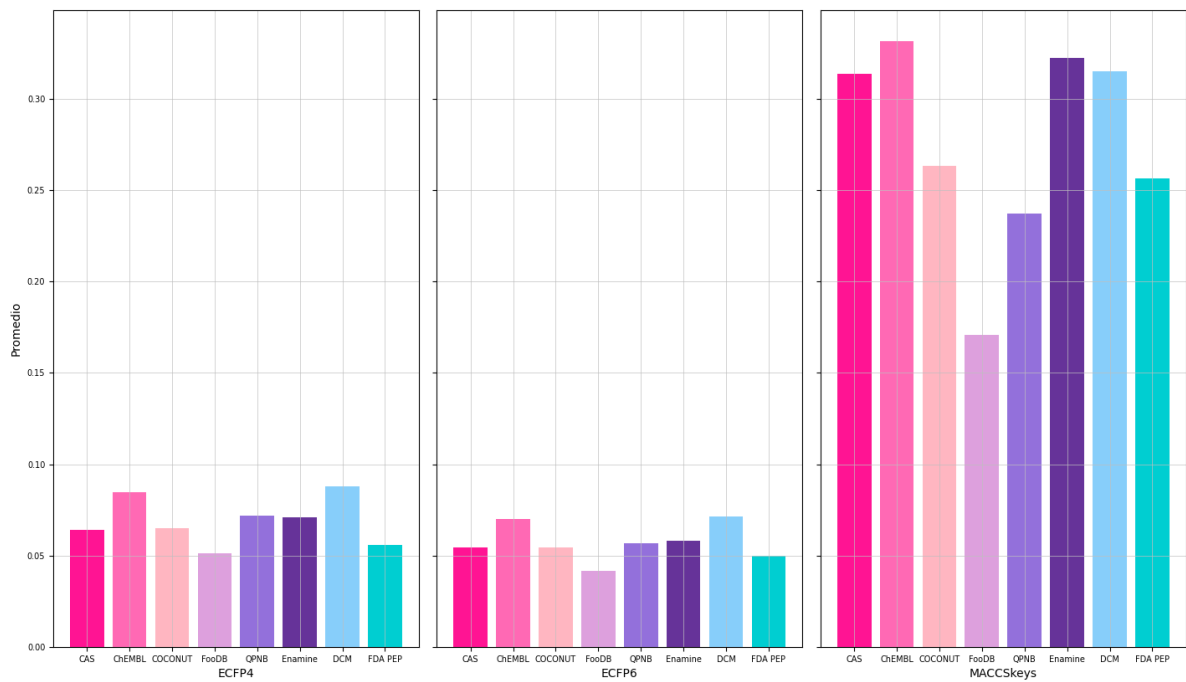


Figura S55. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs pamoato de pirvinio.

Similitud vs Remdesivir

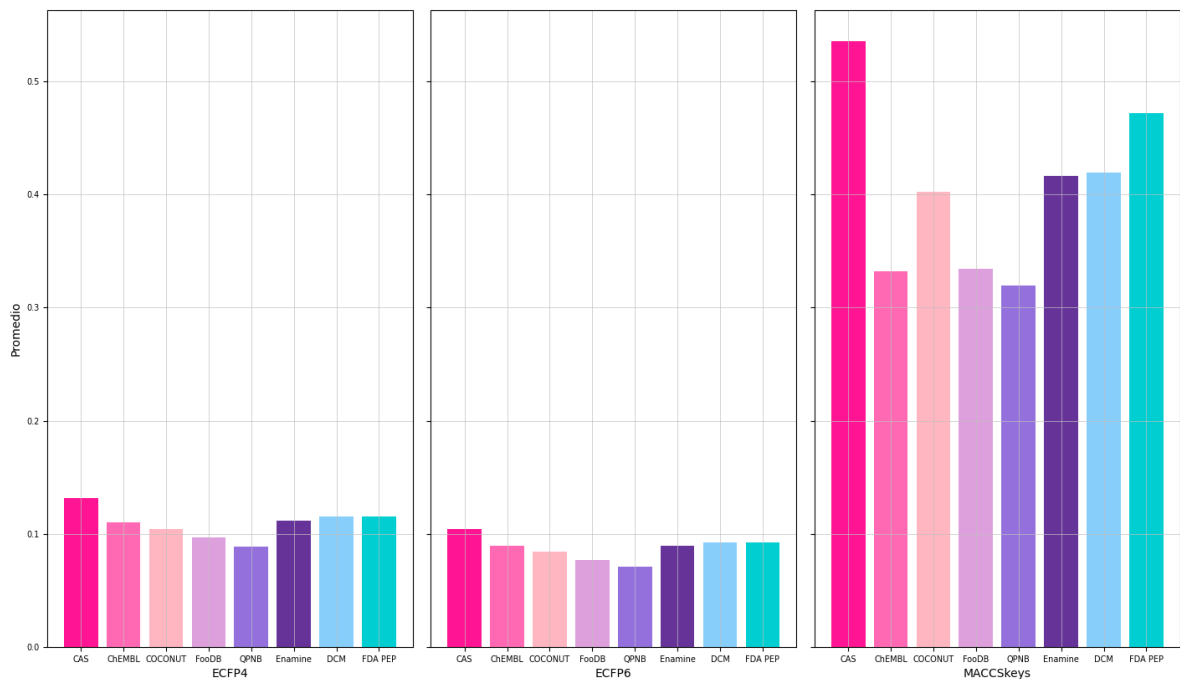


Figura S56. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs remdesivir.

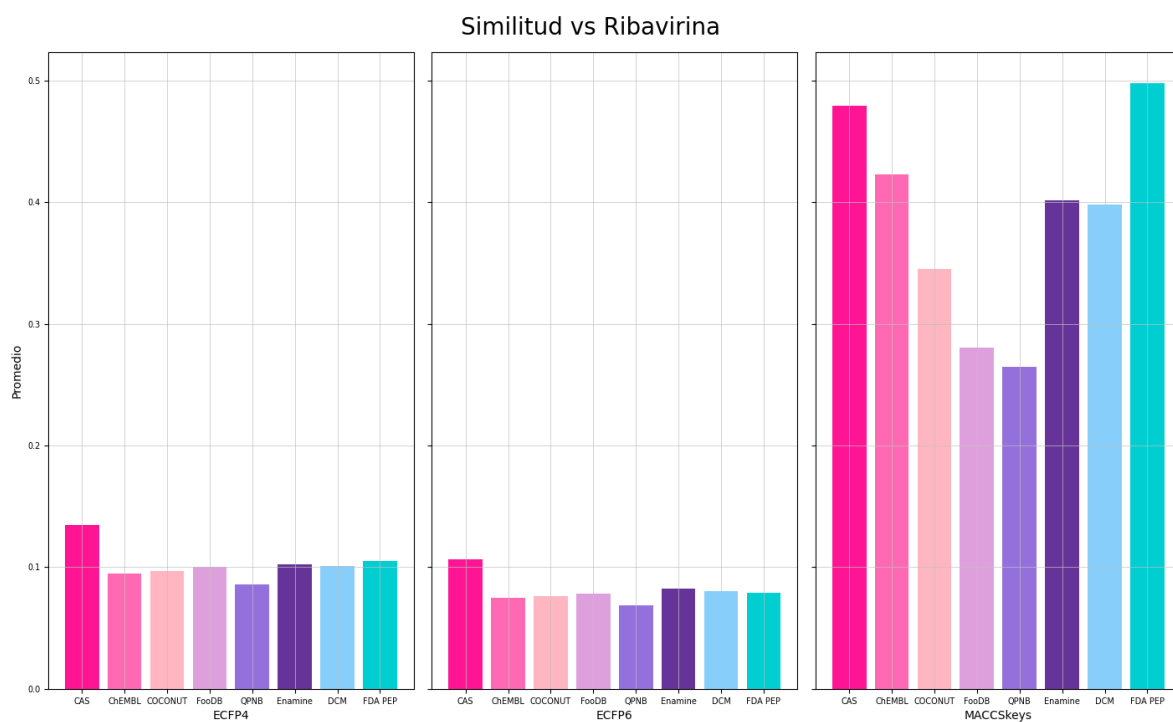


Figura S57. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs ribavirina.

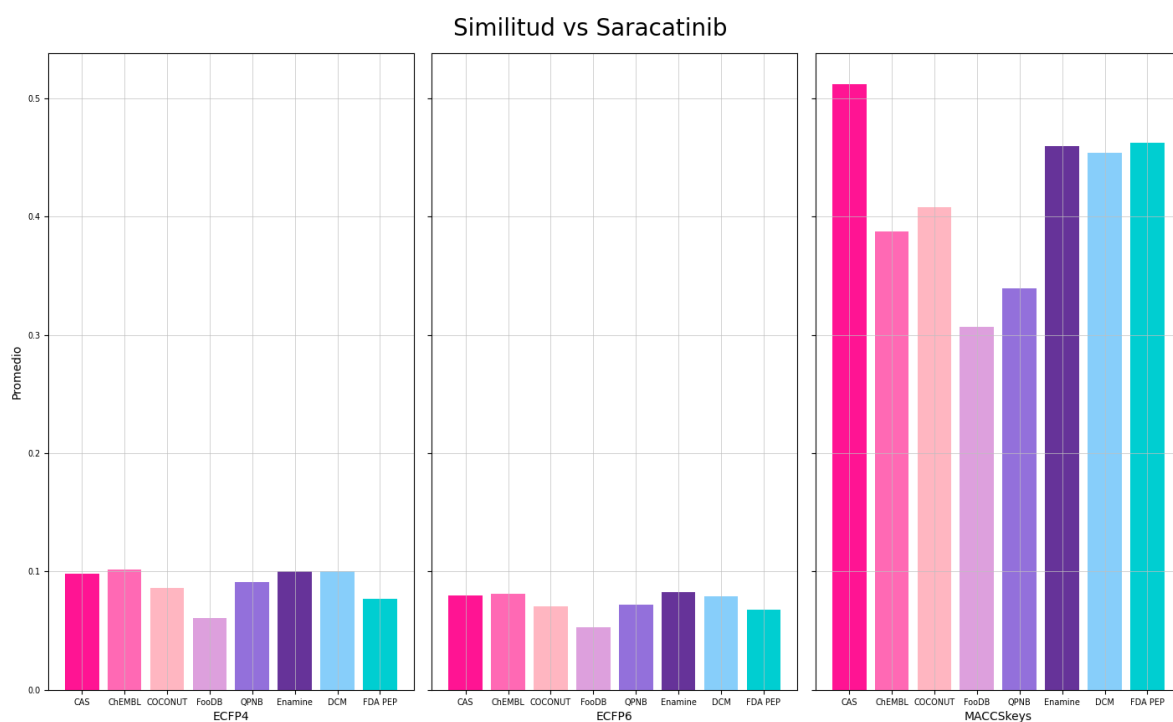


Figura S58. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs saracatinib.

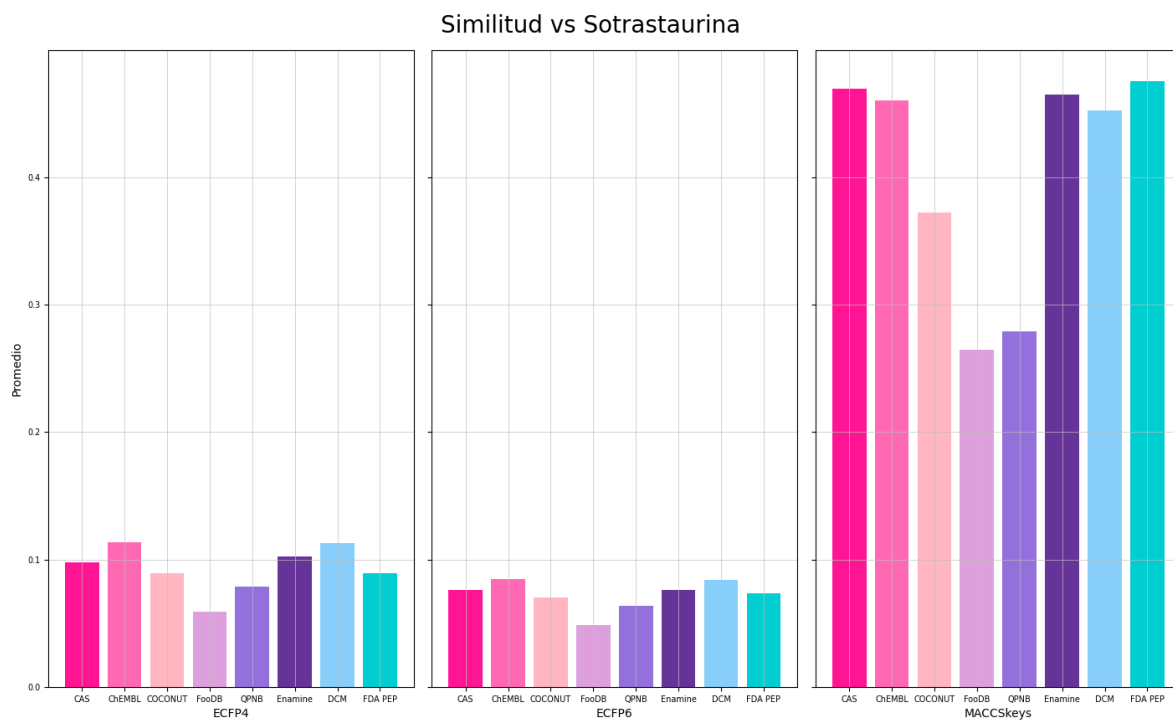


Figura S59. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs sotrastaurina.

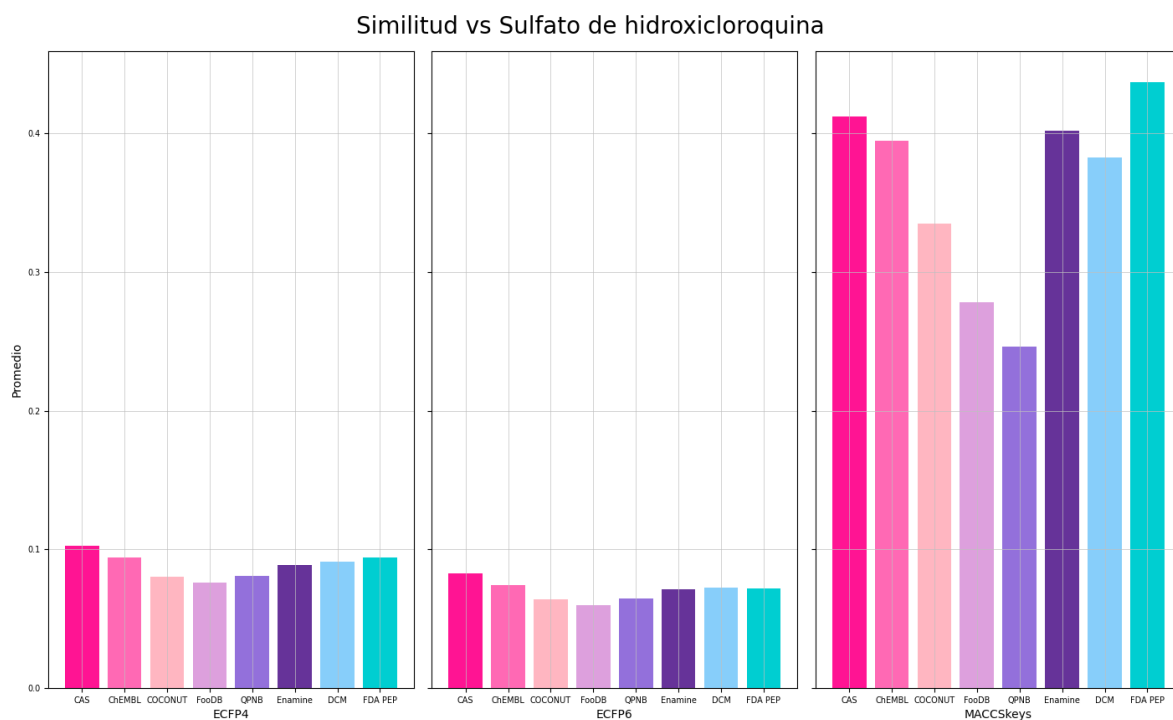


Figura S60. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs sulfato de hidroxicloroquina.

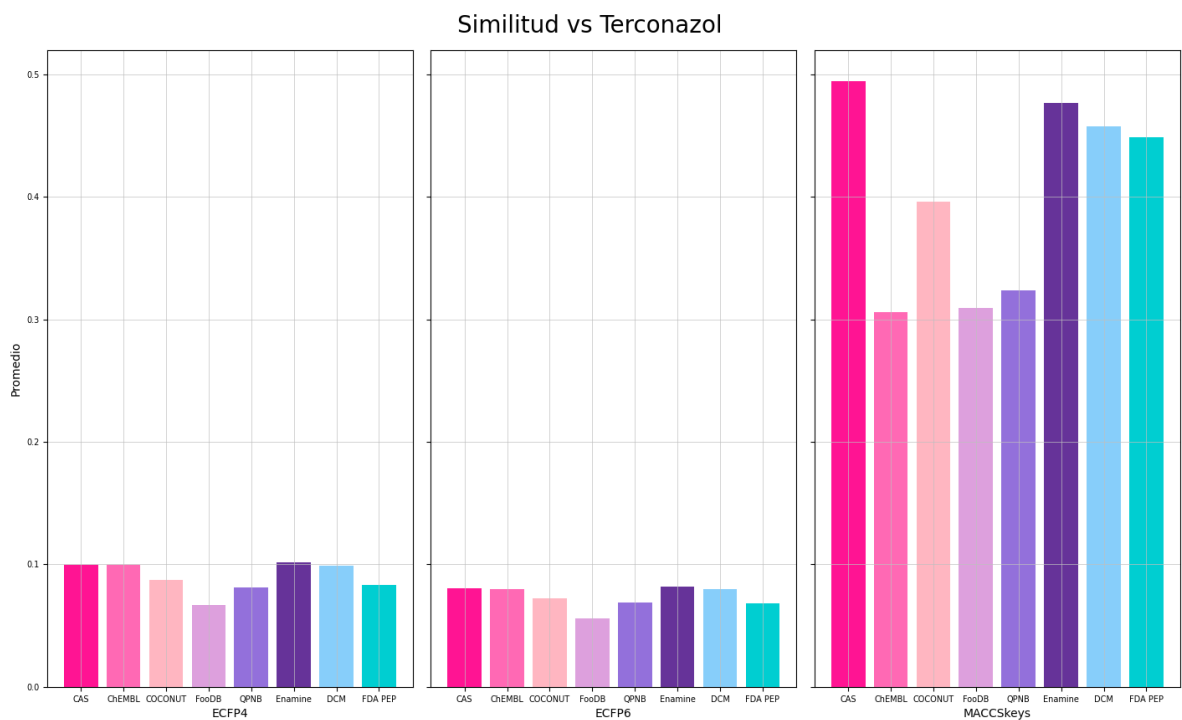


Figura S61. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs terconazol.

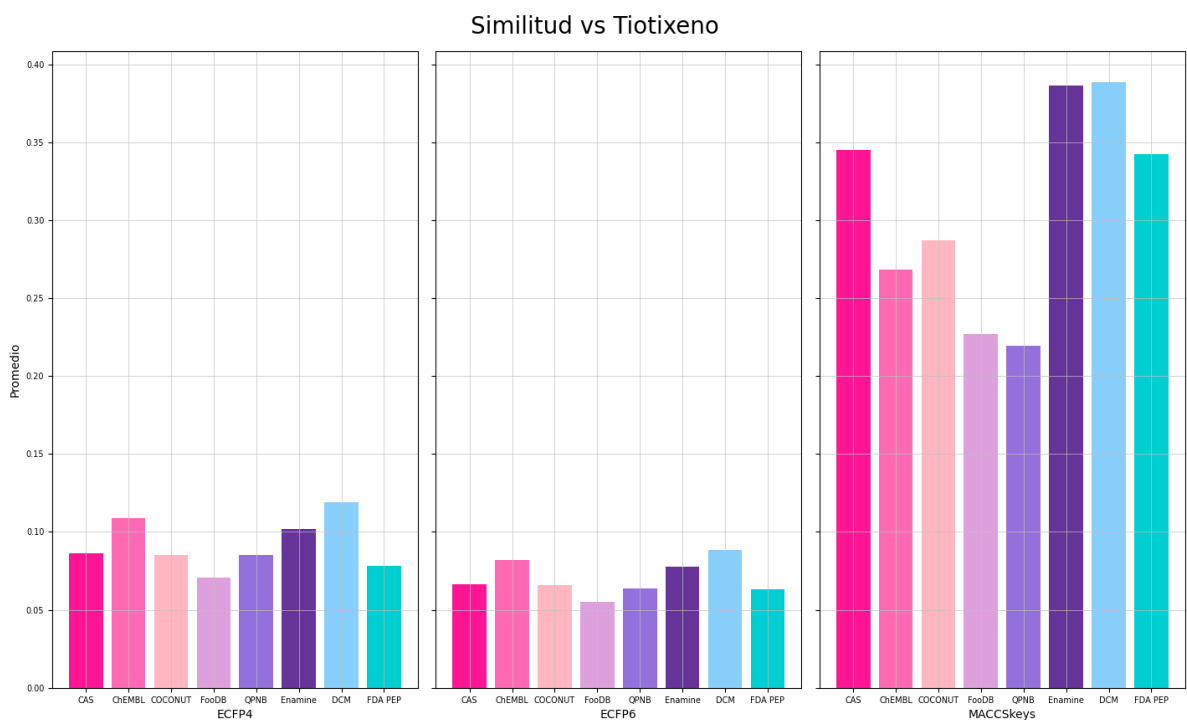


Figura S62. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs tiotixeno.

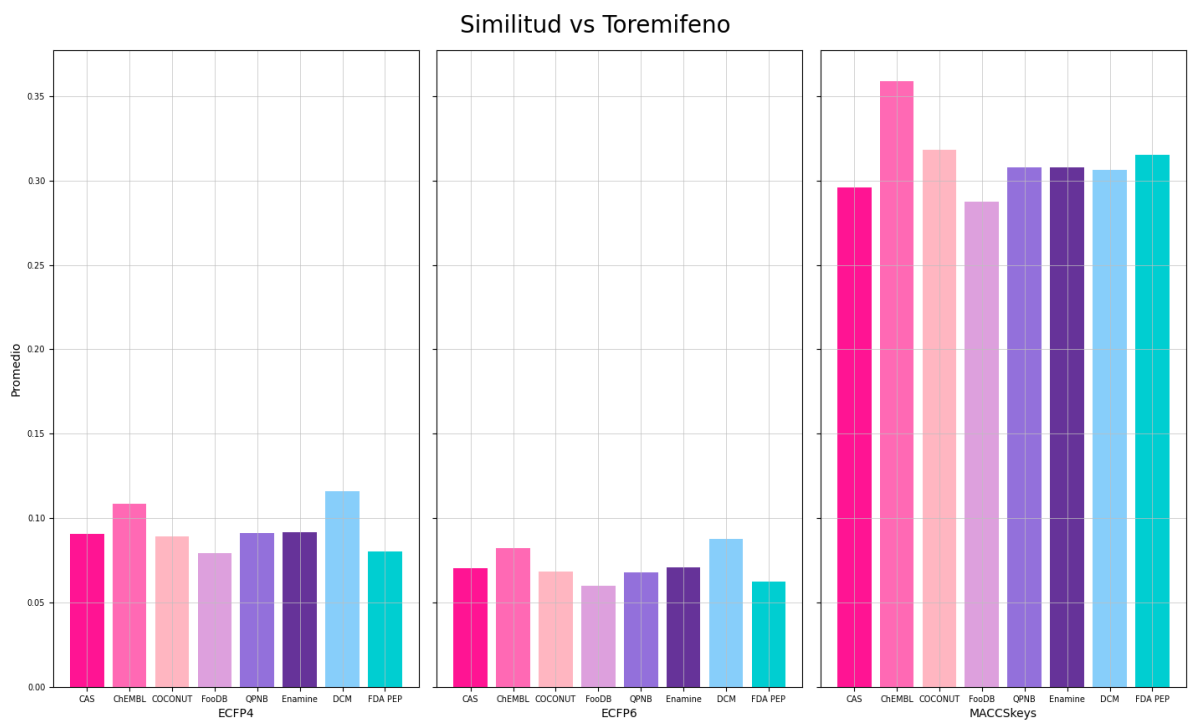


Figura S63. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs toremifeno.

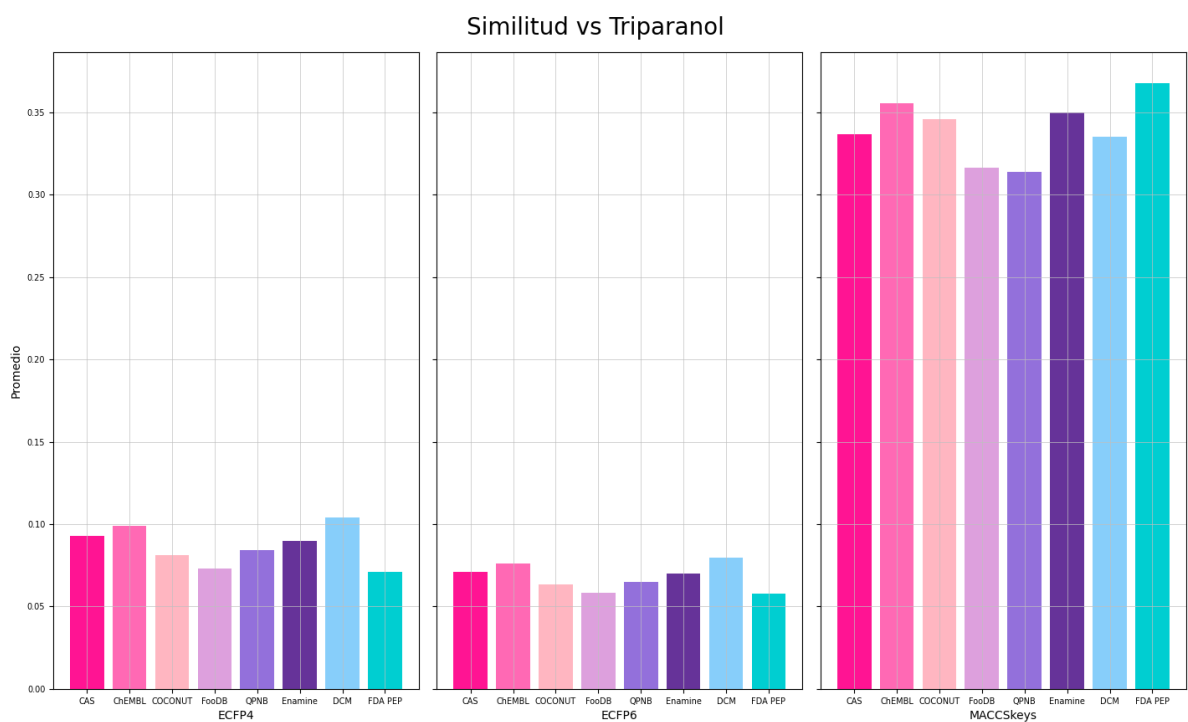


Figura S64. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs triparanol.

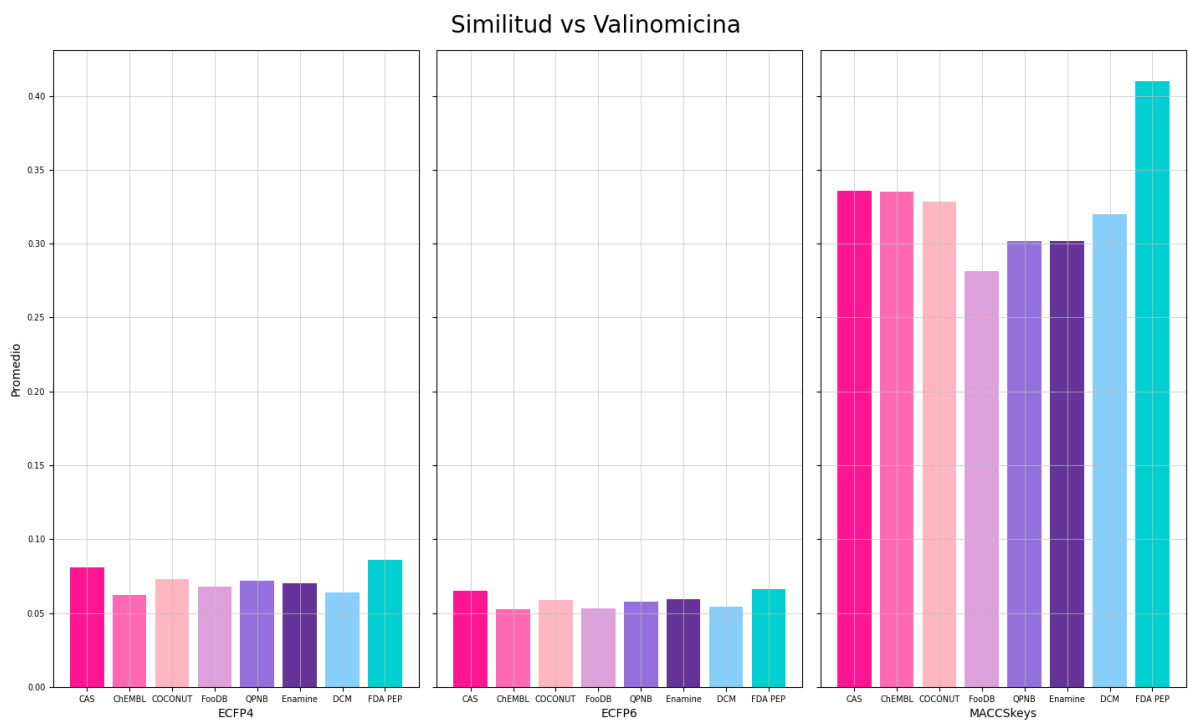


Figura S65. Promedio de los valores de similitud obtenidos con el coeficiente de Tanimoto para cada biblioteca vs valinomicina.

Tabla S1. Parámetros estadísticos del LogS (*Silicos IT LogSw*).

Parámetro estadístico	Biblioteca molecular				
	CAS	COCONUT	DCM	FooDB	QPNB
Promedio	-3.66	-4.65	-4.71	-3.82	-5.03
Desviación estándar	1.26	2.11	1.51	1.94	2.40
Valor mínimo	-5.15	-10.38	-9.47	-8.40	-6.47
1er cuartil	-4.44	-6.26	-5.77	-5.58	-6.42
2do cuartil	-4.20	-4.75	-4.71	-3.74	-6.36
3er cuartil	-2.64	-3.07	-3.75	-3.36	-4.31
Valor máximo	-1.50	0.78	1.03	0.10	-2.25

Tabla S2. Parámetros estadísticos de lipofilidad (*Consensus LogP*).

Parámetro estadístico	Biblioteca molecular				
	CAS	COCONUT	DCM	FooDB	QPNB
Promedio	1.95	2.34	2.40	1.84	2.29
Desviación estándar	1.24	1.25	0.94	1.28	1.28
Valor mínimo	0.02	-2.09	-1.40	-2.09	0.82
1er cuartil	1.18	1.50	1.76	1.04	1.89
2do cuartil	2.09	2.42	2.43	1.92	2.96
3er cuartil	2.82	3.30	3.10	2.73	3.03
Valor máximo	3.62	5.08	4.99	4.54	3.10

Tabla S3. Parámetros estadísticos de absorción intestinal humana (*HIA*).

Parámetro estadístico	Biblioteca molecular				
	CAS	COCONUT	DCM	FooDB	QPNB
Promedio	73.50	91.01	89.87	91.13	92.39
Desviación estándar	33.41	9.04	9.40	11.47	9.40
Valor mínimo	0.78	28.13	3.83	20.88	81.54
1er cuartil	71.40	89.92	89.47	91.67	89.59
2do cuartil	89.58	93.38	92.51	94.09	97.64
3er cuartil	93.44	95.85	94.69	97.45	97.81
Valor máximo	95.68	100.00	100.00	100.00	97.99

Tabla S4. Parámetros estadísticos de violaciones a las reglas de Lipinski.

Parámetro estadístico	Biblioteca molecular				
	CAS	COCONUT	DCM	FooDB	QPNB
Promedio	0.00	0.02	0.01	0.00	0.00
Desviación estándar	0.00	0.14	0.09	0.00	0.00
Valor mínimo	0.00	0.00	0.00	0.00	0.00
1er cuartil	0.00	0.00	0.00	0.00	0.00
2do cuartil	0.00	0.00	0.00	0.00	0.00
3er cuartil	0.00	0.00	0.00	0.00	0.00
Valor máximo	0.00	1.00	1.00	0.00	0.00

Tabla S5. Parámetros estadísticos de violaciones a las reglas de Veber.

Parámetro estadístico	Biblioteca molecular				
	CAS	COCONUT	DCM	FooDB	QPNB
Promedio	0.25	0.03	0.04	0.01	0.00
Desviación estándar	0.46	0.17	0.20	0.09	0.00
Valor mínimo	0.00	0.00	0.00	0.00	0.00
1er cuartil	0.00	0.00	0.00	0.00	0.00
2do cuartil	0.00	0.00	0.00	0.00	0.00
3er cuartil	0.25	0.00	0.00	0.00	0.00
Valor máximo	1.00	1.00	2.00	1.00	0.00

Tabla S6. Parámetros estadísticos de permeabilidad a la barrera hematoencefálica (*BBB permeability*).

Parámetro estadístico	Biblioteca molecular				
	CAS	COCONUT	DCM	FooDB	QPNB
Promedio	-0.60	-0.10	-0.08	-0.06	0.03
Desviación estándar	0.85	0.55	0.52	0.50	0.69
Valor mínimo	-1.68	-1.88	-2.09	-1.83	-0.75
1er cuartil	-1.31	-0.45	-0.44	-0.26	-0.21
2do cuartil	-0.59	-0.04	-0.06	0.05	0.34
3er cuartil	-0.16	0.30	0.30	0.29	0.43
Valor máximo	0.83	1.50	1.65	0.69	0.52

Tabla S7. Parámetros estadísticos de la eliminación total.

Parámetro estadístico	Biblioteca molecular				
	CAS	COCONUT	DCM	FooDB	QPNB
Promedio	0.60	0.56	0.53	0.56	0.57
Desviación estándar	0.35	0.36	0.37	0.31	0.22
Valor mínimo	0.08	-1.05	-0.52	-0.31	0.32
1er cuartil	0.40	0.30	0.24	0.35	0.51
2do cuartil	0.55	0.59	0.54	0.51	0.70
3er cuartil	0.84	0.82	0.83	0.76	0.70
Valor máximo	1.11	1.67	1.96	1.27	0.71

Tabla S8. Nomenclatura SMILES de los 18 *hits* computacionales.

Compuesto	Nomenclatura SMILES
1	<chem>CN(C)c1ccc(C2Nc3cccc4cccc(c34)N2)cc1</chem>
2	<chem>O=C(Oc1cccc1C(=O)Oc1cccc1)c1cco1</chem>
3	<chem>COc1ccc2c(c1)CN(C(=O)CCN1CCC(Cc3cccc3)CC1)CCS2</chem>
4	<chem>CN1CCCN(S(=O))(=O)c2ccc3cccc3c2)CC1</chem>
5	<chem>CN(C)c1ccc(C(c2ccc(N(C)C)cc2)c2ccc(N(C)C)cc2)cc1</chem>
6	<chem>CN(C)c1ccc(C(c2cccc2)c2ccc(N(C)C)cc2)cc1</chem>
7	<chem>CN(C)c1ccc(C(=O)N(Cc2cccc2)Cc2cccc2)cc1</chem>
8	<chem>C[N+](C)(C)CCOC(=O)Oc1ccc2cccc2c1</chem>
9	<chem>COC1CCC2CC3c4cc5c(cc4CCN3CC2C1=O)OCO5</chem>
10	<chem>CCN(Cc1cc2cccc2[nH]c1=O)C(=O)c1cccnc1</chem>
11	<chem>Cc1cc(OCC(=O)N2CCC(O)(c3ccc(Cl)cc3)CC2)c2c(C)cc(=O)oc2c1</chem>
12	<chem>COc1cccc(C(=O)N(C)CC2CCN(CCc3cccc(C(F)(F)F)c3)CC2)c1</chem>
13	<chem>Cc1cc(OC(=O)c2ccc3cccc3c2)c([N+](=O)[O-])c(=O)[nH]1</chem>
14	<chem>COc1cccc1CN1CCN(Cc2ccc3cccc3c2)CC1</chem>
15	<chem>c1ccc(CN2CCN(Cc3ccc4cccc4c3)CC2)cc1</chem>
16	<chem>CN1CCN(Cc2ccc3cccc3c2)CC1</chem>
17	<chem>OCCN1CCN(Cc2ccc3cccc3c2)CC1</chem>
18	<chem>CC(C)CC(=O)N1CCN(Cc2ccc3cccc3c2)CC1</chem>

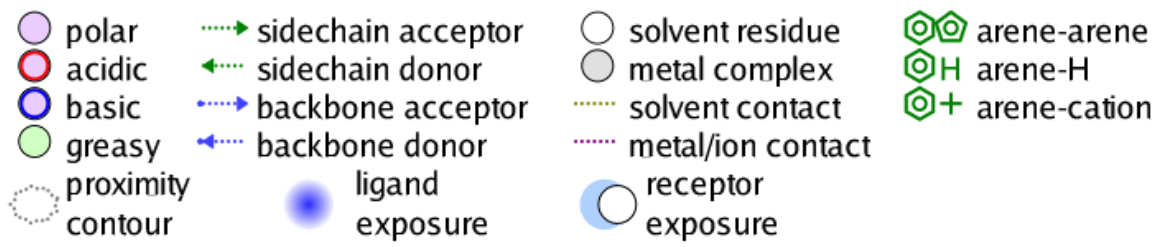


Figura S66. Nomenclatura para los diagramas de interacciones en 2D.

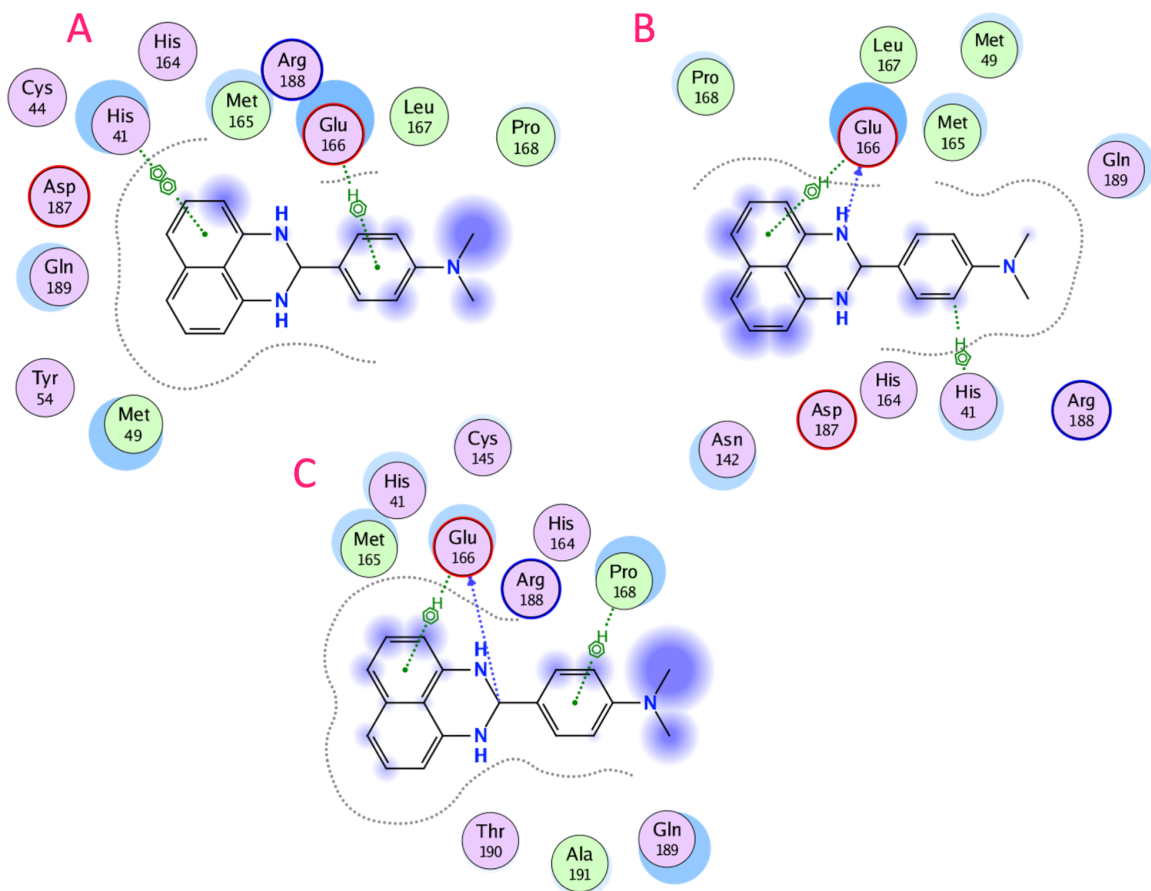


Figura S67. Interacciones predichas entre el *hit* computacional 1 y la Mpro del SARS-CoV-2 en A) MOE, B) LeDock y C) Autodock Vina.

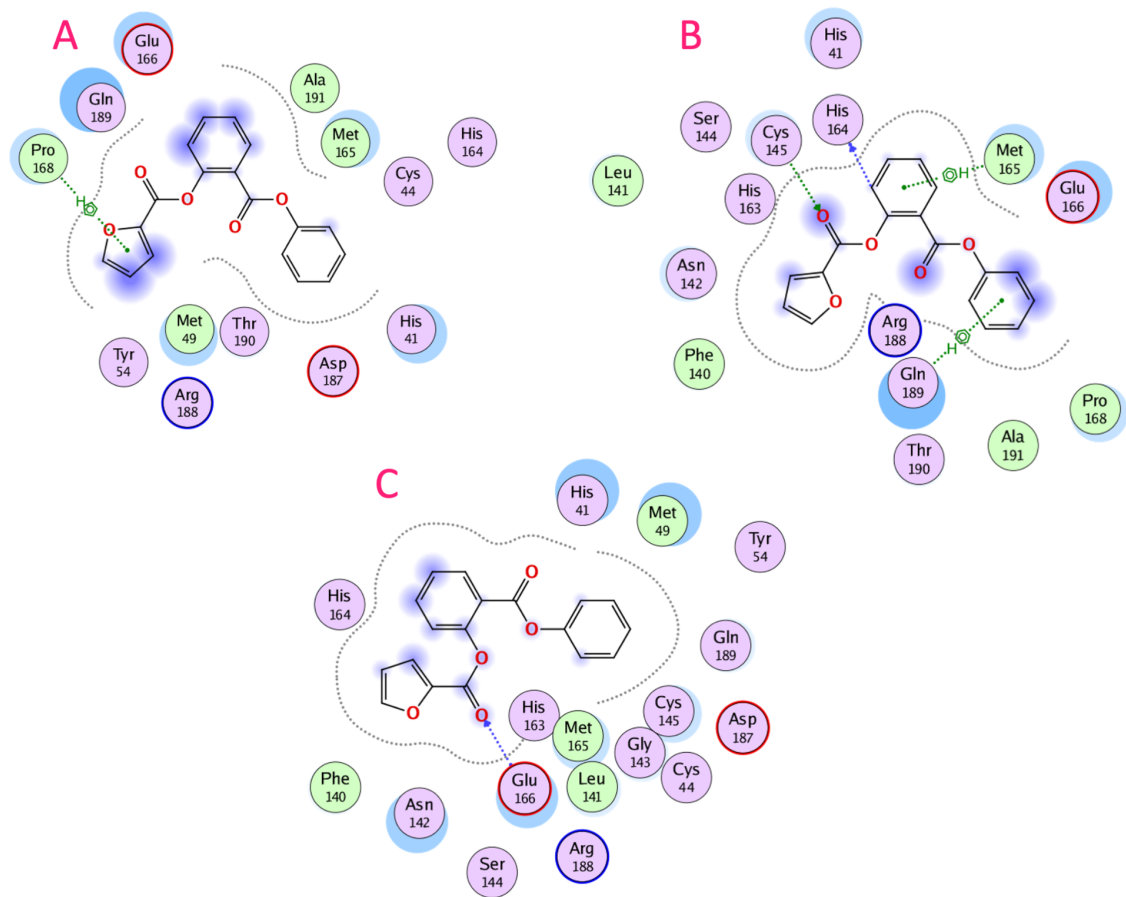


Figura S68. Interacciones predichas entre el *hit* computacional **2** y la Mpro del SARS-CoV-2 en A) MOE, B) LeDock y C) Autodock Vina.

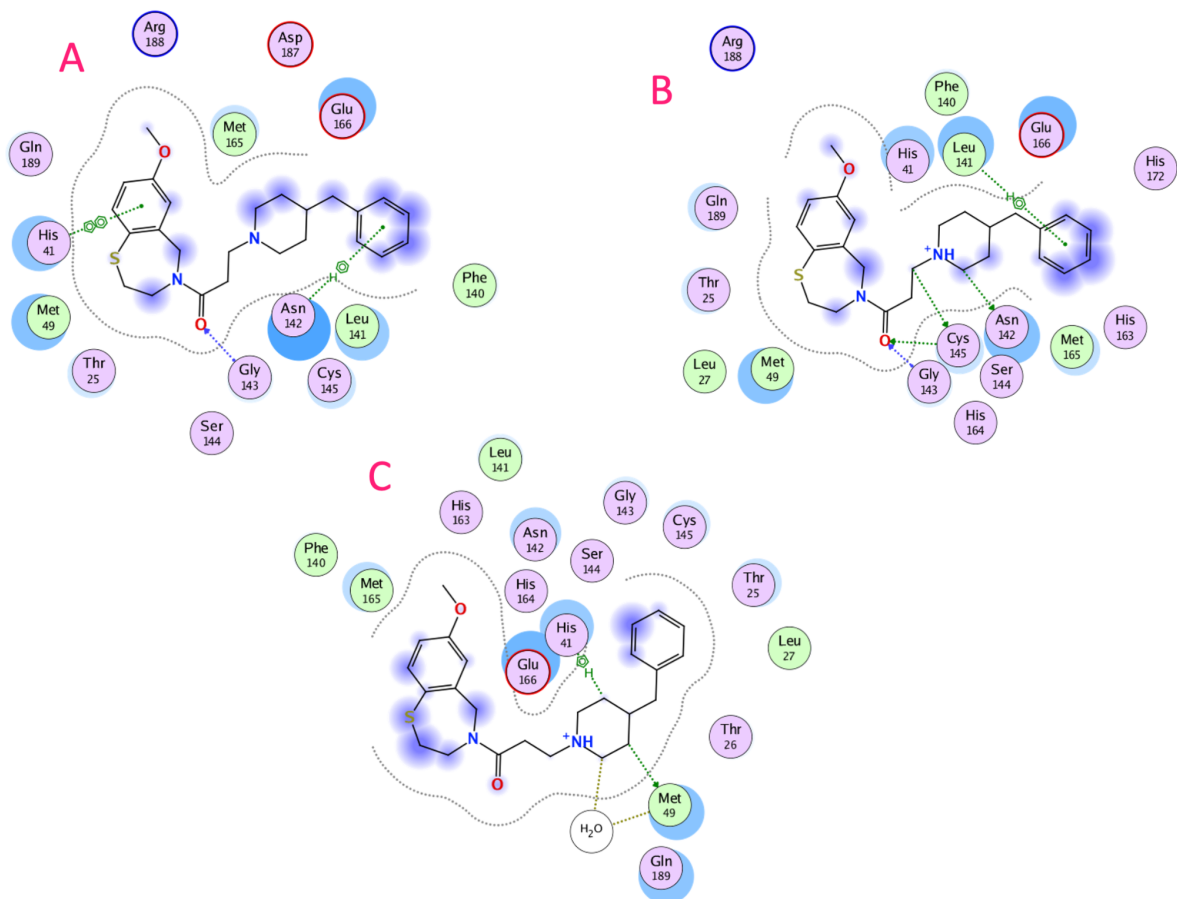


Figura S69. Interacciones predichas entre el *hit* computacional 3 y la Mpro del SARS-CoV-2 en A) MOE, B) LeDock y C) Autodock Vina.

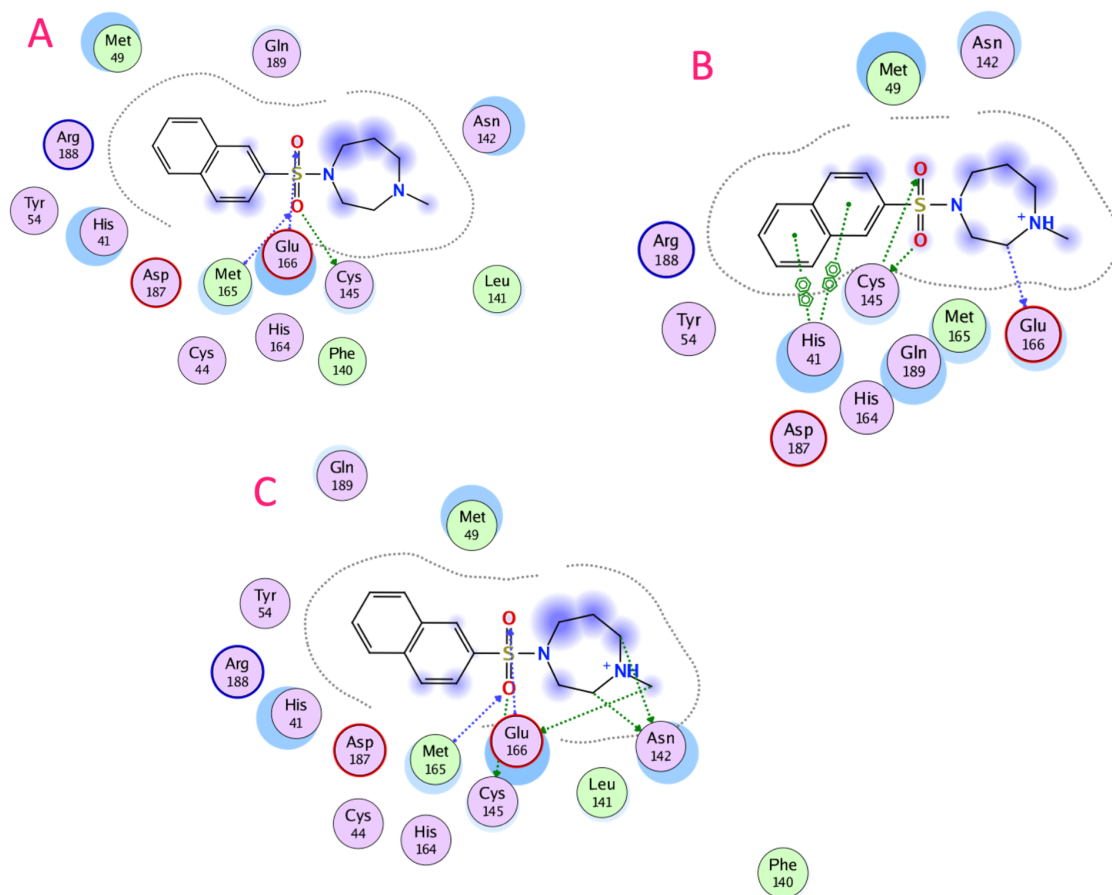


Figura S70. Interacciones predichas entre el *hit* computacional **4** y la Mpro del SARS-CoV-2 en A) MOE, B) LeDock y C) Autodock Vina.

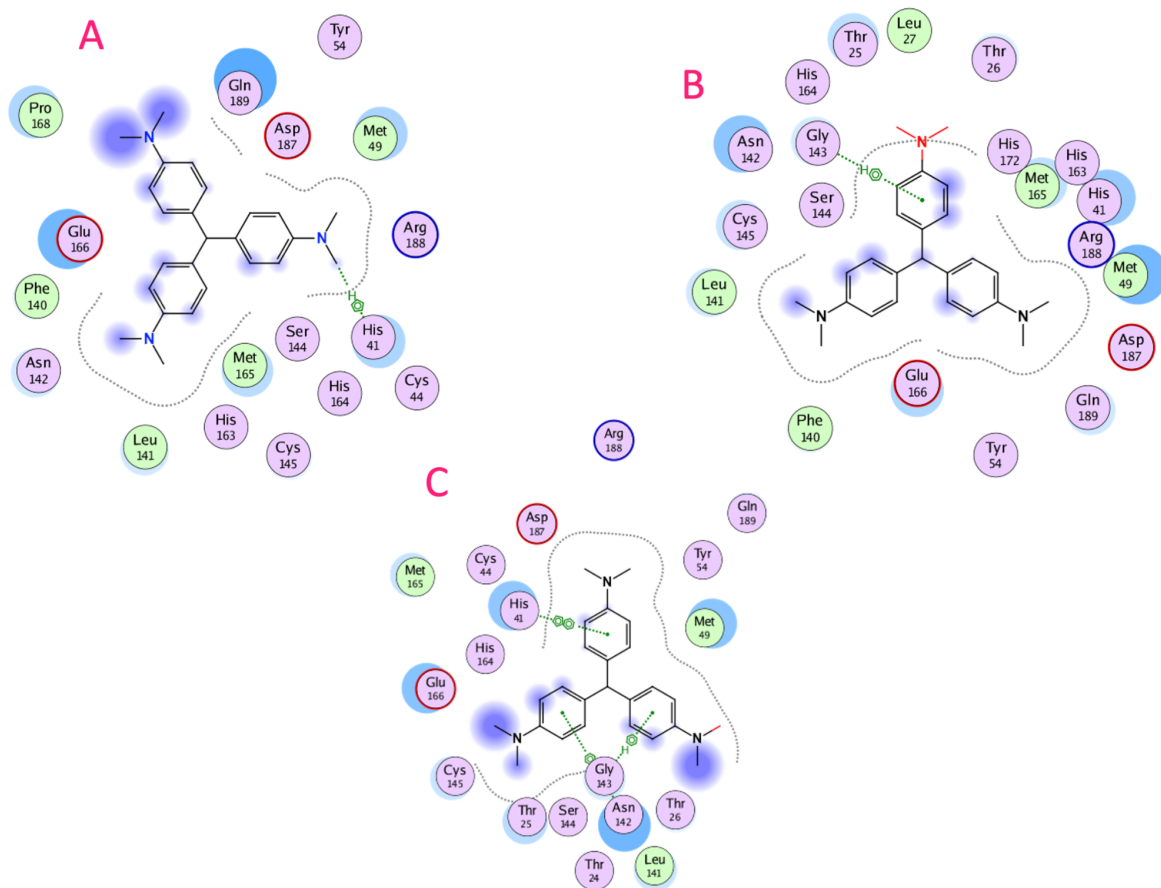


Figura S71. Interacciones predichas entre el *hit* computacional 5 y la Mpro del SARS-CoV-2 en A) MOE, B) LeDock y C) Autodock Vina.

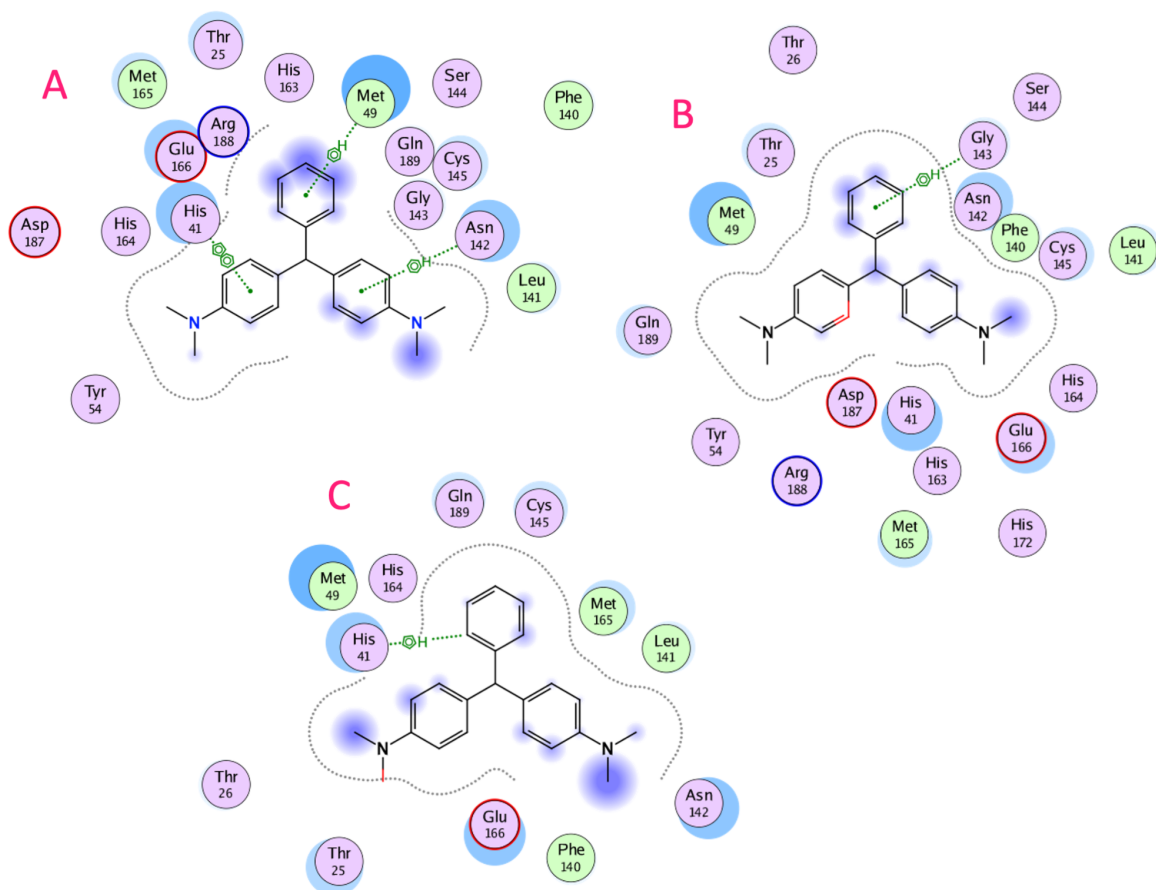


Figura S72. Interacciones predichas entre el *hit* computacional 6 y la Mpro del SARS-CoV-2 en A) MOE, B) LeDock y C) Autodock Vina.

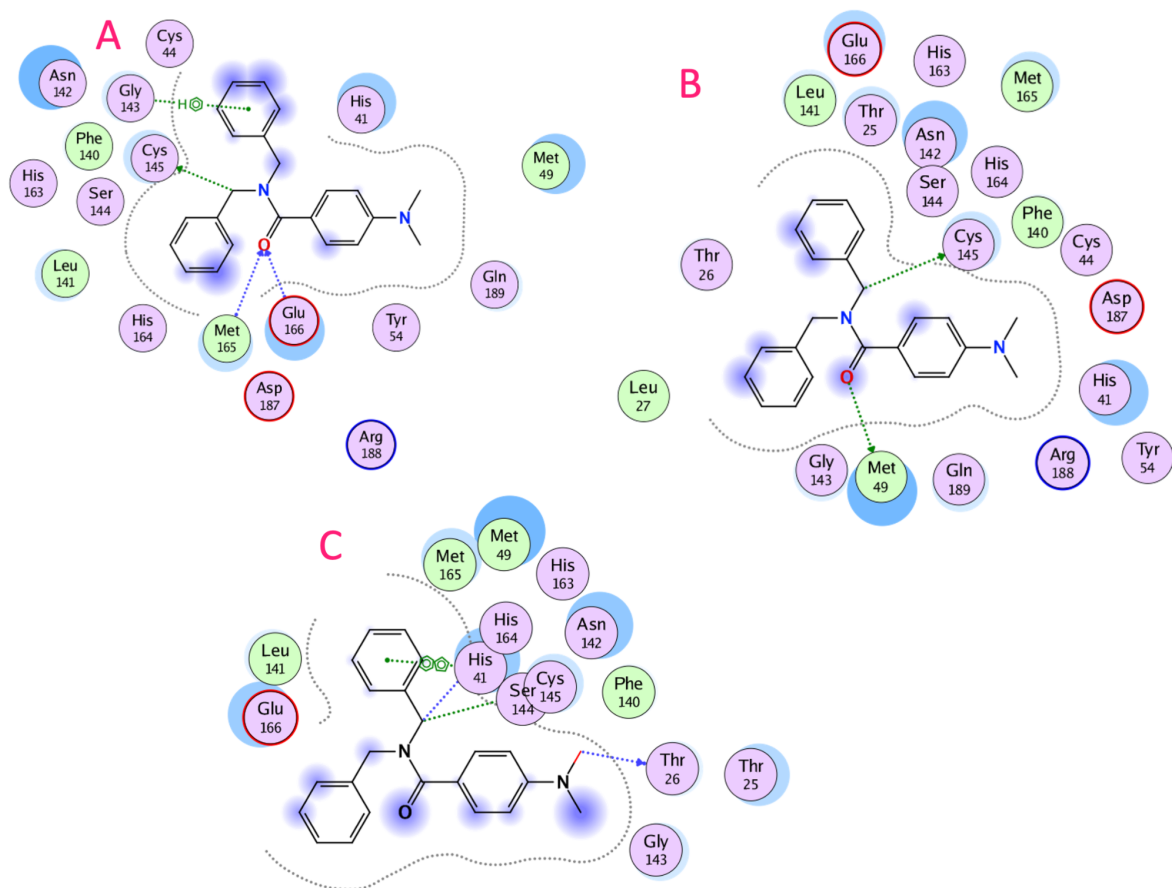


Figura S73. Interacciones predichas entre el *hit* computacional 7 y la Mpro del SARS-CoV-2 en A) MOE, B) LeDock y C) Autodock Vina.

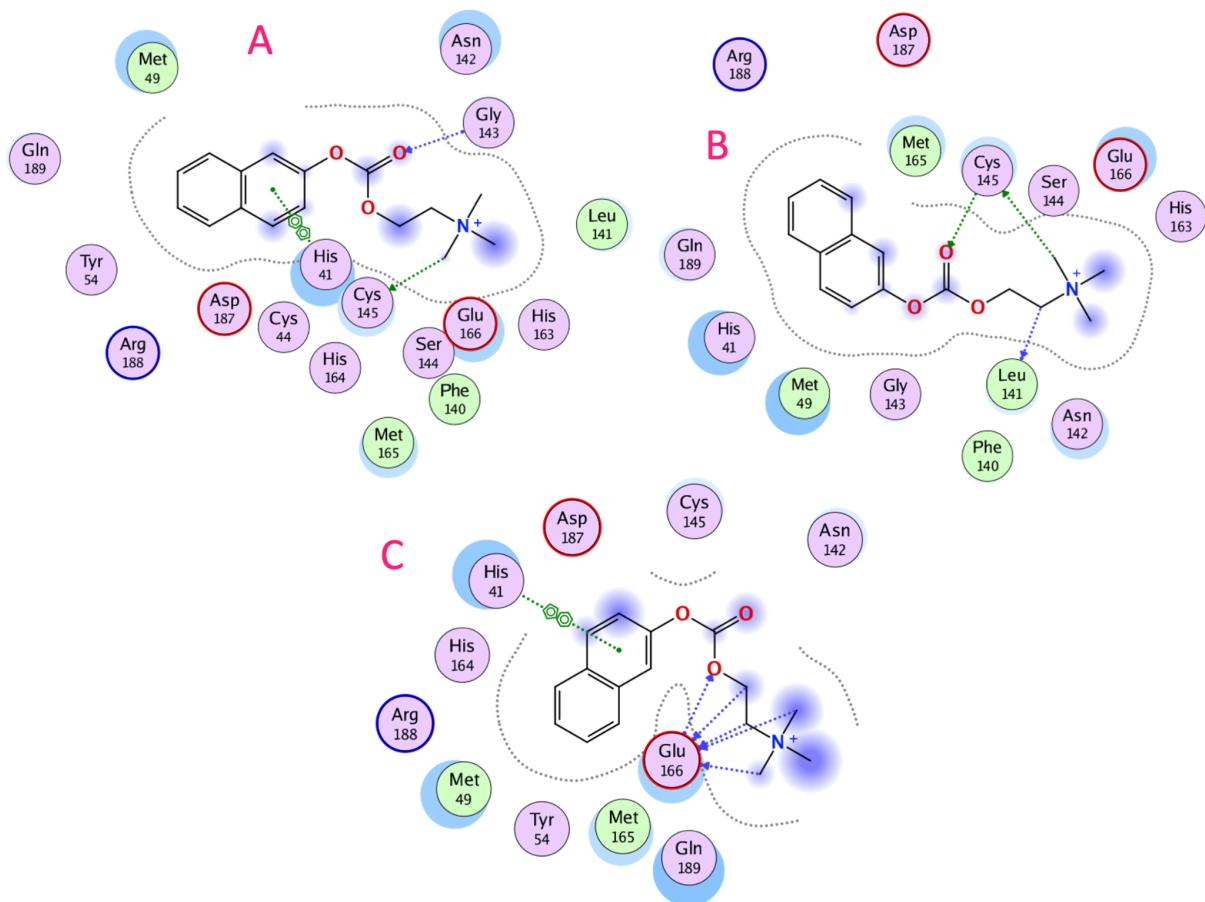


Figura S74. Interacciones predichas entre el *hit* computacional 8 y la Mpro del SARS-CoV-2 en A) MOE, B) LeDock y C) Autodock Vina.

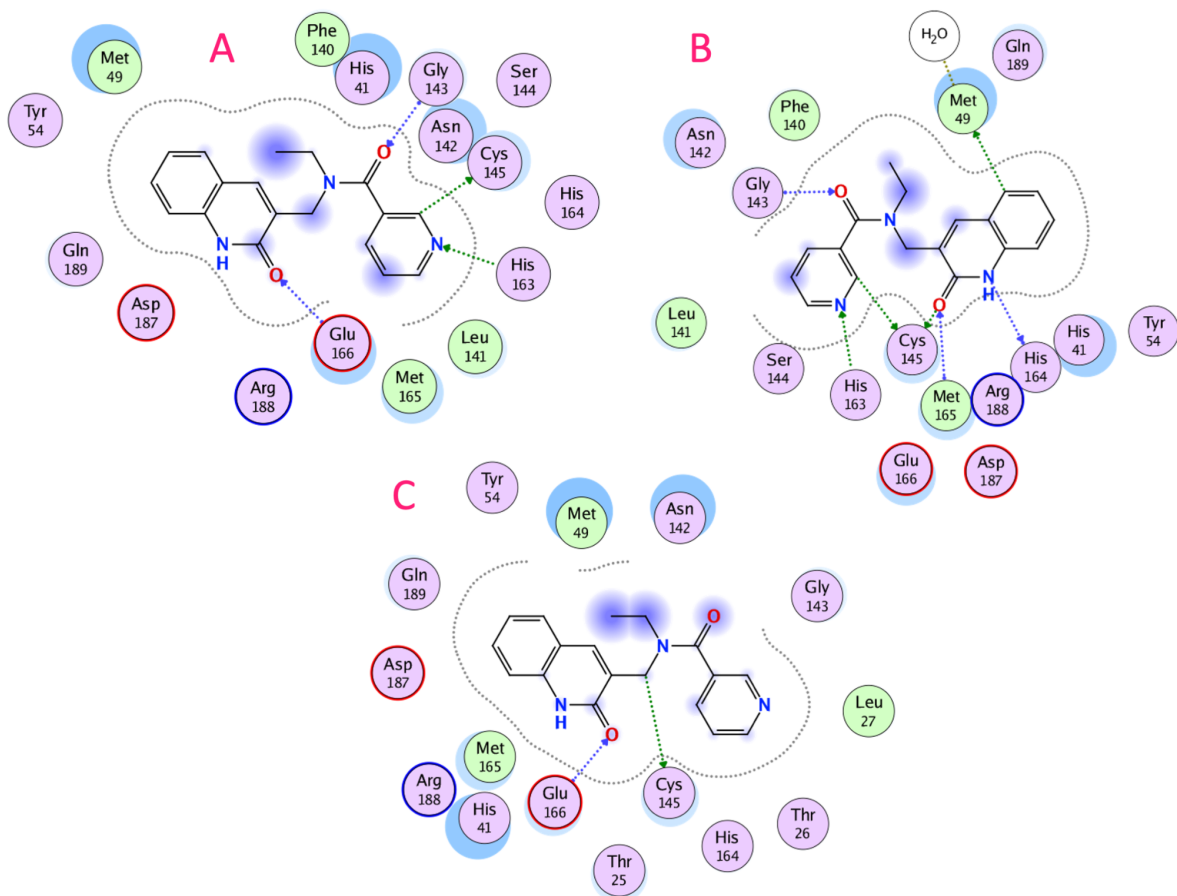


Figura S76. Interacciones predichas entre el *hit* computacional 10 y la Mpro del SARS-CoV-2 en A) MOE, B) LeDock y C) Autodock Vina.

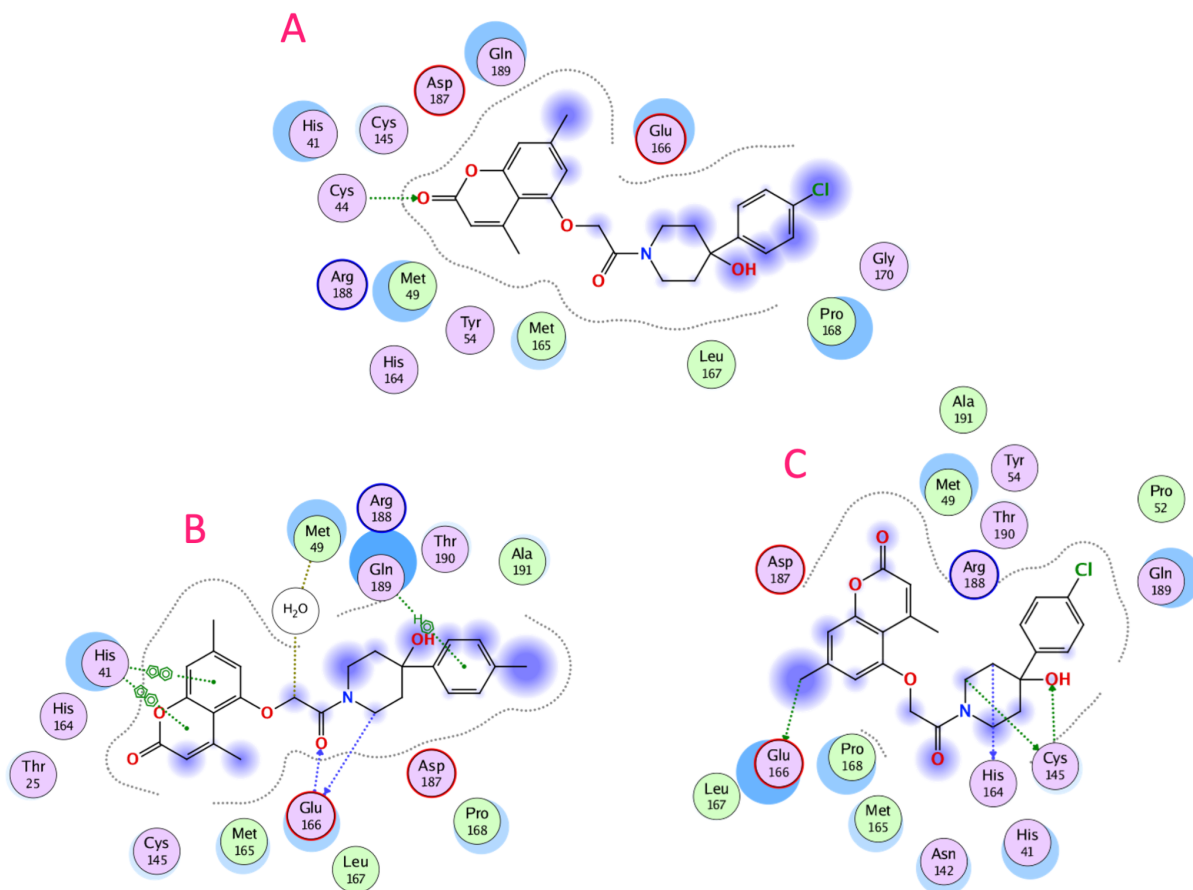


Figura S77. Interacciones predichas entre el *hit* computacional **11** y la Mpro del SARS-CoV-2 en A) MOE, B) LeDock y C) Autodock Vina.

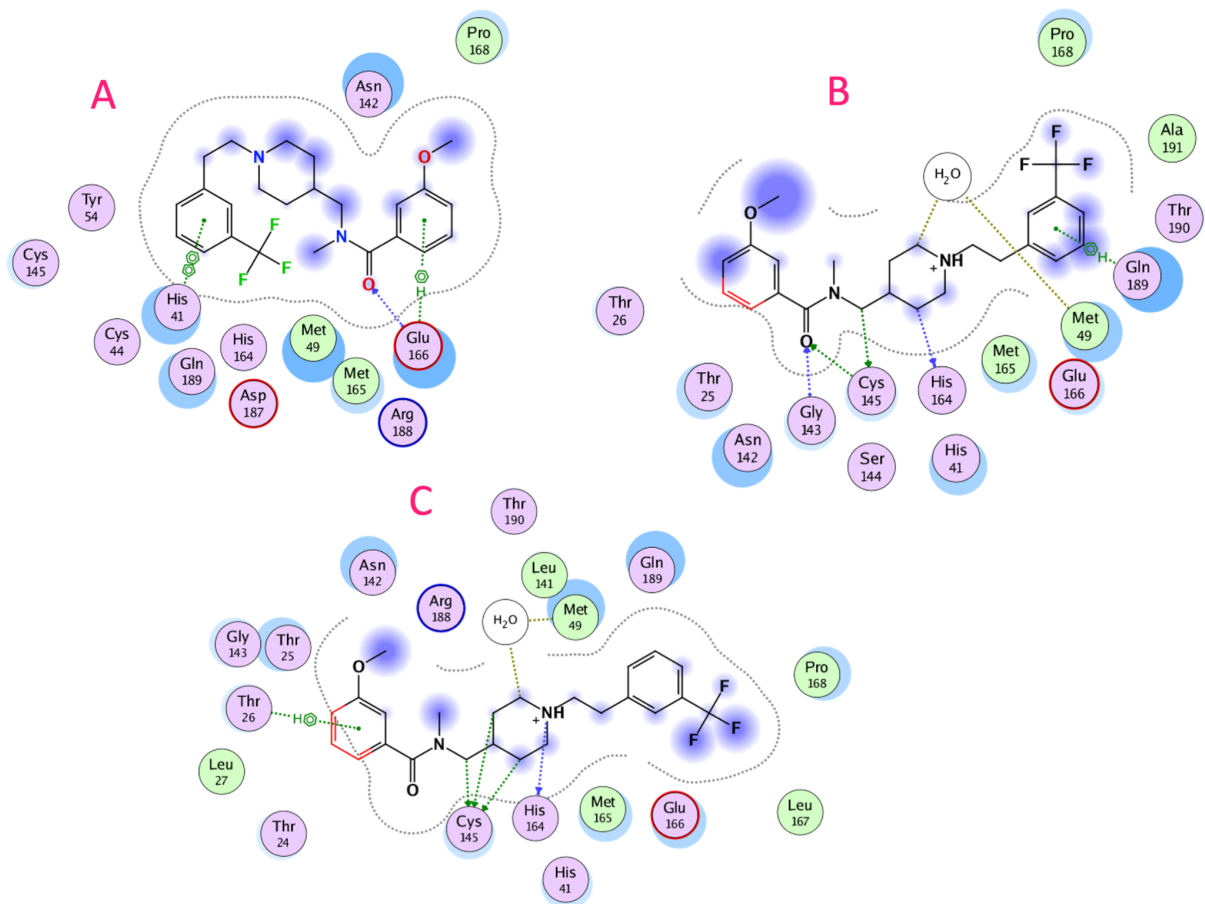


Figura S78. Interacciones predichas entre el *hit* computacional 12 y la Mpro del SARS-CoV-2 en A) MOE, B) LeDock y C) Autodock Vina.

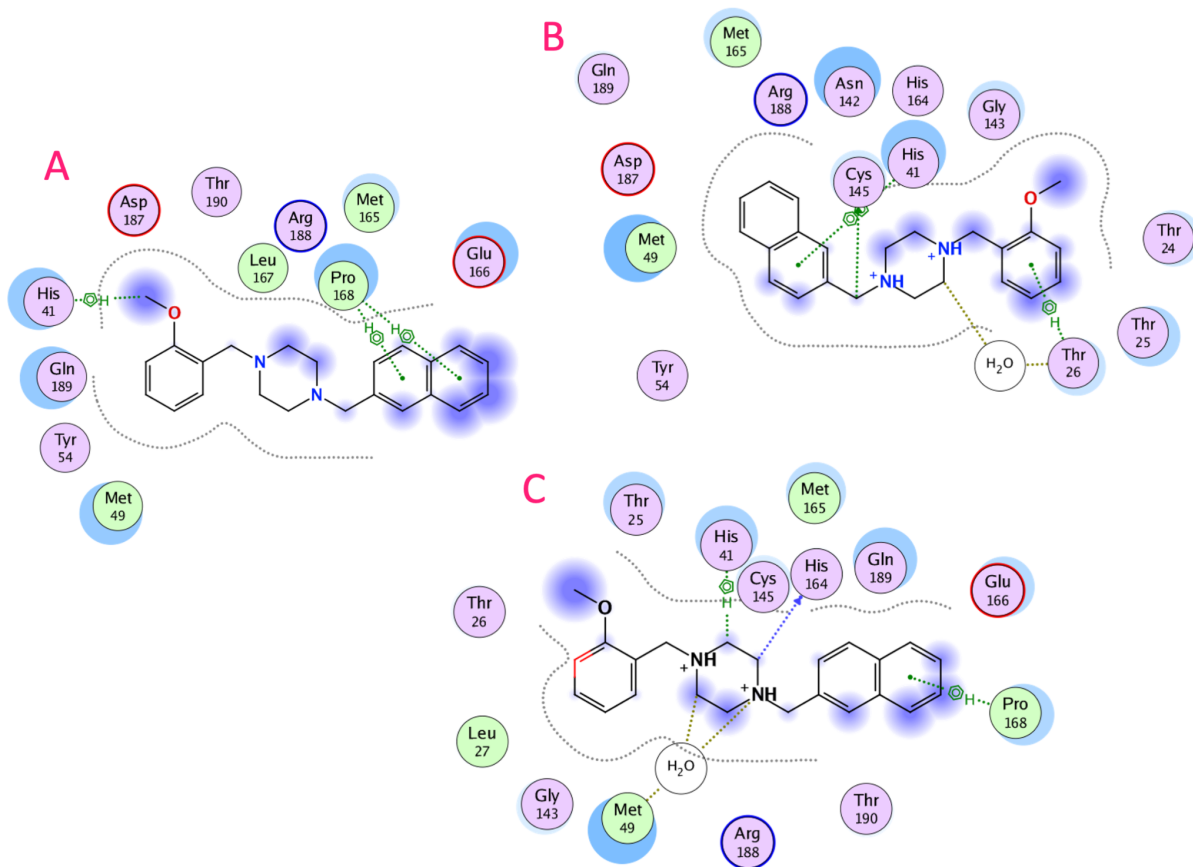


Figura S79. Interacciones predichas entre el *hit* computacional **14** y la Mpro del SARS-CoV-2 en A) MOE, B) LeDock y C) Autodock Vina.

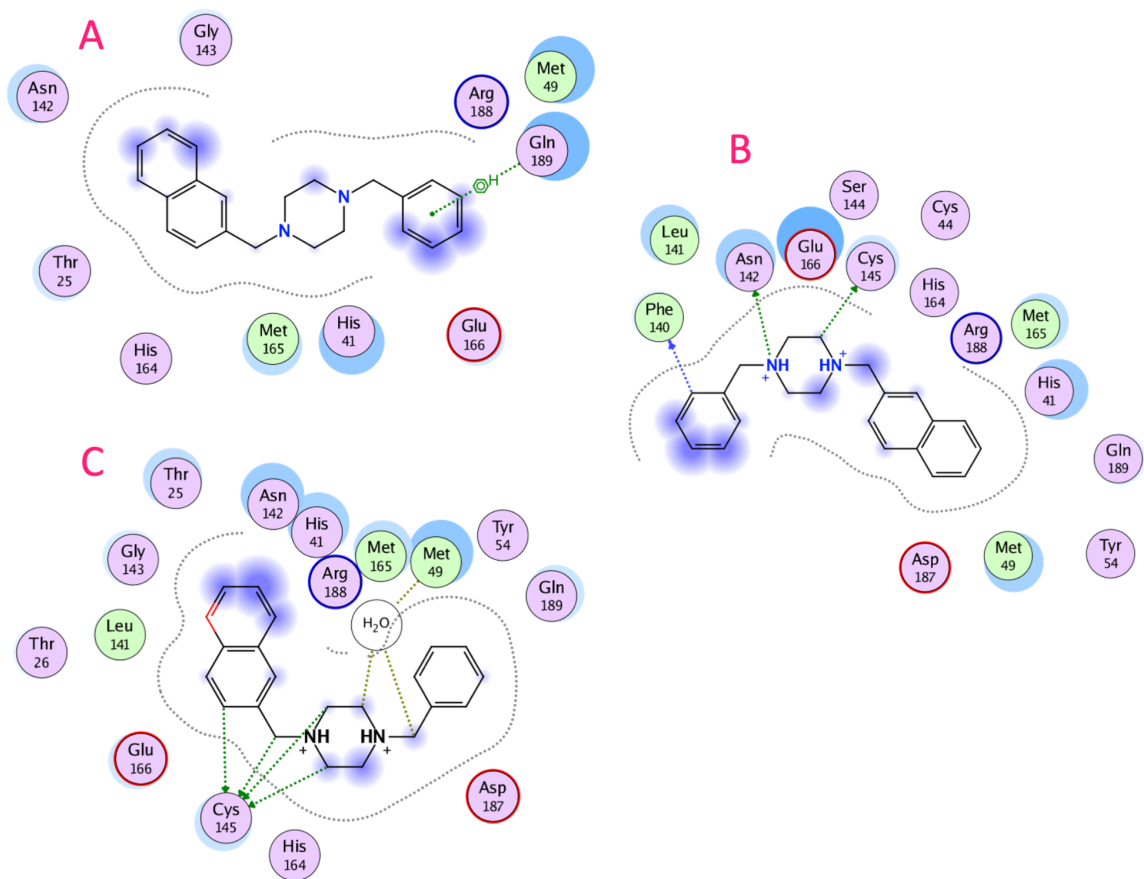


Figura S80. Interacciones predichas entre el *hit* computacional **15** y la Mpro del SARS-CoV-2 en A) MOE, B) LeDock y C) Autodock Vina.

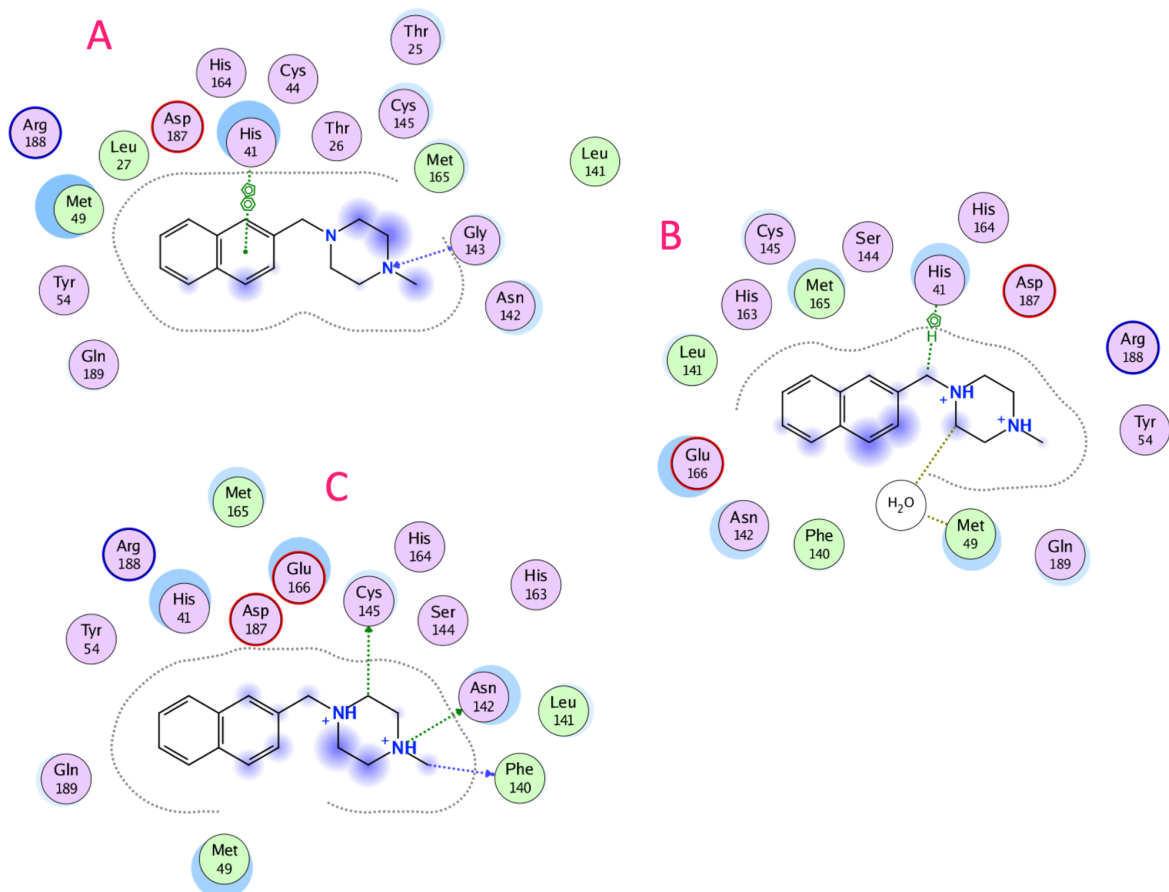


Figura S81. Interacciones predichas entre el *hit* computacional 16 y la Mpro del SARS-CoV-2 en A) MOE, B) LeDock y C) Autodock Vina.

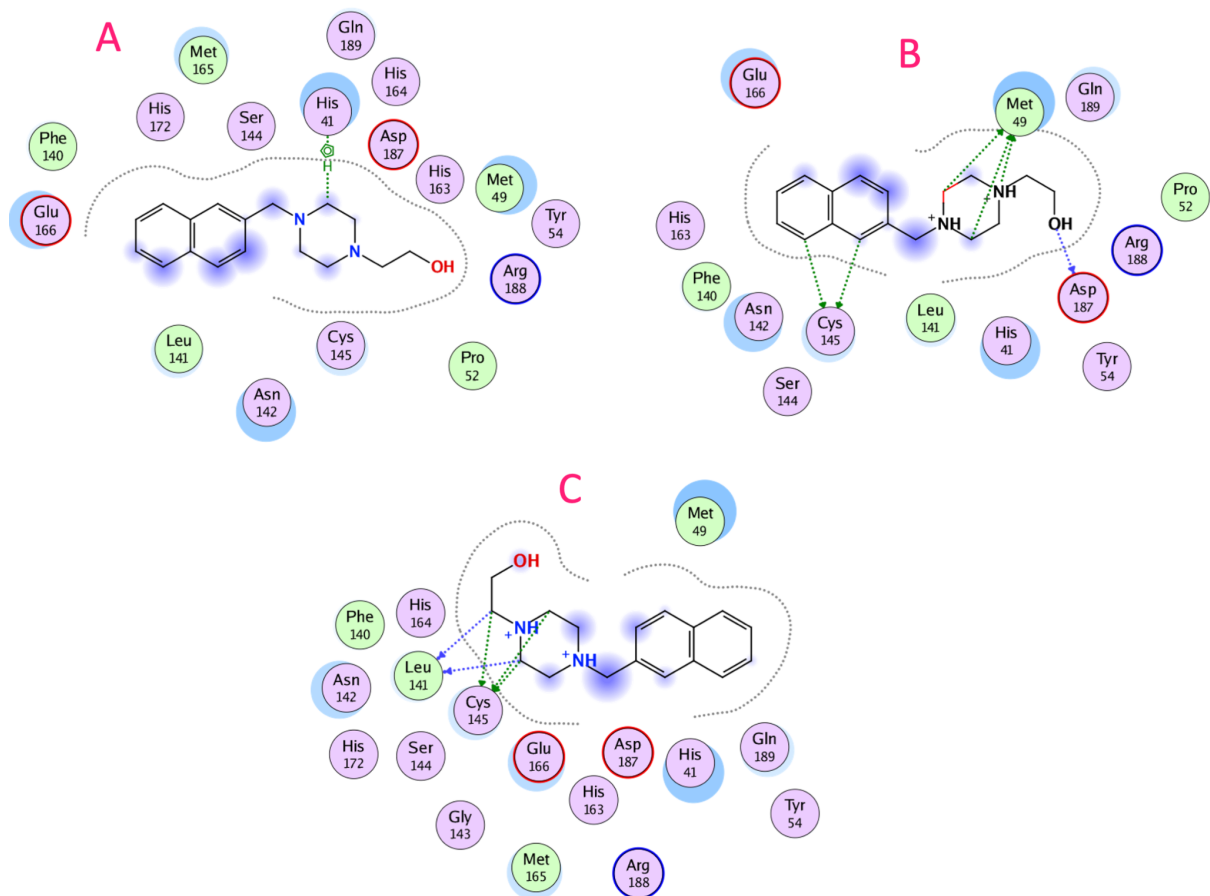


Figura S82. Interacciones predichas entre el *hit* computacional 17 y la Mpro del SARS-CoV-2 en A) MOE, B) LeDock y C) Autodock Vina.

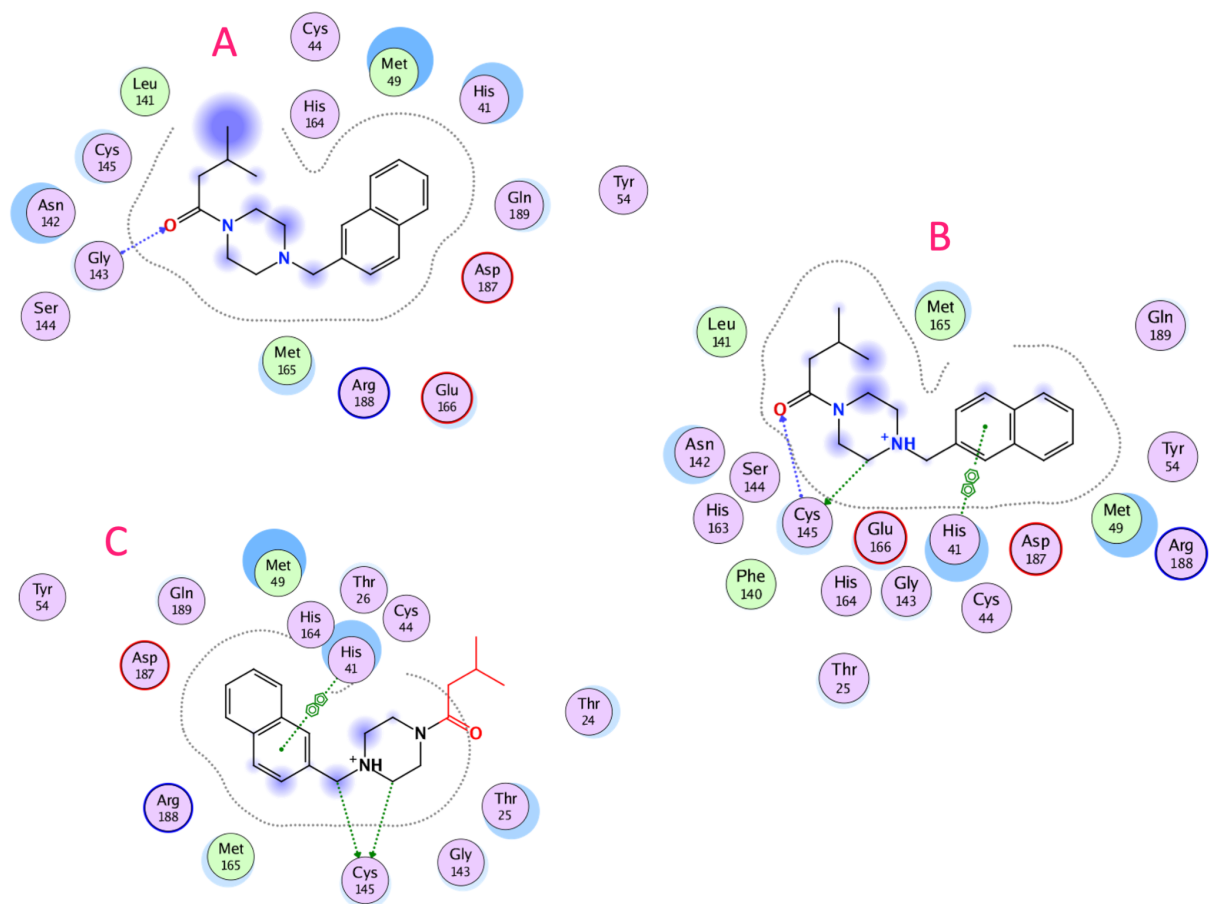


Figura S83. Interacciones predichas entre el *hit* computacional 18 y la Mpro del SARS-CoV-2 en A) MOE, B) LeDock y C) Autodock Vina.



COMPUTATIONAL FILTERING OF COMPOUND DATABASES TO IDENTIFY POTENTIAL DRUG CANDIDATES FOR THE TREATMENT OF COVID-19



K. Eurídice Juárez-Mercado, Bárbara I. Díaz-Eufracio and José L. Medina-Franco
DIFACQUIM research group, Department of Pharmacy, School of Chemistry, Universidad Nacional Autónoma de México, Avenida Universidad 3000, Mexico City 04510, Mexico

Abstract

The goal of this study was to screen systematically a large and broad area of the chemical space using chemoinformatic principles. As a general strategy first, we compiled a database with nearly 18 million molecules from eight different sources including a focused library, bioactive compounds, natural products, peptides, molecules classified as dark chemical matter, and a large on-demand collection with unique chemical structures. Then we performed fast 2D similarity searching with active compounds as queries (a list of 65 compounds were retrieved from the literature with reported activity against different coronavirus vs different targets). After of this, we did a search by commercial availability employ the "For-Sale", "In-Stock," and "Now" subsets for check; and we did the ADME-tox profile at the result compounds. Also Machine Learning and QSAR predictions was employed at the results of commercial availability compounds; obtaining 19 compounds with positive results in both analysis. Finally, these final compounds were submitted at a Molecular Docking analysis in three different softwares (MOE, LeDock and AutoDock Vina) to evaluate their possible interactions with the crystal structure of COVID-19 main protease (PDB: ID: 6LU7).

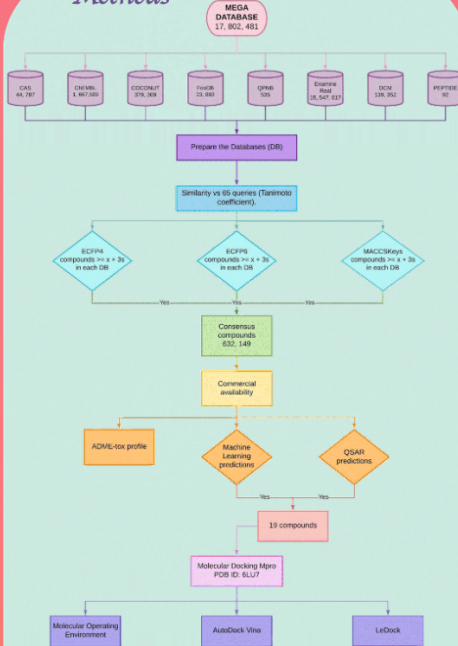
Introduction

A novel Coronavirus named Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) by the International Committee on Taxonomy of Viruses (ICTV) [1], was associated with a cluster of respiratory tract infections named COVID-19 and has rapidly spread across the world, causing that the World Health Organization (WHO) recognized it as a pandemic on March 2020 [2-3].



Because of the emergency of the pandemic, one of the most attractive strategies to identify potential treatment of COVID-19 is to screen compound data sets to identify potential drug candidates hitting one or more of the main molecular targets or directly inhibiting the SARS-CoV-2. Chemoinformatic analysis and computational filtering of compound databases or virtual screening (VS) is a logical step to shorten times and cost of massive experimental screening.

Methods



Results

Total compounds of Megadatabase

17, 802, 481

Similarity Searching

632,149 compounds

The similarity values were computed with the Tanimoto coefficient. Fingerprint and Tanimoto coefficient were computed with the Python API of RDKit. The selection of hit compounds was based on distribution of similarity values for each fingerprint and each query: similarity hits were chosen as those with a similarity value greater than or equal the median plus three standard deviations of the corresponding distribution for each query compound, the final hit list was assembled from a consensus approach: identifying compounds with high similarity values by all three fingerprints.

Commercial availability

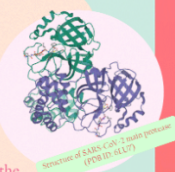


The commercial availability of the hits selected from similarity searching from CAS, COCONUT, FooDB, QPDB, DCM and FDA PEP Databases was assessed by a quick search of the structures in the most recent release of the ZINC database (ZINC 15) [4] and we obtained a total number of **4,423** unique molecules.

The ADME-tox profile evaluated the absorption, distribution, metabolism, elimination and toxicity of the compounds after commercial availability searching and was performed on the SwissADME server [5]. QSAR prediction of hit compounds were done with QSAR models recently published using different molecular descriptors (the models were developed to predict activity as inhibitors of Mpro of SARS-CoV-2) [6].

Molecular Docking

The compounds with positive results in the Machine Learning Predictions and QSAR predictions were selected for a Molecular Docking study, which was performed in three different softwares: Molecular Operating Environment (MOE) version 2018, LeDock and AutoDock Vina to evaluate the possible binding poses between the compounds and the crystallographic structure of SARS-CoV-2 main protease.



Conclusions

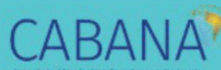
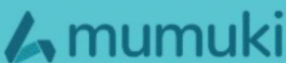
We obtained 19 potential computational candidates as inhibitors of the main protease of SARS-CoV-2. This work suggests evaluating the inhibitory activity of these compounds with in vitro and in vivo studies.

References

- 1) Gorbalenya, A.E., Baker, S.C., Baric, R.S. et al. The species *Severe acute respiratory syndrome-related coronavirus* classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol* 5, 536–544 (2020). <https://doi.org/10.1038/s41564-020-0695-z>.
- 2) Cucinotta D, Vanelli M. WHO Declares COVID-19 a Pandemic. *Acta Bio Med [Internet]*. 2020;Mar;19. Epub 2020Sep;9:91(1):157-60. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7281907/>
- 3) Cheng Ng Kee Kwong, K., Mehta, P. R., Shukla, G., & Mehta, A. R. (2020). COVID-19, SARS and MERS: A neurological perspective. *Journal of Clinical Neuroscience*. Available from: <https://doi.org/10.1016/j.jocn.2020.04.124>
- 4) Sterling T., & Irwin, J. J. (2015). ZINC 15 – Ligand Discovery for Everyone. *Journal of Chemical Information and Modeling*, 55(11), 2324–2337. Available from: <https://doi.org/10.1021/acs.jcim.5b00559>
- 5) Daina, A., Michielin, O., & Zoete, V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci Rep* 7, 42717 (2017). Available from: <https://doi.org/10.1038/srep42717>
- 6) Bekrowski, Tomasz; Alves, Vinícius; Melo-Filhos, Cleber C.; Korn, Daniel; Auerbach, Scott S.; Schmitt, Charles et al. (2020); Computational Models Identify Several FDA Approved or Experimental Drugs as Putative Agents Against SARS-CoV-2. *ChemRxiv*. Preprint available from: <https://doi.org/10.26434/chemrxiv-121535944>

Acknowledgments

This research was supported by the Consejo Nacional de Ciencia y Tecnología (CONACyT), Mexico (support No. 16231) and the Programa de Apoyo a la Investigación y el Posgrado (PAIP) grant 5000-9163, Facultad de Química, UNAM. Authors are grateful to Dirección General de Cómputo y de Tecnologías de Información y Comunicación (DGTIC), UNAM, for the computational resources to use Miztel supercomputer and UNAM, project LANCAD-UNAM-DGTIC-335 and also thank the financial support of the NUATEI (Nuevas Alternativas para el Tratamiento de Enfermedades Infecciosas) program IBT-UNAM to purchase MOE license.



www <https://womenbioinfodatascla.github.io>

WomenBioinfoDataScLA

Womenbioinfodatascl

Bioinfola



Computational filtering of compound databases to identify potential drug candidates for the treatment of COVID-19

Eurídice Juárez-Mercado¹, Bárbara I. Díaz-Eufracio¹, José L. Medina-Franco¹

¹. DIFACQUIM research group, Department of Pharmacy, School of Chemistry, Universidad Nacional Autónoma de México, Avenida Universidad 3000, Mexico City 04510, Mexico. *kaeuridice@gmail.com

Because of the emergency of the pandemic, one of the most attractive strategies to identify potential treatment of COVID-19 is to screen compound data sets to identify potential drug candidates hitting one or more of the main molecular targets or directly inhibiting the SARS-CoV-2. Chemoinformatic analysis and computational filtering of compound databases or virtual screening (VS) is a logical step to shorten times and cost of massive experimental screening. The goal of this study was to screen systematically a large and broad area of the chemical space using chemoinformatic principles. As a general strategy first, we compiled a database with nearly 18 million molecules from eight different sources including a focused library, bioactive compounds, natural products, peptides, molecules classified as *dark chemical matter*, and a large on-demand collection with unique chemical structures. Then we performed fast 2D similarity searching with active compounds as queries (a list of 65 compounds were retrieved from the literature with reported activity against different coronavirus vs different targets). All compounds were curated using an in-house developed protocol using public sources and were represented using SMILES strings. All the preparation was done with RDKit, version 2019.09.1. Similarity searching was performed using three 2D fingerprints, namely: Extended Connectivity Fingerprints ECFP4, ECFP6 and Molecular ACCes System (MACCS) keys (166-bits), and their values were computed with the Tanimoto coefficient. The selection of hit compounds was based on distribution of similarity values for each fingerprint and each query: similarity hits were chosen as those with a similarity value greater than the median plus three standard deviations of the corresponding distribution. For each query compound, the final hit list was assembled from a consensus approach: identifying compounds with high similarity values by all three fingerprints. This work suggests evaluating the activity of these compounds with aid of ADMETox profiling and QSAR-predictions.

Keywords: COVID-19, SARS-CoV-2, chemoinformatics, similarity searching, virtual screening.

Computational Drug Design Methods—Current and Future Perspectives

*Fernando D. Prieto-Martínez**, *Edgar López-López†*,
*K. Eurídice Juárez-Mercado**, *José L. Medina-Franco**

*Department of Pharmacy, School of Chemistry, National Autonomous University of Mexico, Mexico City, Mexico †Medicinal Chemistry Laboratory, University of Veracruz, Veracruz, Mexico

1 INTRODUCTION

The generation of, search for, and experimental evaluation of new molecules with increased potency and selectivity is booming. This is possible with the aid of techniques such as combinatorial chemistry and high-throughput screening (HTS). However, these techniques also generate a large number of false positives, so even with this evaluation capacity it is necessary to further filter screening data sets. Computer-aided drug design (CADD) is a discipline that collects multiple chemical-molecular and quantum strategies with the aim of discovering, designing, and developing therapeutic chemical agents. Many CADD approaches are based on structure-activity relationships (SAR). The main objectives of CADD are part of a multidisciplinary work for the improvement of bioactive molecules, the development of therapeutic alternatives, and the understanding of biological events at the molecular level.

In general, the drug-discovery process includes three key stages: (1) the discovery phase, in which the goal is the identification of relevant molecular targets and active molecules or hits; (2) the development phase, where the compounds are evaluated using *in vitro* and *in vivo* models (this phase includes various stages: preclinical, clinical I, II, and III); and (3) the registry phase that will enable distribution on the market and the clinical use of drugs. Recent estimates indicate that the average cost of the preclinical phase is 3.4 million dollars,

increasing to 8.6 and 21.4 million dollars respectively in clinical phases II and III (Martin, Hutchens, Hawkins, & Radnov, 2017).

The role of CADD in the drug-discovery process lies mainly in the discovery phase, where a primary goal is to filter and select compounds for experimental synthesis or testing. It is expected that this filtering will reduce the time and costs involved in drug development. In addition, CADD enables the possibility of systematically identifying novel potential uses for drugs already approved for other indications. As discussed later in this chapter and this book, this strategy is called drug repurposing. Computational calculations have played a significant role in the investigation of molecules that are currently in clinical use. For example, CADD has made notable contributions to the treatment of acquired immunodeficiency syndrome, influenza virus infections, glaucoma, and patients with nonsmall-cell lung cancer (Medina-Franco, Martínez-Mayorga, Juárez-Gordiano, & Castillo, 2007; Talele, Khedkar, & Rigby, 2010).

With new technological advances and the application of CADD techniques it is possible to solve complex problems in the pharmaceutical area. Recent related reviews have been published (Prieto-Martínez & Medina-Franco, 2018a, 2018b; Talevi, 2018). For instance, Saldívar-González, Prieto-Martínez, and Medina-Franco (2017) commented on the need to further increase these technologies and augment the multidisciplinary investigation of new drugs. Other reviews highlight that CADD represents a systematic manner to merge basic and applied science (Das, 2017). For authors such as Usha, Shanmugarajan, Goyal, Kumar, and Middha (2018), CADD includes a collection of pharmacological, pharmacodynamic, and in silico toxicity predictions, which are useful to identify or filter out active or toxic molecules, respectively.

Currently, the development of new computational techniques is allowing a more comprehensive and detailed study of compounds of clinical interest, such as the application of artificial intelligence, big data, chemical space, chemoinformatics, deep learning, molecular modeling, polypharmacology, structure multiple-activity relationships (SmART), target fishing, and virtual screening. These concepts are further commented on the sections later in this chapter.

The goal of this chapter is to provide a general introduction of CADD, covering their principal methods, recent successful applications in the development of compounds that currently on the market, and major challenges. The chapter is organized in six major sections. After this introduction, an overview of current methods used in CADD and their major applications is presented. Theoretical frameworks in CADD are covered briefly. Section 3 mentions recent successful examples of CADD. Section 4 discusses trending concepts and topics in the area. Section 5 covers the major challenges involved in the development and application of computational approaches. Section 6 presents summary conclusions.

2 OVERVIEW OF CURRENT APPROACHES USED IN COMPUTER-AIDED DRUG DESIGN

During the past 30 years the increase in computational power and the availability of chemogenomic data have allowed computational chemistry methods to become an

1. INTRODUCTION

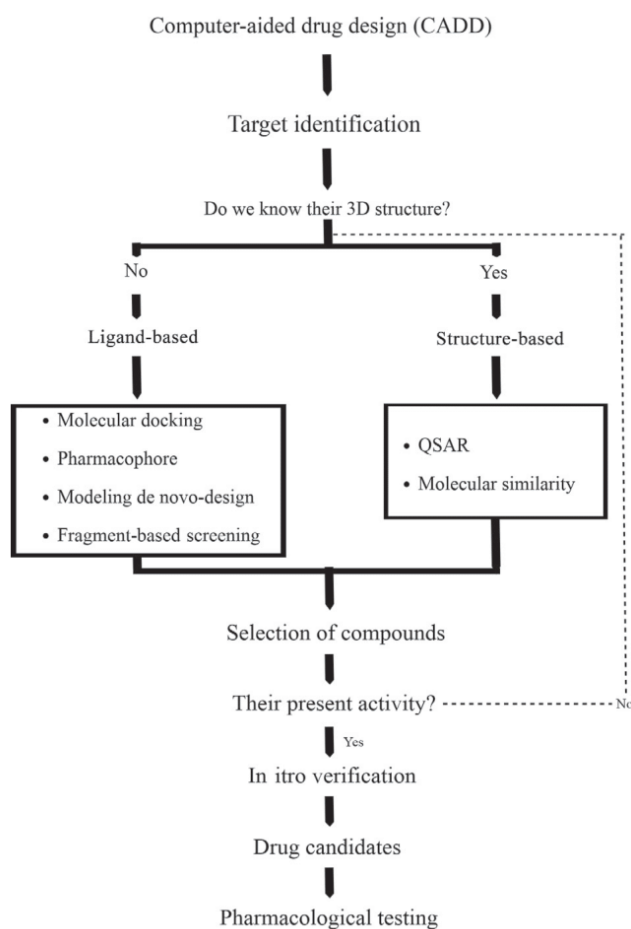


FIG. 1 Schematic overview of a representative computer-aided drug design process.

indispensable part in drug discovery. To date, several marketed drugs, for example, imatinib, zanamivir, and nelfinavir, and several clinical candidates, have been discovered or optimized with the aid of molecular modeling techniques. Fig. 1 outlines a summary of the CADD process with concepts and methods discussed in this chapter.

The concept of “big data” impacts our everyday life, and the area of CADD is not an exception. Through current computational processors it is possible to collect, evaluate, and analyze molecular characteristics in a massive, systematic, and logical manner.

One can make use of the data of each compound to analyze them from different perspectives. In this sense one of the key questions to ask would be: What do I want to analyze? Theoretical chemistry, chemoinformatics, and machine learning (Varnek & Baskin, 2011) provide methods to guide the answers to this question. Using the principles of each major discipline, we can evaluate nearly any kind of “similarity” between molecules. It is through these disciplines that chemistry, biology, and pharmaceutical sciences converge.

For drug-discovery applications, “homology modeling” makes use of structural information available to generate 3D models of biological targets that have not been crystallized

(Sliwoski, Kothiwale, Meiler, & Lowe, 2014). Thus homology modeling is a useful tool to explore and guide, for instance, the structure-based design of novel therapeutic targets or difficult to crystallize, as is the case of calcium channels and some epigenetic and protein complexes.

“Molecular docking” is one of the most used techniques to study 3D ligand-target interactions. Comprehensive reviews of docking have been published recently. One of the main purposes of docking is the generation of models that reveal the possible conformations and, thereafter, evaluate which of them are energetically more viable. However, its application is not limited to characterizing ligand-interactions to increase the potency of active molecules. Docking is also valuable to evaluate specificity and drug resistance, based on 3D structure-property relationships (Pagadala, Syed, & Tuszynski, 2017).

“De novo” design is another group of CADD techniques. De novo design can be roughly compared to a puzzle, where atoms or small fragments are “fitted” into the 3D structure of a binding site. After which these small fragments need to be connected through linkers. One of the challenges may be to find the suitable linkers that can allow molecules to be synthesized in the laboratory. Therefore key questions are: How can we assemble the candidate compounds? How do we evaluate their potential quality? How do we sample the search space effectively? A comprehensive review of de novo design has been covered in the literature (Schneider & Fechner, 2005).

With the advent of technological alternatives for the massive evaluation of compounds or fragments, it has become possible to emphasize the identification of structurally simple hits that can be optimized and generate more powerful ligands. “Fragment-based” screening is based on the fact that a relatively small number of fragments can represent a large fraction of the chemical space. However, for various technical reasons, including the low affinity of the hits fragments and the biophysical methods used for their discovery (e.g., nuclear magnetic resonance, surface plasmon resonance, isothermal calorimetry), fragment-based screening has been limited mainly to in vitro tests with purified proteins (Parker et al., 2017).

It is noteworthy that several CADD techniques make large simplifications of the systems and assume, for the sake of speed, that macromolecules are rigid. A classic example is rigid docking. However, sometimes it is compulsory to consider the dynamics in a model, for example, the binding of small compounds to highly flexibly targets or the simulation of the binding of two macromolecules. To this end molecular dynamic techniques are employed that focus on the use of statistical mechanics, quantum chemistry, and the properties of the electric potential (force field) (Ganesan, Coote, & Barakat, 2017).

The recovery and analysis of chemical information for any type of application in the physical or biological sciences enter the spectrum of chemoinformatics, as well as the relevant computational approaches for the maximized exploration of pharmaceutically relevant compounds. Some of its most used chemoinformatic approaches are quantitative structure-activity relationship (QSAR) and molecular similarity methods (Leach & Gillet, 2007). Overall, QSAR approaches allow the improvement of the pharmacological characteristics of a certain scaffold (basic structure), mediating the determination of the key interactions for a given target, that is, it refines the conformational, spatial, and electronic characteristics of a series of compounds.

“Molecular similarity” is, in principle, a method simpler than QSAR. It is founded in the premise that similar compounds will have a similar activity. One of the main applications of similarity searching is filtering compounds from existing databases, such as a database of compounds approved by the Food and Drug Administration (FDA). Similarity searching of approved drugs is a technique in drug repurposing (Bajorath, 2017).

2.1 Classification of Computer-Aided Drug Design Methods

As described in this chapter, CADD includes multiple approaches to answer questions of a biological-pharmaceutical nature. In general, CADD methods can be classified in three major groups: structure-based, ligand-based, and hybrid methods.

2.1.1 Structure-Based Methods

Structure-based methods depend on the 3D information of the molecular target. Prominent examples of these methods are docking and molecular dynamics (MD) (see also Section 4.5). Applications of structure-based methods include characterization of binding sites, elucidation of the mechanism of action of active molecules at the molecular level, and assessment of the kinetics and thermodynamics involved in ligand-target recognition (Śledź & Caflich, 2018).

2.1.2 Ligand-Based Methods

Ligand-based methods are based on the information of the chemical structures of a set of ligands with known biological activity. One of the main goals of these methods is identifying bioactive compounds or improving the activity of active molecules. Typical examples of ligand-based methods are similarity searching and QSAR modeling (Siju et al., 2017).

2.1.3 Hybrid Methods and Methods Based on End-Points

When the structure of the target is known as well as the structure of active molecules, it is possible to use hybrid or combined methods, i.e., a combination of structure-based and ligand-based methods. An example is certain methods of pharmacophore modeling. Other examples are in silico approaches to predict bioactivity based on the biological profile of compounds tested vs. one or multiple targets (Yongye & Medina-Franco, 2012).

2.2 Main Applications of Computer-Aided Drug Design

CADD has two major applications: identify novel potential active compounds, for example, hit identification, and optimize the bioactivity or ADMETox profile of active compounds, for example, assist in hit-to-lead process.

2.2.1 Hit Finding

A common general CADD approach to identify hit compounds is virtual screening. This technique can be compared to a filtering process: starting from a usually large number of

compounds, structure, ligand, or hybrid approaches are used to select a reduced number of molecules (Siju et al., 2017). The working hypothesis is that the reduced number of compounds have increased probabilities to be active. Of course, experimental validation of selected compounds is mandatory. After the experimental evaluation has been conducted a second round, or more, of filtering steps is performed. In the second, or more, iteration the experimental information of the previous iteration should be considered in the selection of the molecules. Virtual screening is also applicable to filter potential biological targets for a given small molecule. The later process is called inverse virtual screening or target fishing (Yuriev, Holien, & Ramsland, 2015).

2.2.2 Lead Optimization

A number of structure-, ligand-, or hybrid-based methods can be used to improve the potency or reduce side effects of active molecules. Notably, issues with absorption, distribution, metabolism, and excretion (ADME) properties may hamper the development of compounds in the clinic. QSAR and machine learning approaches (see Section 4.4 in this chapter) have been employed to address not only ADME but also toxicity and potency (Caldwell & Yan, 2014).

3 CASE STUDIES: SUCCESSFUL APPLICATIONS OF COMPUTER-AIDED DRUG DESIGN

Nowadays, chemoinformatics and molecular modeling methods are useful in several scientific areas. These approaches are becoming key components in the development of new drugs. Despite the fact the computational results applied to pharmaceutical and medicinal chemistry problems are not 100% accurate, CADD represents an efficient way to help save time and costs as compared to using only experimental approaches. Often two or more methods are used in research projects. This is because CADD complement each other, helping to predict more efficiently active compounds. Table 1 summarizes representative CADD based on 3D structures. The table includes a brief description of their common use with actual approaches in the chemical-biological area. Table 2 summarizes two of the more common methods of CADD based on 2D structures with an example of a recent application.

Several chemoinformatics and CADD methods in general, as exemplified in Tables 1 and 2 (and other sections of this chapter), are being employed to develop drugs that are currently on the market. Some examples are oxymorphone, saquinavir, zanamivir, dorzolamide, and norfloxacin. Table 3 summarizes the information on these drugs and their chemical structures are shown in Fig. 2. Of note, molecular similarity is not stated in Table 3 because this method was employed during the first stages of the design, therefore it is not considered as the principle method for drug discovery. However, molecular similarity principles are commonly used in CADD.

TABLE 1 Representative Computer-Aided Drug Design Approaches Based on 3D Structures

Method	Aim	Example of Recent Approach
Molecular docking	CADD is a promising strategy	Docking suggested that angiotensin II receptor blockers could bind to an active site of kynurenine aminotransferase II. This mechanism action may be advantageous in the treatment of schizophrenia (Zakrocka et al., 2017)
Pharmacophore	Approach to identify in 3D the elements required for the receptor-ligand recognition process. If the ligand is an agonist, the recognition process can lead to the activation of the receptor upon binding. Pharmacophore-based design can be used to guide the chemical modifications to molecules to improve recognition with the receptor and enhance the biological activity. This can be useful to give some indication of the nature of the functional groups in the receptor responsible for binding to the set drugs	Construction of pharmacophore models of the <i>Mycobacterium</i> structural proteome (Lone, Kumar, Athar, & Jha, 2018)
De novo-design	Generation of new molecules with specific and desired properties. In this method it is necessary to use models of the molecular world to produce a trustworthy model that correctly reflects the real world, so it can be used for predicting new molecules that possess the target property reflected in the model	Designed and tested mini-proteins of 37–43 residues that target influenza hemagglutinin and botulinum neurotoxin B (Chevalier et al., 2017)
Fragment-based screening	This approach enables to rapidly scan many molecular fragments that could have very specific interactions with cavities in a binding pocket. The fragments are later connected with linkers with the final goal of generating small molecules that should be synthetically feasible	Recreation of liver tumors that could be avatars for high-throughput drug screening (Fong et al., 2018)

4 TRENDING CONCEPTS AND TECHNOLOGIES

The landscape of drug discovery and development is changing constantly. In this section, we aim to present a concise yet substantial picture of the current trends and emerging concepts on CADD. Table 4 summarizes the main concepts discussed hereafter.

4.1 Big Data

The term “big data” has been mystified. Nowadays in the era of information and the Internet, the quantity of data generated has increased exponentially. Recently it was estimated that the volume of total stored data stacks to nearly two zettabytes, with projections of this

TABLE 2 Representative Computer-Aided Drug Design Approaches Based on 2D Structures

Method	Aim	Example of Recent Approach
Quantitative structure-activity relationship (QSAR)	2D QSAR methods are based on molecular topological descriptors that are represented as graphs; the key elements of the method are the numerical descriptors used to translate a chemical structure into mathematical variables. These methods are focused mainly on the main structural changes that are responsible to modifications in the biological activity In general, the statistical methods for QSAR are used to identify the correlation between the molecular descriptions and biological activity	Molecular docking, QSAR and ADMET studies of withanolide analogs against breast cancer (Yadav et al., 2017)
Molecular similarity	This method is based on the relationship between the chemical structures of two molecules. A common application is similarity searching (similarity-based virtual screening) where active compounds are compared to screening molecules. Selected molecules are tested for biological activity	Application of high-throughput RNA interference screens to problems in cell and developmental biology (Groth, Fish, Nusse, & Calos, 2004)

TABLE 3 Examples of Drugs Recently Discovered With Computer-Aided Drug Design

Drug	Indication	Computer-Aided Drug Design Method	Target	Status
Oxymorphone	Peripheral opioid receptor antagonists	3D molecular docking	Gene name: OPRD1 Gene name: OPRM1	Clinical trials 2015
Saquinavir	Inhibitor of HIV-1 and HIV-2 proteases	Pharmacophore	Uniprot id: Q72874 Gene name: pol	Approved 1995
Zanamivir	Antiviral (influenza A and influenza B)	Modeling de novo-design	Neuraminidase Uniprot id: P27907 Neuraminidase Uniprot id: 06818 Sialidase-2 Uniprot id: P27907	Approved 2000
Dorzolamide	Glaucoma and ocular hypertension	Fragment-based screening	Gene name: CA2 Gene name: CA4 Gene name: CA1 Gene name: CA3	Approved 2012
Norfloxacin	Inhibitor of bacterial DNA gyrase	QSAR	Gene name: gyrA Gene name: parC Gene name: TOP2A	Approved 1998

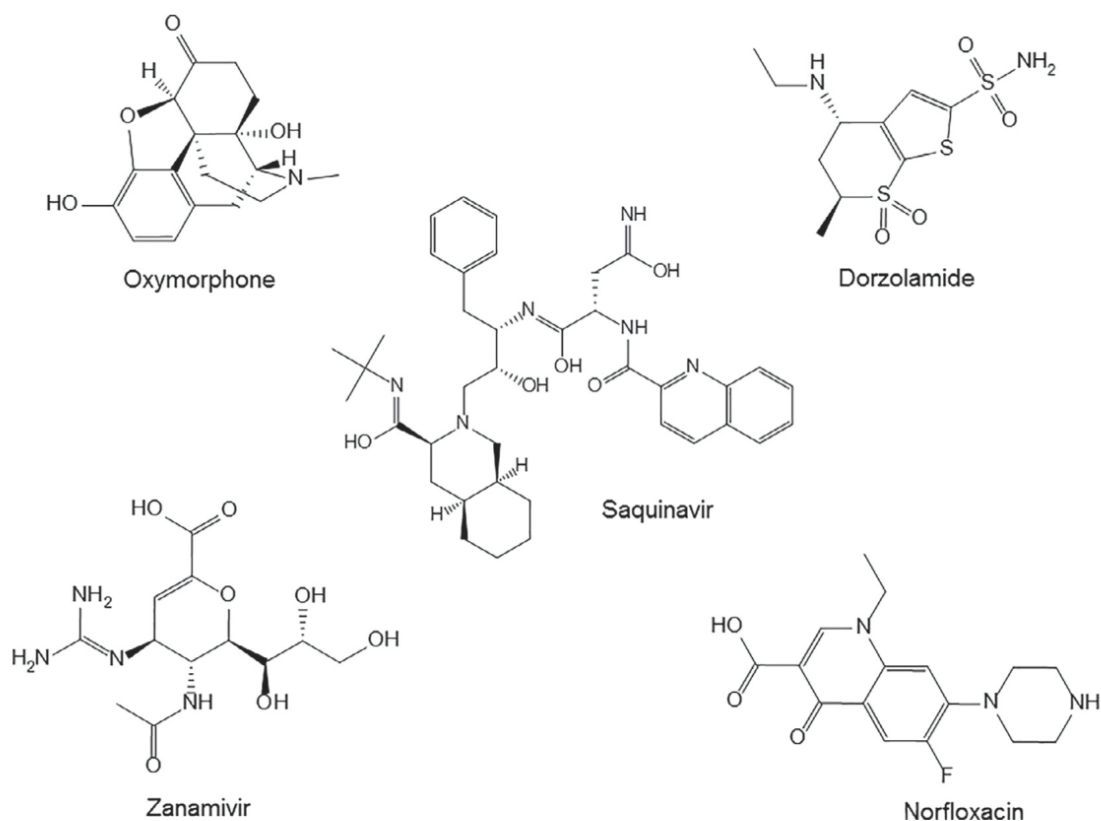


FIG. 2 Chemical structures of approved drugs developed with computer-aided drug design.

TABLE 4 Selected Trending Concepts and Technologies in Computer-Aided Drug Design

Concept	Description	References
Big data	Refers to the complex or large amount of data either in situ or on the public domain	Brown et al. (2018)
Web servers	Web based applications or databases that may be used during a virtual screening campaign	González-Medina et al. (2017)
Workflows	An organized series of algorithms and steps in a virtual screening campaign	Cereto-Massagué et al. (2015)
Machine learning	Algorithms and models that may be selfed, towards optimization and improvement of data	Khamis, Gomaa, and Ahmed (2015)
Deep learning	Implementation of machine learning based on artificial neural networks with nonlinear processing units	Zhang et al. (2017)
Artificial intelligence	Refers to an algorithm or machine, capable of mimicking cognitive functions without supervision or user input	Duch et al. (2007)
Molecular dynamics	Simulation and modeling of molecules (usually proteins) by molecular mechanics and force fields	De Vivo, Masetti, Bottegoni, and Cavalli (2016)

figure doubling every 2 years (Akoka, Comyn-Wattiau, & Laoufi, 2017). Therefore mining this information offers a myriad of possibilities to enhance competitiveness and productivity. But to effectively use big data, one must dive directly in and not just dip one's toes in the shores of this vast ocean. The problem with big data is not just its volume, but its complexity, leading to debate as to the true role and efficiency of statistical thinking in such an arena (Secchi, 2018). Additionally, it often happens that such broad data contains errors, duplicates, and missing values due to overcollection. Hence, preprocessing and curation of data are mandatory to correctly assess the quality of information and avoid any potential bias (Cox, Kartsonaki, & Keogh, 2018). However, data curation may be done differently by different research groups based on experience or previous reports.

The need for a unified or canonical protocol for data curation arises to justify the quality of any given study. Several studies have questioned the role of quality in research due to a major focus on impact (Bornmann, 2012), with others suggesting guidelines towards an objective assessment of the true impact and quality of a given work (Mårtensson, Fors, Wallin, Zander, & Nilsson, 2016). Thus the current paradigm on research stays with a "less-is-more" approach.

Big data has always played a significant role in medicinal chemistry, sometimes indirectly. Methods like combinatorial chemistry and HTS produce large amounts of data over brief periods of time. Previously this was seen like a new dawn for medicinal chemistry, i.e., the ability to process large amounts of new data could reduce the time invested in the drug-development cycle. Consider the following example: human immunodeficiency virus (HIV), a global pandemic for almost 40 years with an estimated of 37 million people infected, and only 57% of patients receiving antiviral therapies (WHO, 2018). Over the years several studies have focused on the inhibition of the viral reverse transcriptase and/or integrase (Cabrerá, Hernández, Chávez, & Medina-Franco, 2018; Ghosh, Osswald, & Prato, 2016). While this approach has proven effective enough, it still has several drawbacks, such as viral resistance and poor bioavailability. During the 1990s, studies on the HIV's entry mechanisms showed the role of CD4⁺ cells and CCR5 chemokine. Chemokines activity is related to their G-coupled receptors (GPCRs), in CCR5's case it is a "C-C" receptor with 75% homology to CCR2 (Barmania & Pepper, 2013). Once CCR5 had been established as an interesting and druggable novel target to tackle HIV, several pharmaceutical companies turned to their GPCRs inhibitor libraries on the hunt for a putative ligand for this protein. Scientists in Pfizer identified an imidazopyridine (UK107,543) as a potent lead based on HTS assays (MacArthur & Novak, 2008). Following molecular optimization, maraviroc (Selzentry) was developed and by 2007 it acquired an approved status for the treatment of HIV-1 from the FDA (Kuritzkes, Kar, & Kirkpatrick, 2008).

This helps to illustrate how medicinal chemistry can be significantly boosted using big data resources. It has been suggested that this has begun to shift the background required for medicinal chemists, towards more data-driven and informatic savvy skills (Lusher, McGuire, Van Schaik, Nicholson, & De Vlieg, 2014). In this regard, scientific research started to develop tools and systems to cope with big data storage and usage; nonetheless one of the main concerns of such platforms is security and privacy for its users (Ebejer, Fulle, Morris, & Finn, 2013). Despite this, public resources offer broad collections that may be exploited beyond the pharmaceutical scope (Kissin, 2018).

In the next section we introduce some of these web-based platforms, especially those used for data mining and virtual screening campaigns.

4.2 Web Servers

As discussed in previous publications, chemoinformatics was developed in direct response to the need for informatic management of chemical data (Chen, 2006). Its role in the drug-discovery process is a stepping stone or gateway to virtual screening. Thus it is very important to know the available resources to aid such ventures. Table 5 presents examples of web-based platforms and their potential role in virtual screening.

As illustrated in Table 5, there are several resources available for CADD, databases being major players in the chemoinformatic landscape. In pharmaceutical research, the nature of this big data is on the complex side, as most drug and/or small molecules have multiple records, for example, biological activity, drug-drug interactions, protein-ligand interactions, etc. Therefore a given analysis can get quite complex as some relations may not be apparent at first glance and naïve assumptions can lead to wrong results or hypotheses (Hu & Bajorath, 2014). For that reason, a virtual screening protocol can be broken down into several steps. The number of cycles of a given campaign depends on its goal, as discussed in the following section.

4.3 Workflows

Let us consider a common case on virtual screening for lead identification and/or optimization. A protocol usually begins by acquiring pertinent data from repositories or *in-house* datasets. The next step is data curation (Fouches, Muratov, & Tropsha, 2015). Next, a query or reference set is selected for comparison and filtering of the datasets (Prieto-Martinez, Fernandez-de Gortari, Mendez-Lucio, & Medina-Franco, 2016). Filtering may be done based on chemical space or similarity metrics, for example, Tanimoto index. Finally, *in silico* testing would involve molecular docking as a means to select lead candidates (Dhananjayan, 2015; Prieto-Martínez & Medina-Franco, 2018c).

The example workflow could be accomplished using several tools, for example, commercial suites of software like Schrödinger or Molecular Operating Environment, as these aim to be complete toolboxes for drug discovery. Academic suites or programs also exist, although their performance is sometimes limited to the available methods or descriptors. However, one of the advantages of open source tools is their flexibility. As these often come as a collection of scripts from a given programming language (R, python, Perl, etc.) advanced users may modify them to cover specific needs.

A third option offers the best of both worlds, found in platforms like the Konstanz information miner (KNIME). Self-described as an open source platform, KNIME is designed to aid data analysis and develop workflows. It offers a wide array of features and integration to many applications by “nodes” that can be connected and rearranged to comply with user needs. Examples of KNIME development include workflows for data integration and cataloguing (Tiwari & Sekhar, 2007) and chemoinformatics (Beisken et al., 2013; Saubern,

TABLE 5 Examples of Web Servers Useful for Chemoinformatics, Drug Discovery and/or Lead Optimization

Name and Website	Description	Type
Binding Database https://www.bindingdb.org/	Public repository containing ~1.1 million reports of binding affinity of protein-ligand complexes	Database
BRENDA https://www.brenda-enzymes.org/	Collection of enzymatic data manually curated from 82,568 proteins	Database
ChEMBL https://www.ebi.ac.uk/chembl/	Public database with more than 14 million activity values from ~ 11,000 targets	Database
Chemspider http://www.chemspider.com/	Repository of small molecules, with >63 million structures	Database
DrugBank https://www.drugbank.ca/	Contains information on more than 10,000 drugs and almost 5000 unique targets	Database
PubChem https://pubchem.ncbi.nlm.nih.gov/	Public database comprising an extensive number of records on compounds, bioactivity, assays and targets	Database
HEMD http://mdl.shsmu.edu.cn/HEMD/	Repository on epigenetic targets data and their chemical modulators	Database
GLIDA http://pharminfo.pharm.kyoto-u.ac.jp/services/glida/	Database providing data on G-coupled receptors and their ligands	Database
PUMA https://www.difacquim.com/d-tools/	Server implementing several analyses on chemical space and diversity of small molecules	Chemoinformatic tool
USR-VS http://usr.marseille.inserm.fr/	Server implementing ultrafast shape recognition algorithms for virtual screening. Queries are compared to almost 94,000 conformers from ZINC	Virtual screening tool
UFSRAT http://opus.bch.ed.ac.uk/ufsrat/	Server for virtual screening, based on the similarity principle. Queries can be compared to more than 10 million conformers from several sources	Virtual screening tool
T-COFFEE http://tcoffee.crg.cat/	Server for homology modeling	Homology modeling tool
SwissModel https://swissmodel.expasy.org/	Server for homology modeling	Homology modeling tool
SwissDock http://www.swissdock.ch/	Web implementation of EADock DSS software, allowing the docking of small molecules based on a manually curated database of protein-ligand interactions	Molecular docking server
Molecular Docking Server https://www.dockingserver.com/web	Web platform for protein and ligand preparation with several methods, while also offering docking and postprocessing capabilities	Molecular docking server
Hex Server http://hexserver.loria.fr/	Server for protein-protein docking based on shape and electrostatics of targets	Protein-protein docking
ZDOCK http://zdock.umassmed.edu/	Server for protein-protein docking based on Fourier transform, to evaluate energy on protein poses	Protein-protein docking

Guha, & Baell, 2011). Additionally, due to the positive reception of the platform, some cloud-based services allow their users to execute KNIME nodes and workflows (Ebejer et al., 2013).

Related tools include the Chemistry Development Kit (CDK, Kuhn, Willighagen, Zielesny, & Steinbeck, 2010), Chembench (Walker, Grulke, Pozefsky, & Tropsha, 2010), Scaffold Hunter (Schäfer et al., 2017), Pipeline pilot (Warr, 2012), and Indigo (Pavlov et al., 2011). Furthermore, these implementations can be enhanced by the introduction of machine learning, enabling batch execution, and self-optimization algorithms.

4.4 Machine Learning

What is learning? Physiologically speaking, learning involves complex cognitive processes all leading to remembering certain conducts or pieces of information for a given purpose. This process can be replicated, to a certain extent, in machines and algorithms (Grosan & Abraham, 2011). For instance, consider junk mail ending in the spam tray. To accomplish this, the server makes a query based on the sender and/or message content, based on a previous registry of user activity it decides the fate of a given message. However, as the reader may know, this process is faulty, as desired mail often ends in spam tray. This helps to illustrate two fundamentals of learning: it begins with the recollection of previous experience/data to make a given choice. But to ensure said choice is the most adequate, superstition and/or false data must be filtered out and discarded (Shalev-Schwartz & Ben-David, 2014).

Based on this, algorithms emulating learning can be developed, but due to this rudimentary form of learning these algorithms will depend on large amounts of data to yield significant choices (Alpaydin, 2004). Fortunately, as seen in Sections 4.2 and 4.3, many sources of information exist.

4.4.1 Applications of Machine Learning in Drug Discovery

With the rise of data repositories like PubChem or ChEMBL, it is now easier to access large amounts of data. This has led to the so-called data-driven discovery: making use of said datasets it is possible to extract and identify patterns yielding predictions. Table 6 presents literature examples of machine learning techniques applied to drug discovery.

As it has been shown, machine learning can be applied to a broad range of problems in drug discovery. Nonetheless, this may be the first step to more complex models. After all, most machine learning methods require user input at some level.

4.4.2 Deep Learning

Deep learning involves scaling machine learning using multilayered neural networks to attempt model abstraction of big data (Cao et al., 2018). Like machine learning, deep learning evolved using a cognitive process as inspiration. In this case, deep learning mimics the perception process using neural connections to extract features based on different observations (Riesenhuber & Poggio, 1999). In other words, deep learning uses neural networks of several topologies trained to identify features corresponding to a complexity scale to recognize patterns in data (Zhang, Tan, Han, & Zhu, 2017).

TABLE 6 Examples of Applications of Machine Learning to Drug Discovery

Application	Result	References
Virtual screening	Lead structures for a novel target in tuberculosis were identified. Also, hit compounds are candidates for drug repurposing	Ekins et al. (2017)
High throughput screening	Fingerprints from imaging data were developed. These allowed the correlation of biological tests and provided activity prediction	Simm et al. (2018)
Protein-ligand interactions	Machine learning methods are good for interaction recognition, although the molecular diversity is not wide enough to allow complete assessment of binding and nonbinding molecules	Colwell (2018)
Side effects prediction	A novel clustering algorithm (K-Seeds) was developed. It showed higher enrichment than other methods	Dimitri and Lió (2017)
Classification of carcinogenic and mutagenic properties	Models showed agreement with experimental knowledge. Multicell descriptors were developed with 70% accuracy	Moorthy, Kumar, and Poongavanam (2017)
Molecular docking accuracy	Effective representations of protein-ligand complexes were developed using the DUD set	Pereira, Caffarena, and Dos Santos (2016)
Binding affinity prediction	Machine learning models can largely benefit from molecular dynamics data. In turn, this can benefit molecular dynamics by means of forcefield improvement	Pérez, Martínez-Rosell, and De Fabritiis (2018)

Artificial neural networks (ANN) are trained to iteratively modify their weight values to be self-optimized. These algorithms were developed in the 1980s and evolved to supporting vector machines and random forests ([Chen, Engkvist, Wang, Olivecrona, & Blaschke, 2018](#)). An important limitation to the efficiency of deep learning methods is their focus on organizing data. ANNs cannot apply their “logic” to decision making beyond pattern recognition, which is a fundamental difference between knowledge and true insight ([Groumpos, 2016](#)).

Nonetheless, deep learning has been applied successfully to organizing, filtering, and mining diverse data sources. Examples include: protein-ligand interaction prediction ([González-Medina, Naveja, Sánchez-Cruz, & Medina-Franco, 2017](#); [Tian, Shao, Wang, Guan, & Zhou, 2016](#)), plant classification by leaf morphology ([Lee, Chan, Mayo, & Remagnino, 2017](#)), nucleotide-protein interactions ([Yi et al., 2018](#)), and HTS data for activity prediction ([Simm et al., 2018](#)).

4.4.3 Artificial Intelligence

Artificial intelligence (AI) is a branch of information and computer science concerned with the embedding of intelligence in machines and computers ([Tzafestas, 2016](#)). Nowadays, we encounter examples of AI in technology such as Apple’s Siri, Amazon’s Alexa, or Microsoft’s Cortana. All of which are digital assistants that serve as data sources and schedule optimizers. How about medical science? AI implementations have made incursion into hospital environments for ICU care and surgical scheduling ([Bini, 2018](#)).

Because of the success of said methods and due to the rise of data-driven paradigms, AI has been applied to diagnosis and medical care (Patel et al., 2009). Notable results have been found in imaging analysis and recognition, where it has been suggested that, although in its infancy, AI may replace physicians in a near future (Krittawong, 2018). However, it is debatable if there is a real difference (at least for now) between deep learning and AI methods (Miller & Brown, 2018).

Hence current developments in AI have been inspired by neuroscience developments to increase model complexity and its capabilities (Hassabis, Kumaran, Summerfield, & Botvinick, 2017). On the other hand, informatic studies on artificial learning methods have shown that the predictive power of models follows Occam's razor: less assumptions yield better results (Pieters & Wiering, 2018).

In medicinal chemistry, AI was used for QSAR and SAR studies (Klopman, 1984) to enhance their performance and predictive power. This has extended to drug design, molecular docking, and predictive toxicology (Duch, Swaminathan, & Meller, 2007). Also, most AI implementations for drug discovery are based on so-called metaheuristics (Ivanciuc, 2009). Metaheuristics involve high-level algorithmic frameworks, which in turn are capable of heuristic optimization, and include: tabu search, simulated annealing, genetic algorithms, ant colony optimization, particle swarm optimization, and neighborhood search (Glover & Sörensen, 2015).

In summary, there is no denying that machine learning and its derivatives can improve drug discovery and development. After all, as of 2017 the pharmaceutical industry as a whole has invested more than 400 million USD in AI development. Nevertheless, it is necessary to remain mildly skeptical on the true merits of AI, as with any other method.

4.5 Molecular Dynamics

Molecular dynamics (MD) have evolved slowly but steadily since the 1970s. A quick search on academic repositories shows the exponential growth of interest on this matter (see Fig. 3). But why now? Considering the background on the subject, this may seem more a late blooming than a breakthrough.

One of the inherent limitations of MD is its scalability. This is due to the reliance on intensive calculations for many particles, thus requiring a high CPU capability to handle the number of calculations needed to model the system. This has led to the development of a dedicated super clusters like Anton (DESRES group, Shaw et al., 2008) that achieves simulations on a microsecond scale. To accomplish this MD often approaches parallelization by domain decomposition. Briefly, this takes the simulation box and splits it into smaller systems assigned to different processors. In practice often several CPUs take a heavy load while most remain idle due to an imbalance in the atom density of the subdomains (Rissland & Deng, 2005). Another issue is latency, as processors must be in constant communication, an inefficient infrastructure results in slow computing (Younge et al., 2015) as CPUs pass more time waiting for responses than running calculations.

An unexpected solution came with the development of graphics processing units (GPUs). Most of the GPU optimization is based on the compute unified device architecture (CUDA) language as developed by NVIDIA. One of the benefits of CUDA is its flexibility, allowing the

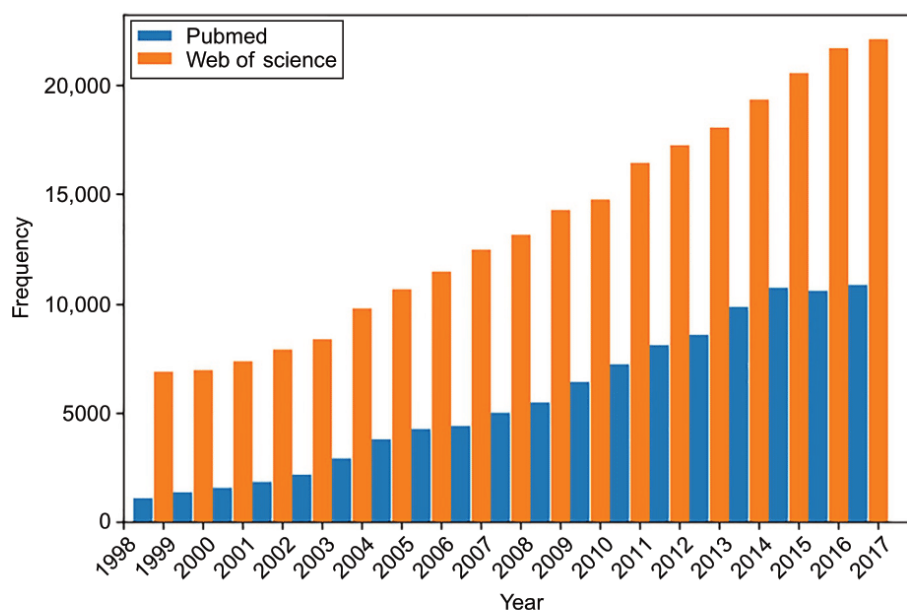


FIG. 3 Frequency of the keywords “molecular dynamics” on two main academic repositories.

use of parallelization outside graphic performance, essentially converting the GPU into a cluster. Hence, most MD software has adopted GPU support to increase their productivity and scalability. A notable example being an all atom simulation of the complete capsid of HIV-1, using nearly 4000 Tesla GPUs (Perilla & Schulten, 2017). The number of CPUs used for such a task is roughly 20,000, a figure hardly accessible to most academic or public clusters. Thus a desktop computer equipped with a GPU can produce around 500ns/day on systems of 20,000 to 30,000 atoms (Fig. 4).

Now we take a general look into the theory behind MD and their general methods.

4.5.1 General Aspects of Molecular Dynamics

MD involves the application of the laws of motion to molecules. To accomplish this several simplifications are due. First, a molecule is considered as a set of spheres while bonds are represented as springs. From here it is possible to assign certain values based on experimental and theoretical information. This process is known as parametrization. The parameter set includes charges, bonded interactions, nonbonding interactions, polarization, and torsions (Vanommeslaeghe, Guvench, & MacKerell, 2014). Take, for example, the general amber forcefield (GAFF), charges are obtained from restrained electrostatic potential (RESP) at HF/6-31G* basis set or semiempirical AM1-BCC method (Wang, Wolf, Caldwell, Kollman, & Case, 2004).

Thus forcefields are born: mathematical expressions used to describe energy dependence with particle location (see Eq. 1). Ideally these expressions must be simple to ensure a quick computation, but substantial enough to reproduce the properties of a given system (González, 2011):

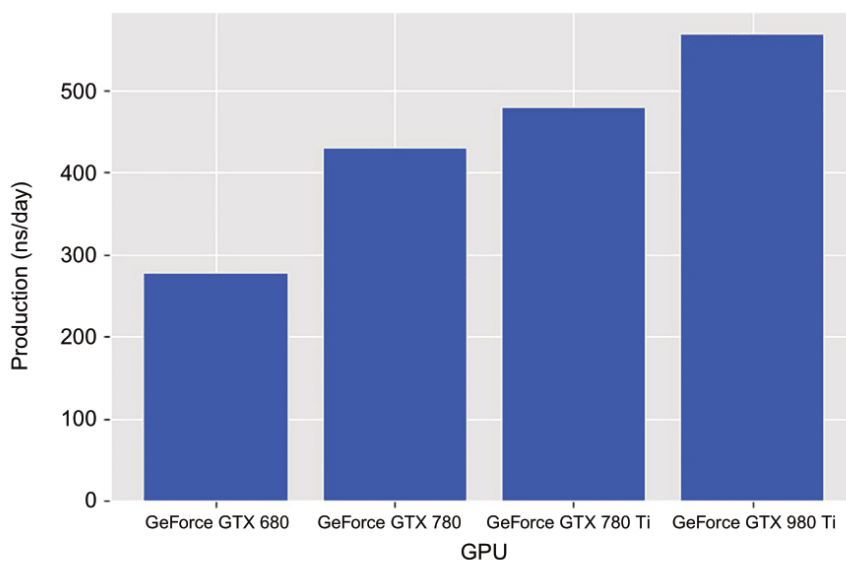


FIG. 4 Molecular dynamics performance with central processing units and graphics processing units, values come from the 2016 Benchmark of Desmond.

EXPRESSION FOR POTENTIAL ENERGY, AS IMPLEMENTED IN GENERAL FORCEFIELDS

$$V(r) = \sum_{\text{bonds}} \kappa_b (b - b_0)^2 + \sum_{\text{angles}} \kappa_\theta (\theta - \theta_0)^2 + \sum_{\text{torsions}} \kappa_\phi [\cos(n\phi + \delta) + 1] + \sum_{\text{nonbond}} \left[\frac{q_i q_j}{r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^6} \right] \quad (1)$$

Therefore it follows that any forcefield requires extensive evaluation and updates to improve their performance. Suffice to say, for the case of proteins, the most prominent examples: AMBER, CHARMM, and OPLS, were initially developed during the 1980s and continue their development today to include lipids or even small molecules (Ponder & Case, 2003).

4.5.2 Applications of Molecular Dynamics in Drug Discovery

In drug discovery, MD has been adopted as the successor of molecular docking. It has been used mainly in the modeling of putative allosteric sites (Papaleo et al., 2016; Spiliotopoulos & Caflisch, 2014) and binding mode studies (Arcon et al., 2017; Clark et al., 2016; Leone, Marinelli, Carloni, & Parrinello, 2010). Recently, MD scalability allows for newer techniques that can solve problems beyond protein flexibility. For example, consider free energy perturbations, which allow calculation of the energy involved in the binding of a ligand and several derivatives (Mortier et al., 2015). Other methods include pH-REMD, an enhanced sampling method used to assess the environment of active sites and correct pKa on sidechains (Sabri Dashti, Meng, & Roitberg, 2012), and QM/MM, a hybrid approach to model electronic

changes based on *ab initio* and physical changes caused by molecular mechanics (Aldeghi, Heifetz, Bodkin, Knapp, & Biggin, 2016). This method allows the study of enzymatic mechanisms such as covalent inhibition.

In summary, MD has come a long way since its inception more than 40 years ago. Gaining a prominent position in drug discovery and molecular modeling to complement other techniques and methods (Alonso, Bliznyuk, & Gready, 2006). This has led to mainstream excitement about its capabilities; however, MD simulations need to improve (i.e., the performance of forcefields) to correctly model protein secondary structures (Beauchamp, Lin, Das, & Pande, 2012; Best, Buchete, & Hummer, 2008). Nonetheless, we can be cautiously optimistic and acknowledge that MD can help us to reach new highs in drug discovery (Ganesan et al., 2017).

5 CHALLENGES AND EMERGING PROBLEMS IN COMPUTER-AIDED DRUG DESIGN

CADD still faces many challenges, which include, but are not limited to: (a) increasing the efficiency of virtual screening; (b) augmenting the number and quality of online computational resources; (c) further developing the field of computational chemogenomics; (d) strengthening the design of drugs aimed at multiple molecular targets; (e) improving the predictive capacity of toxicity models and side effects, and (f) strengthening the interaction with other disciplines to optimize the search for bioactive molecules for the treatment and/or prevention of diseases. The authors' opinions on some of the areas that need improvement in CADD are given in the following section.

5.1 Integration With Other Techniques

CADD should be considered as an integral part of a multidisciplinary effort and not a single approach that will bring drugs to the market in its own (Saldívar-González, Prieto-Martínez, et al., 2017). In most practical applications this is the case and CADD is employed in combination with other technologies, such as HTS and/or combinatorial chemistry. However, there is still room for improvement because there is still a reluctance in some research groups to guide their experiments with the aid of computational methods. As discussed in Section 5, we anticipate that the education of young scientists and established researchers regarding the capabilities and limitations of computational methods will increase productivity and synergy. A brief list of examples of the integration of computational methods with other technologies include:

- CADD and HTS: the computer-guided selection of screening libraries (including the informatic filtering of frequent hitters), informatic analysis of screening data, development of structure-activity relationships, data visualization, organization, storage, and management of output data.
- Natural product-based drug discovery (Medina-Franco, 2015): building, organization, and mining of natural product databases, target fishing, structure elucidation, etc. (González-Medina et al., 2016).

- Combinatorial chemistry-HTS-CADD: computer-guided design of combinatorial libraries for HTS, characterization of the chemical space of combinatorial collections (Lopez-Vallejo et al., 2011), etc.
- Design and analysis (including deconvolution) of phenotypic screening data (Moffat, Vincent, Lee, Eder, & Prunotto, 2017), etc.

5.2 Absorption, Distribution, Metabolism, and Excretion, and Toxicity Prediction

ADME is one of the bottle necks in drug development. Computational methods are playing a key role in anticipating potential ADME and toxicity problems and reducing the number of experiments that involve animal testing. However, predicting toxicity accurately is not a simple task due to the complexity of several toxicity mechanism. A recent review on in silico prediction of chemical toxicity has been published (Yang, Sun, Li, Liu, & Tang, 2018). As such further development of the field of informatics toxicology is anticipated (Mangiatordi et al., 2016).

5.3 Difficult and Emerging Targets

CADD is a promising strategy to tackle the targets that are currently considered difficult, such as protein-protein interactions (PPIs) (Díaz-Eufracio, Naveja, & Medina-Franco, 2018). The rational design and optimization of inhibitors of PPIs represent major challenges due to the nonclassical drug-like properties of these compounds. Similarly, CADD methods offer potential new avenues to address orphan, neglected, or emerging targets with potential applications in the clinic, for example, epigenetic targets (Oprea et al., 2018; Santos et al., 2016).

5.4 Neglected Diseases

There is active research towards the treatment of neglected diseases. The overall lower cost of computational methods compared to traditional experimental approaches can be of significant importance in identifying novel hits and/or optimizing their activity (Andrade et al., 2018; Melo-Filho et al., 2016; Varela, Cobo, Pawar, & Yadav, 2017; Volkamer, Kuhn, Grombacher, Rippmann, & Rarey, 2012). Here, a wide range of CADD methods are playing a crucial role, including virtual screening and database management.

5.5 Chemical Space

Delineating and exploring efficiently the medicinally relevant chemical space (López-Vallejo, Giulianotti, Houghten, & Medina-Franco, 2012) continues to be a challenge. This applies to exploring the chemical space of molecular targets (e.g., epigenetic targets) or exploring the chemical space from the point of view of the compounds that are commercially available (Lucas, Grüning, Bleher, & Günther, 2015; Opassi, Gesù, & Massarotti, 2018) and/or commonly used in HTS campaigns, for example, the dark chemical matter (Macarron, 2015; Siramshetty & Preissner, 2018; Wassermann et al., 2015).

5.6 Advance Multitarget Drug Discovery and Polypharmacology

A current trend in drug discovery is shifting toward multitarget drug discovery (Medina-Franco, Giulianotti, Welmaker, & Houghten, 2013). In this regard, CADD can further address polypharmacology and associated topics:

- Chemogenomics (including virtual screening) (Bajorath, 2013; Rognan, 2013)
- Deconvolution of phenotypic screening: find associations molecules-targets-cell/based data-toxicity, etc. (Prieto et al., 2013)
- Dual- and multitarget drug design (Atkinson et al., 2014; Cabrera et al., 2018)
- Structure multiple-activity relationships (SmARTs) (Saldívar-González, Naveja, Palomino-Hernández, & Medina-Franco, 2017)
- Drug repurposing (Brindha et al., 2017; Liu et al., 2013).

5.7 Training, Teaching, and Divulcation

In addition to continuing to advance the generation of knowledge and techniques to improve CADD, it is necessary to improve education at undergraduate and graduate levels (Díaz-Eufracio et al., 2018). Current learning is on the fly (driven by specific needs) but training can be greatly aided with the development of text books focused on CADD. Formal training can be complemented with the current availability of conferences focused on CADD.

6 CONCLUSIONS

CADD is part of multidisciplinary efforts that have proved to be useful in the development of drugs that are currently in clinical use or in advanced clinical trials. As such CADD is a combination of several theoretical and computational disciplines, such as molecular modeling, chemoinformatics, theoretical chemistry, among others. There are a broad range of computational approaches used in CADD that have been used for several years, such as molecular docking, dynamics, QSAR, similarity searching, to name a few. However, these approaches continue to be improved and refined. Also, there are several novel concepts and approaches that are driving CADD into the future. Clear instances of the latter are “big data,” artificial intelligence, machine and deep learning, among others. All new and classical approaches are being used in combination with novel trends in drug discovery, such as polypharmacology and drug repurposing. Like other multidisciplinary approaches, CADD faces several challenges that include not only the refinement of the theoretical basis, but also its rational application (with the knowledge of the limitations), and training and education of the users of these technologies.

Acknowledgments

This work was supported by the *Programa de Apoyo a Proyectos para la Innovación y Mejoramiento de la Enseñanza* (PAPIME) grant PE200118, *Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica* (PAPIIT) grant IA203718 and National Council of Science and Technology (CONACyT), Mexico grant number 282785. FD P-M acknowledges the PhD scholarship from CONACyT no. 660465/576637. KE J-M thanks CONACyT for the support (No. 16231).

References

- Akoka, J., Comyn-Wattiau, I., & Laoufi, N. (2017). Research on big data—a systematic mapping study. *Computer Standards & Interfaces*, 54, 105–115.
- Aldeghi, M., Heifetz, A., Bodkin, M. J., Knapp, S., & Biggin, P. C. (2016). Accurate calculation of the absolute free energy of binding for drug molecules. *Chemical Science*, 7(1), 207–218.
- Alonso, H., Bliznyuk, A. A., & Gready, J. E. (2006). Combining docking and molecular dynamic simulations in drug design. *Medicinal Research Reviews*, 26(5), 531–568.
- Alpaydin, E. (2004). *Introduction to machine learning (adaptive computation and machine learning)*. Cambridge, MA: The MIT Press.
- Andrade, C. H., Neves, B. J., Melo-Filho, C. C., Rodrigues, J., Silva, D. C., Braga, R. C., & Cravo, P. V. L. (2018). In silico chemogenomics drug repositioning strategies for neglected tropical diseases. *Current Medicinal Chemistry*. <https://dx.doi.org/10.2174/0929867325666180309114824> [in press].
- Arcon, J. P., Defelipe, L. A., Modenutti, C. P., López, E. D., Alvarez-Garcia, D., Barril, X., ... Martí, M. A. (2017). Molecular dynamics in mixed solvents reveals protein-ligand interactions, improves docking, and allows accurate binding free energy predictions. *Journal of Chemical Information and Modeling*, 57, 846–863.
- Atkinson, S. J., Soden, P. E., Angell, D. C., Bantscheff, M., Chung, C., Giblin, K. A., ... Prinjha, R. K. (2014). The structure based design of dual HDAC/BET inhibitors as novel epigenetic probes. *Medicinal Chemical Communications*, 5, 342–351.
- Bajorath, J. (2013). A perspective on computational chemogenomics. *Molecular Informatics*, 32, 1025–1028.
- Bajorath, J. (2017). *Molecular similarity concepts for informatics applications*: (pp. 231–245). New York, NY: Humana Press. https://dx.doi.org/10.1007/978-1-4939-6613-4_13.
- Barmania, F., & Pepper, M. S. (2013). C-C chemokine receptor type five (CCR5): an emerging target for the control of HIV infection. *Applied & Translational Genomics*, 2, 3–16.
- Beauchamp, K. A., Lin, Y.-S., Das, R., & Pande, V. S. (2012). Are protein force fields getting better? A systematic benchmark on 524 diverse NMR measurements. *Journal of Chemical Theory and Computation*, 8, 1409–1414.
- Beisken, S., Meinel, T., Wiswedel, B., de Figueiredo, L. F., Berthold, M., & Steinbeck, C. (2013). KNIME-CDK: workflow-driven cheminformatics. *BMC Bioinformatics*, 14, 257.
- Best, R. B., Buchete, N.-V., & Hummer, G. (2008). Are current molecular dynamics force fields too helical? *Biophysical Journal*, 95, L07–L09.
- Bini, S. A. (2018). Artificial intelligence, machine learning, deep learning, and cognitive computing: what do these terms mean and how will they impact health care? *The Journal of Arthroplasty*, 33, 2358–2361.
- Bornmann, L. (2012). Measuring the societal impact of research. *EMBO Reports*, 13, 673–676.
- Brindha, S., Vincent, S., Velmurugan, D., Ananthakrishnan, D., Sundaramurthi, J. C., & Gnanadoss, J. J. (2017). Bioinformatics approach to prioritize known drugs towards repurposing for tuberculosis. *Medical Hypotheses*, 103, 39–45.
- Brown, N., Cambruzzi, J., Cox, P. J., Davies, M., Dunbar, J., Plumbley, D., ... Sheppard, D. W. (2018). *Big data in drug discovery*. <https://dx.doi.org/10.1016/bs.pmch.2017.12.003>.
- Cabrera, A., Hernández, L. H., Chávez, D., & Medina-Franco, J. L. (2018). Molecular modeling of potential dual inhibitors of HIV reverse transcriptase and integrase. *Computational Molecular Bioscience*, 8, 1–41.
- Caldwell, G. W., & Yan, Z. (Eds.), (2014). *Optimization in drug discovery*. Totowa, NJ: Humana Press. <https://dx.doi.org/10.1007/978-1-62703-742-6>.
- Cao, C., Liu, F., Tan, H., Song, D., Shu, W., Li, W., ... Xie, Z. (2018). Deep learning and its applications in biomedicine. *Genomics, Proteomics & Bioinformatics*, 16, 17–32.
- Cereto-Massagué, A., Ojeda, M. J., Valls, C., Mulero, M., Pujadas, G., & Garcia-Vallve, S. (2015). Tools for in silico target fishing. *Methods*, 71, 98–103.
- Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., & Blaschke, T. (2018). The rise of deep learning in drug discovery. *Drug Discovery Today*, 23, 1241–1250.
- Chen, W. L. (2006). Chemoinformatics: past, present, and future. *Journal of Chemical Information and Modeling*, 46, 2230–2255.
- Chevalier, A., Silva, D. -A., Rocklin, G. J., Hicks, D. R., Vergara, R., Murapa, P., ... Baker, D. (2017). Massively parallel de novo protein design for targeted therapeutics. *Nature*, 550, 74.
- Clark, A. J., Tiwary, P., Borrelli, K., Feng, S., Miller, E. B., Abel, R., ... Berne, B. J. (2016). Prediction of protein–ligand binding poses via a combination of induced fit docking and metadynamics simulations. *Journal of Chemical Theory and Computation*, 12, 2990–2998.

1. INTRODUCTION

- Colwell, L. J. (2018). Statistical and machine learning approaches to predicting protein–ligand interactions. *Current Opinion in Structural Biology*, 49, 123–128.
- Cox, D. R., Kartsonaki, C., & Keogh, R. H. (2018). Big data: some statistical issues. *Statistics & Probability Letters*. <https://dx.doi.org/10.1016/J.SPL.2018.02.015> [in press].
- Das, P. S. (2017). A review on computer aided drug design in drug discovery. *World Journal of Pharmacy and Pharmaceutical Sciences*, 6, 279–291.
- De Vivo, M., Masetti, M., Bottegoni, G., & Cavalli, A. (2016). Role of molecular dynamics and related methods in drug discovery. *Journal of Medicinal Chemistry*, 59, 4035–4061.
- Dhananjayan, K. (2015). Molecular docking study characterization of rare flavonoids at the Nac-binding site of the first Bromodomain of BRD4 (BRD4 BD1). *Journal of Cancer Research*, 2015, 1–15.
- Díaz-Eufracio, B. I., Naveja, J. J., & Medina-Franco, J. L. (2018). Protein–protein interaction modulators for epigenetic therapies. *Advances in Protein Chemistry and Structural Biology*, 110, 65–84.
- Dimitri, G. M., & Lió, P. (2017). DrugClust: a machine learning approach for drugs side effects prediction. *Computational Biology and Chemistry*, 68, 204–210.
- Duch, W., Swaminathan, K., & Meller, J. (2007). Artificial intelligence approaches for rational drug design and discovery. *Current Pharmaceutical Design*, 13, 1497–1508.
- Ebejer, J.-P., Fulle, S., Morris, G. M., & Finn, P. W. (2013). The emerging role of cloud computing in molecular modelling. *Journal of Molecular Graphics and Modelling*, 44, 177–187.
- Ekins, S., Godbole, A. A., Kéri, G., Orfi, L., Pato, J., Bhat, R. S., ... Nagaraja, V. (2017). Machine learning and docking models for Mycobacterium tuberculosis topoisomerase I. *Tuberculosis (Edinburgh, Scotland)*, 103, 52–60.
- Fong, E. L. S., Toh, T. B., Lin, Q. X. X., Liu, Z., Hooi, L., Mohd Abdul Rashid, M. B., ... Yu, H. (2018). Generation of matched patient-derived xenograft in vitro-in vivo models using 3D macroporous hydrogels for the study of liver cancer. *Biomaterials*, 159, 229–240.
- Fourches, D., Muratov, E., & Tropsha, A. (2015). Curation of chemogenomics data. *Nature Chemical Biology*, 11(8), 535. <https://dx.doi.org/10.1038/nchembio.1881>.
- Ganesan, A., Coote, M. L., & Barakat, K. (2017). Molecular dynamics-driven drug discovery: leaping forward with confidence. *Drug Discovery Today*, 22, 249–269.
- Ghosh, A. K., Osswald, H. L., & Prato, G. (2016). Recent progress in the development of HIV-1 protease inhibitors for the treatment of HIV/AIDS. *Journal of Medicinal Chemistry*, 59, 5172–5208.
- Glover, F., & Sörensen, K. (2015). Metaheuristics. *Scholarpedia*, 10, 6532.
- González, M. A. (2011). Force fields and molecular dynamics simulations. *École Thématique de La Société Française de La Neutronique*, 12, 169–200.
- González-Medina, M., Prieto-Martínez, F. D., Naveja, J. J., Méndez-Lucio, O., El-Elimat, T., Pearce, C. J., ... Medina-Franco, J. L. (2016). Chemoinformatic expedition of the chemical space of fungal products. *Future Medicinal Chemistry*, 6(March), 1113–1126.
- González-Medina, M., Naveja, J. J., Sánchez-Cruz, N., & Medina-Franco, J. L. (2017). Open chemoinformatic resources to explore the structure, properties and chemical space of molecules. *RSC Advances*, 7, 54153–54163.
- Grosan, C., & Abraham, A. (Eds.), (2011). *Machine learning* (pp. 261–268). In *Intelligent systems*. Springer.
- Groth, A. C., Fish, M., Nusse, R., & Calos, M. P. (2004). Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. *Genetics*, 166, 1775–1782.
- Groumpos, P. P. (2016). Deep learning vs. wise learning: a critical and challenging overview. *IFAC-PapersOnLine*, 49, 180–189.
- Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. (2017). Neuroscience-inspired artificial intelligence. *Neuron*, 95, 245–258.
- Hu, Y., & Bajorath, J. (2014). Learning from “big data”: compounds and targets. *Drug Discovery Today*, 19, 357–360.
- Ivanciuc, O. (2009). Drug design with artificial intelligence methods. In R. A. Meyers (Ed.), *Encyclopedia of complexity and systems science* (pp. 2113–2139). New York, NY: Springer.
- Khamis, M. A., Gomaa, W., & Ahmed, W. F. (2015). Machine learning in computational docking. *Artificial Intelligence in Medicine*. <https://dx.doi.org/10.1016/j.artmed.2015.02.002>.
- Kissin, I. (2018). What can big data on academic interest reveal about a drug? Reflections in three major US databases. *Trends in Pharmacological Sciences*, 39, 248–257.
- Klopman, G. (1984). Artificial intelligence approach to structure-activity studies. Computer automated structure evaluation of biological activity of organic molecules. *Journal of the American Chemical Society*, 106, 7315–7321.

1. INTRODUCTION

- Krittanawong, C. (2018). The rise of artificial intelligence and the uncertain future for physicians. *European Journal of Internal Medicine*. (February 1). <https://dx.doi.org/10.1016/j.ejim.2017.06.017> [in press].
- Kuhn, T., Willighagen, E. L., Zielesny, A., & Steinbeck, C. (2010). CDK-Taverna: an open workflow environment for cheminformatics. *BMC Bioinformatics*, 11, 159.
- Kuritzkes, D., Kar, S., & Kirkpatrick, P. (2008). Maraviroc. *Nature Reviews Drug Discovery*, 7, 15–16.
- Leach, A. R., & Gillet, V. J. (2007). *An introduction to chemoinformatics*. Dordrecht: Springer.
- Lee, S. H., Chan, C. S., Mayo, S. J., & Remagnino, P. (2017). How deep learning extracts and learns leaf features for plant classification. *Pattern Recognition*, 71, 1–13.
- Leone, V., Marinelli, F., Carloni, P., & Parrinello, M. (2010). Targeting biomolecular flexibility with metadynamics. *Current Opinion in Structural Biology*, 20, 148–154.
- Liu, Z., Fang, H., Reagan, K., Xu, X., Mendrick, D. L., Slikker, W., & Tong, W. (2013). In silico drug repositioning-what we need to know. *Drug Discovery Today*, 18, 110–115.
- Lone, M. Y., Kumar, S. P., Athar, M., & Jha, P. C. (2018). Exploration of Mycobacterium tuberculosis structural proteome: An in-silico approach. *Journal of Theoretical Biology*, 439, 14–23.
- Lopez-Vallejo, F., Caulfield, T., Martinez-Mayorga, K. A., Giulianotti, M., Nefzi, A., Houghten, R., & Medina-Franco, J. (2011). Integrating virtual screening and combinatorial chemistry for accelerated drug discovery. *Combinatorial Chemistry & High Throughput Screening*, 14, 475–487.
- López-Vallejo, F., Giulianotti, M. A., Houghten, R. A., & Medina-Franco, J. L. (2012). Expanding the medicinally relevant chemical space with compound libraries. *Drug Discovery Today*, 17, 718–726.
- Lucas, X., Grüning, B. A., Bleher, S., & Günther, S. (2015). The purchasable chemical space: a detailed picture. *Journal of Chemical Information and Modeling*, 55, 915–924.
- Lusher, S. J., McGuire, R., Van Schaik, R. C., Nicholson, C. D., & De Vlieg, J. (2014). Data-driven medicinal chemistry in the era of big data. *Drug Discovery Today*, 19, 859–868.
- Macarron, R. (2015). Chemical libraries: how dark is HTS dark matter? *Nature Chemical Biology*, 11, 904–905.
- MacArthur, R. D., & Novak, R. M. (2008). Reviews of anti-infective agents: maraviroc: the first of a new class of antiretroviral agents. *Clinical Infectious Diseases*, 47, 236–241.
- Mangiatiordi, G. F., Alberga, D., Altomare, C. D., Carotti, A., Catto, M., Cellamare, S., ... Nicolotti, O. (2016). Mind the gap! A journey towards computational toxicology. *Molecular Informatics*, 35, 294–308.
- Mårtensson, P., Fors, U., Wallin, S. B., Zander, U., & Nilsson, G. H. (2016). Evaluating research: a multidisciplinary approach to assessing research practice and quality. *Research Policy*, 45, 593–603.
- Martin, L., Hutchens, M., Hawkins, C., & Radnov, A. (2017). How much do clinical trials cost? *Nature Reviews Drug Discovery*, 16(6), 381–382.
- Medina-Franco, J. L. (2015). Discovery and development of lead compounds from natural sources using computational approaches. In P. K. Mukherjee (Ed.), *Evidence-based validation of herbal medicine* (pp. 455–475): Elsevier.
- Medina-Franco, J. L., Martínez-Mayorga, K., Juárez-Gordiano, C., & Castillo, R. (2007). Pyridin-2(1H)-ones: a promising class of HIV-1 non-nucleoside reverse transcriptase inhibitors. *ChemMedChem*, (August 13).
- Medina-Franco, J. L., Giulianotti, M. A., Welmaker, G. S., & Houghten, R. A. (2013). Shifting from the single to the multitarget paradigm in drug discovery. *Drug Discovery Today*, 18, 495–501.
- Melo-Filho, C. C., Dantas, R. F., Braga, R. C., Neves, B. J., Senger, M. R., Valente, W. C. G., ... Andrade, C. H. (2016). QSAR-driven discovery of novel chemical scaffolds active against *Schistosoma mansoni*. *Journal of Chemical Information and Modeling*, 56, 1357–1372.
- Miller, D. D., & Brown, E. W. (2018). Artificial intelligence in medical practice: the question to the answer? *The American Journal of Medicine*, 131, 129–133.
- Moffat, J. G., Vincent, F., Lee, J. A., Eder, J., & Prunotto, M. (2017). Opportunities and challenges in phenotypic drug discovery: an industry perspective. *Nature Reviews Drug Discovery*, 16, 531–543.
- Moorthy, N. H. N., Kumar, S., & Poongavanam, V. (2017). Classification of carcinogenic and mutagenic properties using machine learning method. *Computational Toxicology*, 3, 33–43.
- Mortier, J., Rakers, C., Bermudez, M., Murgueitio, M. S., Riniker, S., & Wolber, G. (2015). The impact of molecular dynamics on drug design: applications for the characterization of ligand-macromolecule complexes. *Drug Discovery Today*, 20, 686–702.
- Opassi, G., Gesù, A., & Massarotti, A. (2018). The hitchhiker's guide to the chemical-biological galaxy. *Drug Discovery Today*, 23, 565–574.

1. INTRODUCTION

- Oprea, T. I., Bologa, C. G., Brunak, S., Campbell, A., Gan, G. N., Gaulton, A., ... Zahoránszky-Köhalmi, G. (2018). Unexplored therapeutic opportunities in the human genome. *Nature Reviews Drug Discovery*, 17, 317–332.
- Pagadala, N. S., Syed, K., & Tuszynski, J. (2017). Software for molecular docking: a review. *Biophysical Reviews*, 9, 91–102.
- Papaleo, E., Saladino, G., Lambrugh, M., Lindorff-Larsen, K., Gervasio, F. L., & Nussinov, R. (2016). The role of protein loops and linkers in conformational dynamics and allostery. *Chemical Reviews*, 116, 6391–6423.
- Parker, C. G., Galmozzi, A., Wang, Y., Correia, B. E., Sasaki, K., Joslyn, C. M., ... Cravatt, B. F. (2017). Ligand and target discovery by fragment-based screening in human cells. *Cell*, 168, 527–541.e29.
- Patel, V. L., Shortliffe, E. H., Stefanelli, M., Szolovits, P., Berthold, M. R., Bellazzi, R., & Abu-Hanna, A. (2009). The coming of age of artificial intelligence in medicine. *Artificial Intelligence in Medicine*, 46, 5–17.
- Pavlov, D., Rybalkin, M., Karulin, B., Kozhevnikov, M., Savelyev, A., & Churinov, A. (2011). Indigo: universal cheminformatics API. *Journal of Cheminformatics*, 3(Suppl. 1), P4.
- Pereira, J. C., Caffarena, E. R., & Dos Santos, C. N. (2016). Boosting docking-based virtual screening with deep learning. *Journal of Chemical Information and Modeling*, 56, 2495–2506.
- Pérez, A., Martínez-Rosell, G., & De Fabritiis, G. (2018). Simulations meet machine learning in structural biology. *Current Opinion in Structural Biology*, 49, 139–144.
- Perilla, J. R., & Schulten, K. (2017). Physical properties of the HIV-1 capsid from all-atom molecular dynamics simulations. *Nature Communications*, 8, 15959.
- Pieters, M., & Wiering, M. (2018). *Comparison of machine learning techniques for multi-label genre classification* (pp. 131–144). Cham: Springer.
- Ponder, J. W., & Case, D. A. (2003). Force fields for protein simulations. *Advances in Protein Chemistry*, 66, 27–85.
- Prieto, P., Kinsner-Ovaskainen, A., Stanzel, S., Albella, B., Artursson, P., Campillo, N., ... Kopp-Schneider, A. (2013). The value of selected in vitro and in silico methods to predict acute oral toxicity in a regulatory context: results from the European Project ACuteTox. *Toxicology In Vitro*, 27, 1357–1376.
- Prieto-Martínez, F. D., & Medina-Franco, J. L. (2018a). Charting the bromodomain BRD4: towards the identification of novel inhibitors with molecular similarity and receptor mapping. *Letters in Drug Design & Discovery*, 15, 1002–1011.
- Prieto-Martínez, F. D., & Medina-Franco, J. L. (2018b). Diseño de fármacos asistido por computadora: cuando la informática, la química y el arte se encuentran. *TIP Revista Especializada En Ciencias Químico-Biológicas*, 21, 124–134.
- Prieto-Martínez, F. D., & Medina-Franco, J. L. (2018c). Molecular docking: Current advances and challenges. *TIP Revista Especializada en Ciencias Químico-Biológicas*, 21, 65–87.
- Prieto-Martínez, F. D., Fernández-de Gortari, E., Méndez-Lucio, O., & Medina-Franco, J. L. (2016). A chemical space odyssey of inhibitors of histone deacetylases and bromodomains. *RSC Advances*, 6, 56225–56239.
- Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2, 1019–1025.
- Rissland, P., & Deng, Y. (2005). Supercomputers and their effect on molecular dynamics. *IEEE Potentials*, 24, 9–12.
- Rognan, D. (2013). Towards the next generation of computational chemogenomics tools. *Molecular Informatics*, 32, 1029–1034.
- Sabri Dashti, D., Meng, Y., & Roitberg, A. E. (2012). pH-replica exchange molecular dynamics in proteins using a discrete protonation method. *The Journal of Physical Chemistry B*, 116, 8805–8811.
- Saldívar-González, F. I., Naveja, J. J., Palomino-Hernández, O., & Medina-Franco, J. L. (2017). Getting SMART in drug discovery: cheminformatics approaches for mining structure–multiple activity relationships. *RSC Advances*, 7, 632–641.
- Saldívar-González, F. I., Prieto-Martínez, F. D., & Medina-Franco, J. L. (2017). Descubrimiento y desarrollo de fármacos: un enfoque computacional. *Educación Química*, 28, 51–58.
- Santos, R., Ursu, O., Gaulton, A., Bento, A. P., Donadi, R. S., Bologa, C. G., ... Overington, J. P. (2016). A comprehensive map of molecular drug targets. *Nature Reviews Drug Discovery*, 16, 19–34.
- Saubern, S., Guha, R., & Baell, J. B. (2011). KNIME workflow to assess PAINS filters in SMARTS format. Comparison of RDKit and indigo cheminformatics libraries. *Molecular Informatics*, 30, 847–850.
- Schäfer, T., Kriege, N., Humbeck, L., Klein, K., Koch, O., & Mutzel, P. (2017). Scaffold hunter: a comprehensive visual analytics framework for drug discovery. *Journal of Cheminformatics*, 9, 28.
- Schneider, G., & Fechner, U. (2005). Computer-based de novo design of drug-like molecules. *Nature Reviews Drug Discovery*, 4, 649–663.

1. INTRODUCTION

- Secchi, P. (2018). On the role of statistics in the era of big data: a call for a debate. *Statistics & Probability Letters*, 136, 10–14.
- Shalev-Schwartz, S., & Ben-David, S. (2014). *Understanding machine learning: from theory to algorithms*. New York: Cambridge University Press.
- Shaw, D. E., Chao, J. C., Eastwood, M. P., Gagliardo, J., Grossman, J. P., Ho, C. R., ... Bowers, K. J. (2008). Anton, a special-purpose machine for molecular dynamics simulation. *Communications of the ACM*, 51, 91.
- Siju, E. N., Rajalakshmi, G. R., Paulose, A. P., Dhanya, F. T., Hariraj, N., & Rahul, K. (2017). CADD: pharmacological approaches in drug design and drug discovery. *World Journal of Pharmacy and Pharmaceutical Sciences*, 6, 892–908.
- Simm, J., Klambauer, G., Arany, A., Steijaert, M., Wegner, J. K., Gustin, E., ... Ceulemans, H. (2018). Repurposing high-throughput image assays enables biological activity prediction for drug discovery. *Cell Chemical Biology*, 25, 611–618.e3.
- Siramshetty, V. B., & Preissner, R. (2018). Drugs as habitable planets in the space of dark chemical matter. *Drug Discovery Today*, 23, 481–486.
- Śledź, P., & Caflisch, A. (2018). Protein structure-based drug design: from docking to molecular dynamics. *Current Opinion in Structural Biology*, 48, 93–102.
- Sliwoski, G., Kothiwale, S., Meiler, J., & Lowe, E. W. (2014). Computational methods in drug discovery. *Pharmacological Reviews*, 66, 334–395.
- Spiliotopoulos, D., & Caflisch, A. (2014). Molecular dynamics simulations of bromodomains reveal binding-site flexibility and multiple binding modes of the natural ligand acetyl-lysine. *Israel Journal of Chemistry*, 54, 1084–1092.
- Talele, T., Khedkar, S., & Rigby, A. (2010). Successful applications of computer aided drug discovery: moving drugs from concept to the clinic. *Current Topics in Medicinal Chemistry*, 10, 127–141.
- Talevi, A. (2018). *Computer-aided drug design: an overview* (pp. 1–19). New York, NY: Humana Press.
- Tian, K., Shao, M., Wang, Y., Guan, J., & Zhou, S. (2016). Boosting compound-protein interaction prediction by deep learning. *Methods*, 110, 64–72.
- Tiwari, A., & Sekhar, A. K. T. (2007). Workflow based framework for life science informatics. *Computational Biology and Chemistry*, 31, 305–319.
- Tzafestas, S. G. (2016). *Artificial intelligence*: (pp. 25–33). Cham: Springer. https://dx.doi.org/10.1007/978-3-319-21714-7_3.
- Usha, T., Shanmugarajan, D., Goyal, A. K., Kumar, C. S., & Middha, S. K. (2018). Recent updates on computer-aided drug discovery: time for a paradigm shift. *Current Topics in Medicinal Chemistry*, 17, 3296–3307.
- Vanommeslaeghe, K., Guvench, O., & MacKerell, A. D. (2014). Molecular mechanics. *Current Pharmaceutical Design*, 20, 3281–3292.
- Varela, J. N., Cobo, M. F. L., Pawar, S. V., & Yadav, V. G. (2017). Cheminformatic analysis of antimalarial chemical space illuminates therapeutic mechanisms and offers strategies for therapy development. *Journal of Chemical Information and Modeling*, 57, 2119–2131.
- Varnek, A., & Baskin, I. I. (2011). Chemoinformatics as a theoretical chemistry discipline. *Molecular Informatics*, 30, 20–32.
- Volkamer, A., Kuhn, D., Grombacher, T., Rippmann, F., & Rarey, M. (2012). Combining global and local measures for structure-based druggability predictions. *Journal of Chemical Information and Modeling*, 52, 360–372.
- Walker, T., Grulke, C. M., Pozefsky, D., & Tropsha, A. (2010). Chembench: a cheminformatics workbench. *Bioinformatics*, 26, 3000–3001.
- Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., & Case, D. A. (2004). Development and testing of a general amber force field. *Journal of Computational Chemistry*, 25, 1157–1174.
- Warr, W. A. (2012). Scientific workflow systems: pipeline pilot and KNIME. *Journal of Computer-Aided Molecular Design*, 26, 801–804.
- Wassermann, A. M., Lounkine, E., Hoepfner, D., Le Goff, G., King, F. J., Studer, C., ... Glick, M. (2015). Dark chemical matter as a promising starting point for drug lead discovery. *Nature Chemical Biology*, 11, 958–966.
- WHO. (2018). *World Health Organization*. Retrieved from <http://www.who.int/en/>. Accessed 24 March 2018.
- Yadav, D. K., Kumar, S., Saloni, Singh, H., Kim, M. H., Sharma, P., ... Khan, F. (2017). Molecular docking, QSAR and ADMET studies of withanolide analogs against breast cancer. *Drug Design, Development and Therapy*, 11, 1859–1870.
- Yang, H., Sun, L., Li, W., Liu, G., & Tang, Y. (2018). In silico prediction of chemical toxicity for drug design using machine learning methods and structural alerts. *Frontiers in Chemistry*, 6, 30.

1. INTRODUCTION






- Yi, H.-C., You, Z.-H., Huang, D.-S., Li, X., Jiang, T.-H., & Li, L.-P. (2018). A deep learning framework for robust and accurate prediction of ncRNA-protein interactions using evolutionary information. *Molecular Therapy—Nucleic Acids*, *11*, 337–344.
- Yongye, A. B., & Medina-Franco, J. L. (2012). Data mining of protein-binding profiling data identifies structural modifications that distinguish selective and promiscuous compounds. *Journal of Chemical Information and Modeling*, *52*, 2454–2461.
- Younge, A. J., Walters, J. P., Crago, S. P., Fox, G. C., Younge, A. J., Walters, J. P., ... Fox, G. C. (2015). Supporting high performance molecular dynamics in virtualized clusters using IOMMU, SR-IOV, and GPUDirect. In: *Vol. 50. Proceedings of the 11th ACM SIGPLAN/SIGOPS international conference on virtual execution environments—VEE '15*, (pp. 31–38). New York, NY: ACM Press.
- Yuriev, E., Holien, J., & Ramsland, P. A. (2015). Improvements, trends, and new ideas in molecular docking: 2012–2013 in review. *Journal of Molecular Recognition*, *28*, 581–604.
- Zakrocka, I., Targowska-Duda, K. M., Wnorowski, A., Kocki, T., Józwiak, K., & Turski, W. A. (2017). Angiotensin II type 1 receptor blockers inhibit KAT II activity in the brain—its possible clinical applications. *Neurotoxicity Research*, *32*, 639–648.
- Zhang, L., Tan, J., Han, D., & Zhu, H. (2017). From machine learning to deep learning: progress in machine intelligence for rational drug discovery. *Drug Discovery Today*, *22*, 1680–1685.

Further Reading

- Gonczarek, A., Tomczak, J. M., Zaręba, S., Kaczmar, J., Dąbrowski, P., & Walczak, M. J. (2018). Interaction prediction in structure-based virtual screening using deep learning. *Computers in Biology and Medicine*, *100*, 253–258.

Article

Expanding the Structural Diversity of DNA Methyltransferase Inhibitors

K. Eurídice Juárez-Mercado ¹, Fernando D. Prieto-Martínez ¹, Norberto Sánchez-Cruz ¹,
Andrea Peña-Castillo ¹, Diego Prada-Gracia ² and José L. Medina-Franco ^{1,*}

¹ DIFACQUIM Research Group, Department of Pharmacy, School of Chemistry, National Autonomous University of Mexico, Avenida Universidad 3000, Mexico City 04510, Mexico; kaeuridice@gmail.com (K.E.J.-M.); ferdpm4@hotmail.com (F.D.P.-M.); norberto.sc90@gmail.com (N.S.-C.); andrea.pecas93@gmail.com (A.P.-C.)

² Research Unit on Computational Biology and Drug Design, Children's Hospital of Mexico Federico Gomez, Mexico City 06720, Mexico; prada.gracia@gmail.com

* Correspondence: medinajl@unam.mx

Abstract: Inhibitors of DNA methyltransferases (DNMTs) are attractive compounds for epigenetic drug discovery. They are also chemical tools to understand the biochemistry of epigenetic processes. Herein, we report five distinct inhibitors of DNMT1 characterized in enzymatic inhibition assays that did not show activity with DNMT3B. It was concluded that the dietary component theaflavin is an inhibitor of DNMT1. Two additional novel inhibitors of DNMT1 are the approved drugs glyburide and panobinostat. The DNMT1 enzymatic inhibitory activity of panobinostat, a known pan inhibitor of histone deacetylases, agrees with experimental reports of its ability to reduce DNMT1 activity in liver cancer cell lines. Molecular docking of the active compounds with DNMT1, and re-scoring with the recently developed extended connectivity interaction features approach, led to an excellent agreement between the experimental IC₅₀ values and docking scores.

Keywords: dietary component; epigenetics; enzyme inhibition; focused library; epi-informatics; multitarget epigenetic agent; natural products; chemoinformatics



check for updates

Citation: Juárez-Mercado, K.E.; Prieto-Martínez, F.D.; Sánchez-Cruz, N.; Peña-Castillo, A.; Prada-Gracia, D.; Medina-Franco, J.L. Expanding the Structural Diversity of DNA Methyltransferase Inhibitors. *Pharmaceuticals* **2021**, *14*, 17. <http://dx.doi.org/10.3390/ph14010017>

Received: 20 November 2020
Accepted: 23 December 2020
Published: 27 December 2020

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Historically, the term “epigenetics” is rooted in Waddington and Nanney’s work, where it was initially defined to denote a cellular memory, persistent homeostasis in the absence of an original perturbation, or an effect on cell fate not attributable to changes in DNA [1,2]. However, “epigenetics” is now used with multiple meanings, for instance, to describe the heritable phenotype (cellular memory) without modification of DNA sequences [3], or the mechanism in which the environment conveys its influence to the cell, tissue, or organism [4]. Regardless of the different definitions, the interest in epigenetic drug discovery has increased, as revealed by the multiple approved epigenetic drugs or compounds in clinical development for epigenetic targets [5,6].

DNA methyltransferases (DNMTs) are one of the primary epigenetic modifiers. This enzyme family is responsible for promoting the covalent addition of a methyl group from *S*-adenosyl-*L*-methionine (SAM) to the 5-carbon of cytosine, mainly within CpG dinucleotides, yielding *S*-adenosyl-*L*-homocysteine (SAH) [7]. DNMT1, DNMT3A, and DNMT3B participate in DNA methylation in mammals to regulate embryo development, cell differentiation, gene transcription, and other normal biological functions. Abnormal functions of DNMTs are associated with tumorigenesis and other diseases [7,8].

DNMTs were the first epigenetic targets for which inhibitors received the approval of the Food and Drug Administration (FDA) of the USA: the nucleoside analogs 5-azacitidine (Vidaza) and decitabine or 5-aza-2'-deoxycytidine (Dacogen) (Figure 1), approved in 2004 and 2006, respectively, for the treatment of the myelodysplastic syndrome [9]. DNMTs

are promising epigenetic targets for the treatment of several types of cancer, including acute myeloid leukemia and colorectal, pancreatic, lung, ovarian, and breast cancer, which have been reviewed comprehensively [8,10]. Furthermore, DNMTs are also attractive targets for the investigation or treatment of other diseases such as diabetes [11], as well as autoimmune [12] and neurological disorders [13]. Inhibitors of DNMTs are also emerging as programs to develop combination therapies in drug cocktails or compounds targeting more than one epigenetic target simultaneously [14]. For instance, Rabal et al. recently reported dual inhibitors of DNMT1 and G9a histone methyltransferase [15,16]. Yuan et al. described a dual DNMT and histone deacetylase inhibitor (HDAC) inhibitor [17]. Hydralazine is an antihypertensive drug and a weak inhibitor of DNMT1 (Figure 1) [18]. This compound has been proved in several cancer cell lines, and there are studies that demonstrated that it is an inhibitor of DNMT1 with an IC_{50} of 2 μ M for A549 cell line (human lung cancer cell line harboring wild-type p53), an IC_{50} of 20 μ M for U373MG cell line (human glioblastoma cell line harboring inactive mutant p53) [19], and an IC_{50} of 30 μ M for Hut78 cell line (cutaneous T cell lymphoma) [20,21].

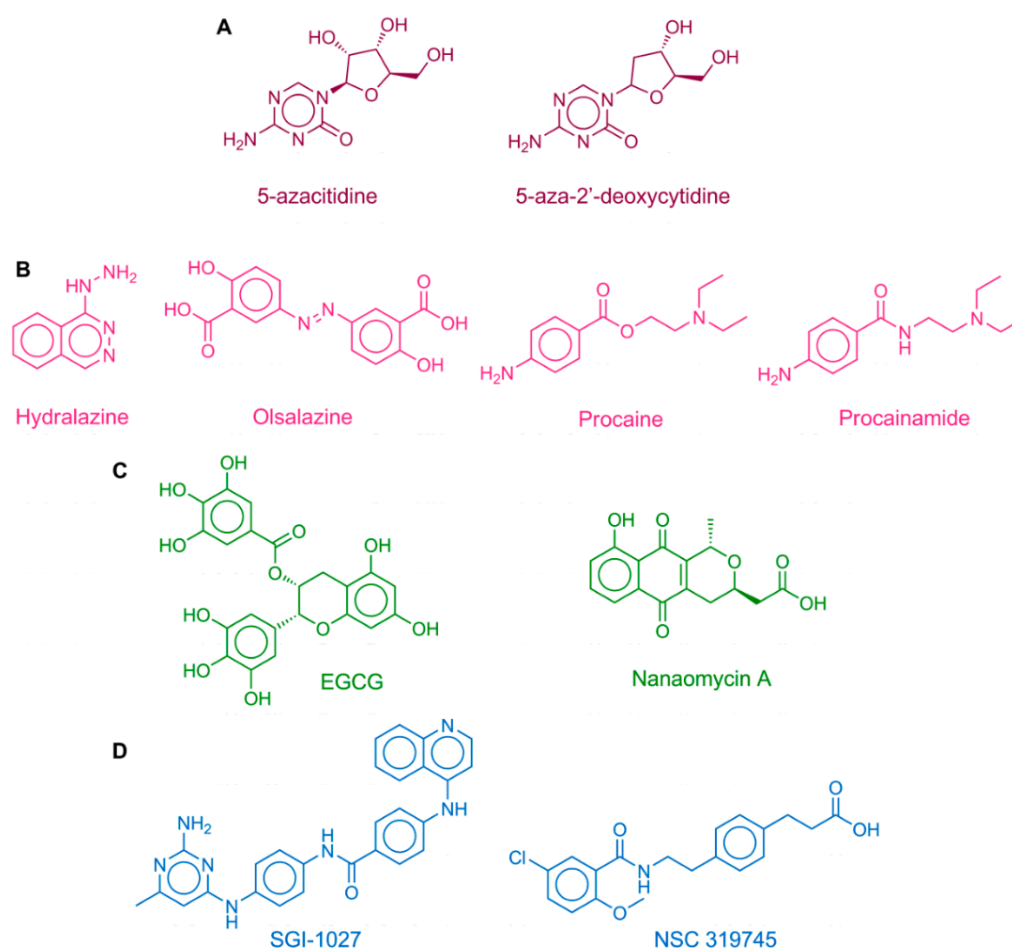


Figure 1. Chemical structures of representative DNA methyltransferase (DNMT) inhibitors and other compounds with associated hypomethylating properties from different sources. (A) approved drugs; (B) drug repurposing; (C) natural products and dietary components; (D) small molecules: synthetic compounds.

Despite the fact there are two DNMT inhibitors approved for clinical use, both azacitidine and decitabine have low specificity, poor bioavailability, and instability in physiological conditions and toxicity. Therefore, it has been the interest of our [22–29] and several other research groups [30–37] to identify DNMT inhibitors with novel chemical scaffolds for further development. Inhibition of DNMTs remains a major topic of research not only because of its potential therapeutic benefits but also to understand the essential mechanisms of epigenetic events in cells. There are currently more than 256 compounds in annotated public chemical databases [38] with measured activity vs. DNMTs. Figure S1 in the Supplementary Materials shows the most frequent scaffolds of active molecules. Figure 1 shows the chemical structures of representative DNMT inhibitors or compounds with DNA demethylation activity from different sources, including drugs for other indications, screening compounds from synthetic origin, and natural products [39–41]. Moreover, there are several compounds from dietary origin [42–44]. Of note, strong evidence indicates that environmental factors and nutrients play a major role in establishing epigenetic mechanisms, including irregular DNA methylation patterns. Thus, a regular uptake of DNA demethylating agents (which are not necessarily very potent DNMT inhibitors) is hypothesized to have a chemopreventive effect [45].

The DNMT inhibitors have been identified from different approaches or their combination [45,46] such as virtual screening [22,31], high-throughput screening, lead optimization [35,47], and structure-guided design [37], to name a few. Amongst the most promising inhibitors are molecules with “long scaffolds” such as the quinolone-based **SGI-1027** (Figure 1) and analogs. With the aid of molecular modeling, it has been hypothesized that such compounds with long scaffolds occupy the catalytic site and SAM’s cavity [48]. It has also been proposed that analogs of **SGI-1027** exert their mechanism through interaction with DNA [49].

As part of an ongoing effort to identify novel DNMT1 inhibitors from different sources and further increase the availability of novel scaffolds, herein, we report five new inhibitors of DNMT1 with distinct chemical scaffolds. Two compounds, glyburide and Panobinostat, are approved drugs with potential drug-repurposing applications. In particular, panobinostat is a pan-histone deacetylase inhibitor (HDAC), another major epigenetic target, and could be used as a dual epigenetic agent.

2. Results

2.1. Biochemical DNMT Assays

Table 1 summarizes the relative enzymatic activities of DNMT1 and DNMT3B in the presence of 100 μ M compound. Compounds that had more than 20% inhibition were regarded as inhibitors and were moved forward to dose–response evaluations. We used a similar criterion in a previous identification of novel chemical scaffolds [22]. Seven molecules showed detectable inhibition with DNMT1, of which the five most active were theaflavin (65% inhibition), **CSC027694519** (65%), panobinostat (63%), **7936171** (62%), and glyburide (60%). The least active were **CSC027480404** (29%) and **CSC026286840** (27% inhibition). Theaflavin and **7936171** showed detectable inhibition with DNMT3B (33 and 22% inhibition, respectively). All other eight molecules were inactive with DNMT3B.

The five compounds with the highest percentage of inhibition at 100 μ M with DNMT1 were tested in a dose–response manner. Theaflavin that showed the best activity at a single dose with DNMT1 and DNMT3B was also tested in a dose–response manner with DNMT3B and DNMT3B/3L. Table 2 summarizes the results.

Table 1. Results of the relative enzymatic activity of DNMT1 and DNMT3B as percentages ^a.

Set	Compound	DNMT1	DNMT3B
Approved drug	Glyburide	40.04 (±0.78)	95.97 (±4.76)
Approved drug	Panobinostat	37.31 (±1.80)	103.05 (±0.59)
Dietary component	Theaflavin	34.62 (±0.06)	66.75 (±1.11)
Inhibitor of the viral NS5 RNA methyltransferase	7936171	37.74 (±2.15)	78.20 (±0.30)
DNMT-focused library	CSC027480404	70.63 (±0.19)	96.39 (±0.28)
DNMT-focused library	CSC026286840	73.45 (±3.06)	90.59 (±2.49)
DNMT-focused library	CSC027694519	35.42 (±1.78)	97.42 (±2.49)
DNMT-focused library	6631802	93.26 (±7.97)	93.54 (±0.54)
DNMT-focused library	CSC027796832	107.25 (±0.59)	95.28 (±0.47)
DNMT-focused library	CSC027083851	105.71 (±1.27)	97.85 (±5.66)

^a Mean value of two measurements ± standard deviation. Compounds with an inhibition greater than 20% are marked in blue.

Table 2. Results of dose–response evaluations for selected compounds (IC₅₀) with DNMT1 ^a.

Set	Compound	DNMT1 (IC ₅₀ μM)
Approved drug	Glyburide	55.85 (±1.11)
Approved drug	Panobinostat	76.78 (±0.23)
Dietary component	Theaflavin ^b	85.33 (±0.14)
Inhibitor of the viral NS5 RNA methyltransferase	7936171	78.53 (±8.60)
DNMT-focused library	CSC027694519	85.11 (±4.10)

^a Mean value of two measurements. SAH was included as a positive control (IC₅₀ of 0.26 μM); ^b also evaluated with DNMT3B and DNMT3B/3L. The IC₅₀ was > 100 μM.

The two approved drugs, glyburide and panobinostat, inhibited DNMT1 with IC₅₀ values of 55.85 and 76.78 μM, respectively. Theaflavin had an IC₅₀ value of 85.33 μM. The other two small molecules **793617** and **CSC027694519** had IC₅₀ values of 78.53 and 85.1 μM, respectively. SAH was used as a non-specific positive control and confirmed its effective inhibition of DNMT1 with an IC₅₀ of 0.26 μM under the assay's conditions. Theaflavin, which showed the best activity at a single dose with DNMT1 and DNMT3B (Table 1), was also tested in a dose–response manner with DNMT3B and DNMT3B/3L showing, in both cases, IC₅₀ values > 100 μM.

2.2. Molecular Docking and Re-Scoring

The docking was performed in Molecular Operating Environment (MOE) v.2018 for the 10 compounds (shown in Figure 2) in the active site of the crystal structure of the catalytic domain of DNMT1 (PDB ID: 4WXX), as described in the Materials and Methods section. The docking scores ranged from −8.35 to −7.09 kcal/mol. The scores for the five compounds tested in the biochemical assays in dose–response evaluations ranged from −8.20 to −7.42 kcal/mol. Table 3 summarizes the results of docking with MOE and re-scoring with extended connectivity interaction features (ECIF) (see the Materials and Methods section).

Figure 3 shows the 2D interaction maps of the predicted binding mode between the human catalytic domain of DNMT1 and the five compounds evaluated in a dose–response manner: glyburide, panobinostat, theaflavin, **793617**, and **CSC027694519**. These interactions correspond to the docking poses with the most favorable docking scores, as calculated with MOE. All the compounds had predicted interactions with catalytic residues and showed hydrogen bond interactions with Glu 1266. Furthermore, glyburide, panobinostat, theaflavin, and **CSC027694519** also showed π–H interactions.

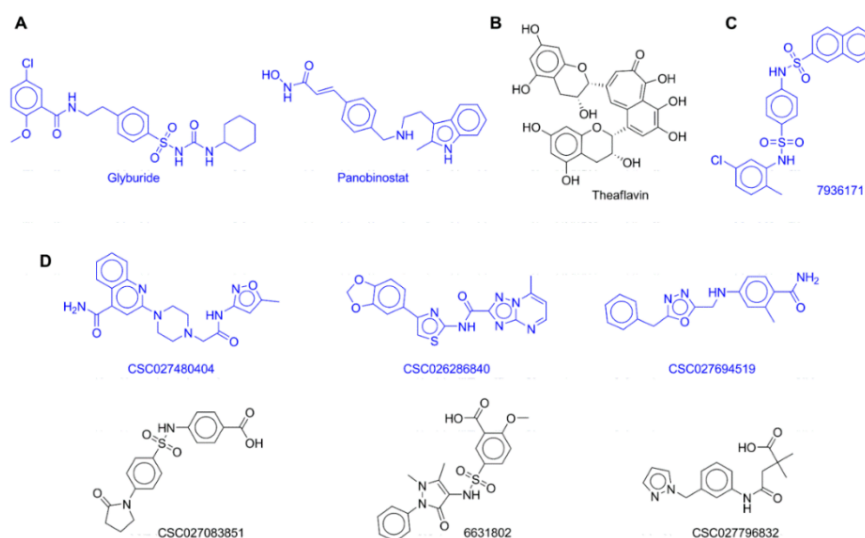


Figure 2. Chemical structures of 10 compounds from different sources experimentally tested. Compounds are grouped by their source: (A) approved drugs; (B) dietary source (natural product); (C) inhibitor of the viral NS5 RNA methyltransferase; (D) DNMT-focused library.

Table 3. Summary results of docking with Molecular Operating Environment (MOE) and re-scoring with extended connectivity interaction features (ECIF).

ID	ECIF Score ^a	MOE Score	pIC ₅₀	% Inhibition
Glyburide	7.38	−8.20	4.25	59.96
Panobinostat	6.36	−7.34	4.11	62.69
Theaflavin	6.48	−8.16	4.07	65.38
7936171	6.52	−7.67	4.11	62.26
CSC027694519	5.85	−7.71	4.07	64.58
6631802	7.26	−8.35	ND ^b	6.74
CSC027796832	5.50	−7.09	ND	−7.25
CSC027480404	6.62	−8.15	ND	29.97
CSC026286840	6.69	−8.28	ND	26.55
CSC027083851	6.81	−7.42	ND	−5.71
SAH	5.25	−9.48	6.59	ND

^a Predictions from ECIF6::LD-GBT; ^b ND: not determined.

The docking poses of the five compounds were re-scored with ECIF, as described in the Materials and Methods section. Table 3 summarizes the results of the predictions from ECIF6::LD-GBT for the docking poses generated with MOE. For the five compounds tested in a dose–response manner, the table shows the results of the IC₅₀ as the $-\log$ value (pIC₅₀). The table indicates that among the five compounds with pIC₅₀ values, the most active, glyburide, was predicted correctly by the ECIF re-scoring method. Moreover, the re-scoring of MOE poses ranked glyburide as the one with the highest affinity overall, while MOE score predicted a higher affinity for compounds 6631802 and CSC026286840. Figure 4 shows the association between the experimental pIC₅₀ of the five compounds evaluated in biochemical inhibition assays of DNMT1 with the MOE's docking scores (Figure 4A) and re-scoring with ECIF6::LD-GBT (Figure 4B). Clearly, the ECIF re-scoring scheme improved the correlation with the experimental pIC₅₀ values.

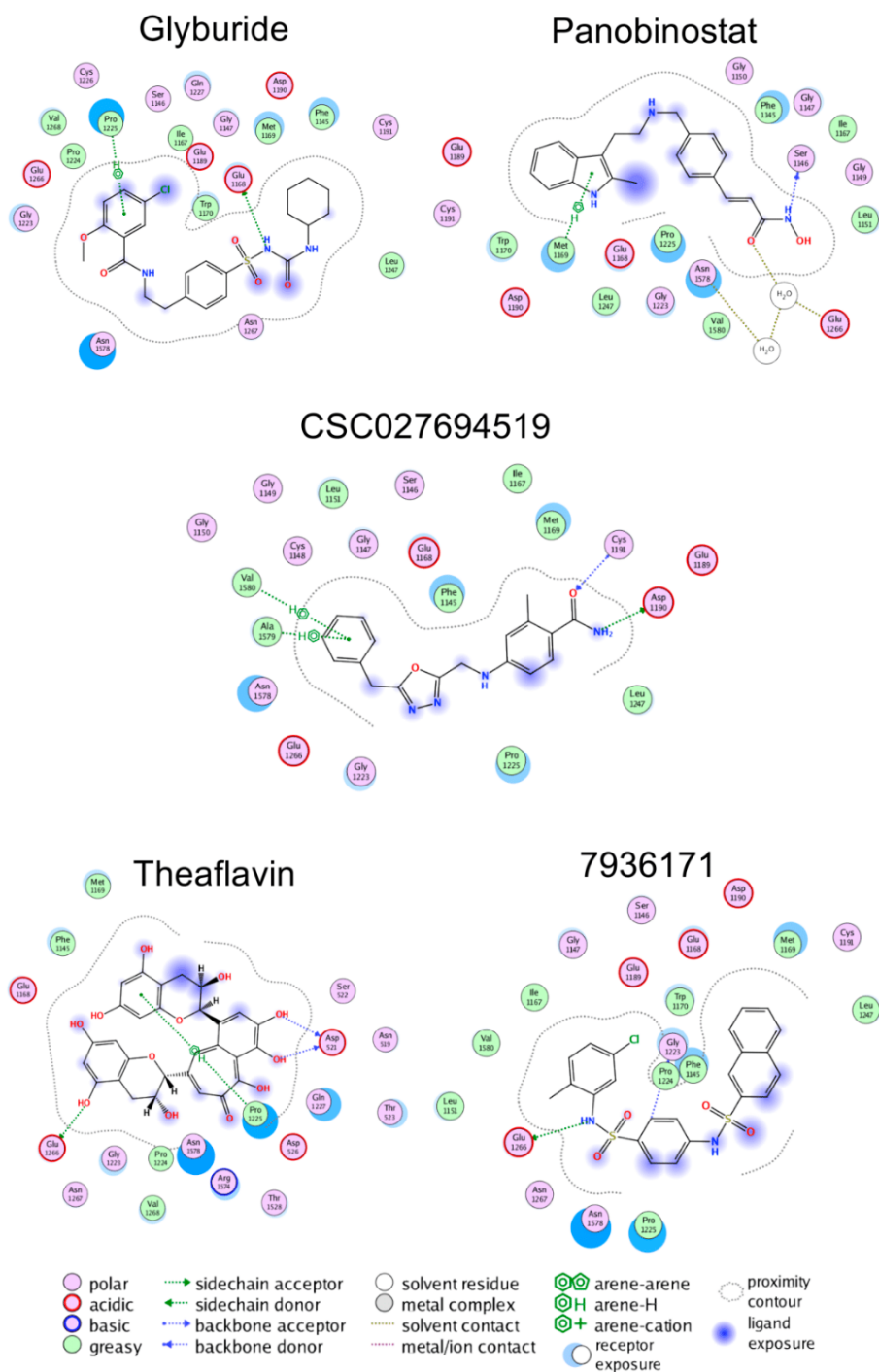


Figure 3. Binding poses predicted with the Molecular with Molecular Operating Environment v.2018 of the five active compounds with the catalytic domain of DNA methyltransferase 1.

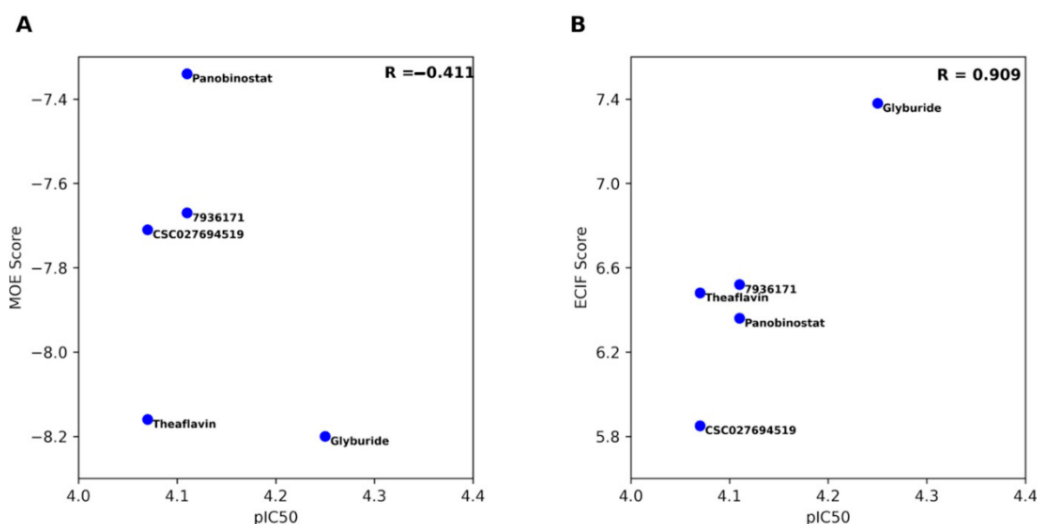


Figure 4. Correlation between the experimental pIC₅₀ of the five compounds evaluated in biochemical inhibition assays of DNMT1 with (A) the docking scores calculated with MOE, and (B) the ECIF6::LD-GBT scores.

2.3. Molecular Dynamics and Adaptive Sampling

Molecular dynamics (MD) simulations were conducted to determine the putative binding mode and interactions of glyburide and DNMT1. To this end, six docking poses (Figure S3 in the Supplementary Materials) were selected as the starting points of the MD simulations in question. With these, kinetic transition networks (KTNs) were constructed to determine microstates (on the basis of protein–ligand contacts obtained with YASARA) and their associated probability. Only the trajectory from “pose 6” showed a microstate network (Figure S4 in the Supplementary Materials) with higher probability than other poses (data not shown). To confirm this observation, we conducted “random walks” using implicit solvation, as this often favors faster transitions [50].

Simulation data were postprocessed with PyEMMA, using different features for microstate definition and construction of preliminary Markov state models (MSMs). At this stage, only inverse distances between the carbonyl atoms of glyburide and DNMT1 (C α atoms) showed metastability by means of time-lagged independent component analysis (TICA) and k-means clustering (Figure 5A). Using spectral analysis along eigenvalues with the Robust Perron Cluster Analysis (PCCA+) [51], we determined microstate density and distribution (Figure 5B); finally, contact frequency along these samples was calculated for comparison (Figure 5C).

Microstate distributions showed that glyburide favored hydrophobic contacts within the SAH binding site, with transitions between hydrogen bonding and hydrophobic contacts for Phe1145, Trp1170, and Asn1578. On the basis of these observations, we decided to conduct end-point calculations by means of molecular mechanics/Poisson-Boltzmann Surface Area (MM/PBSA) using pose 6 trajectory (in explicit solvent), yielding a ΔG value of -48.7 ± 4.6 kcal/mol.

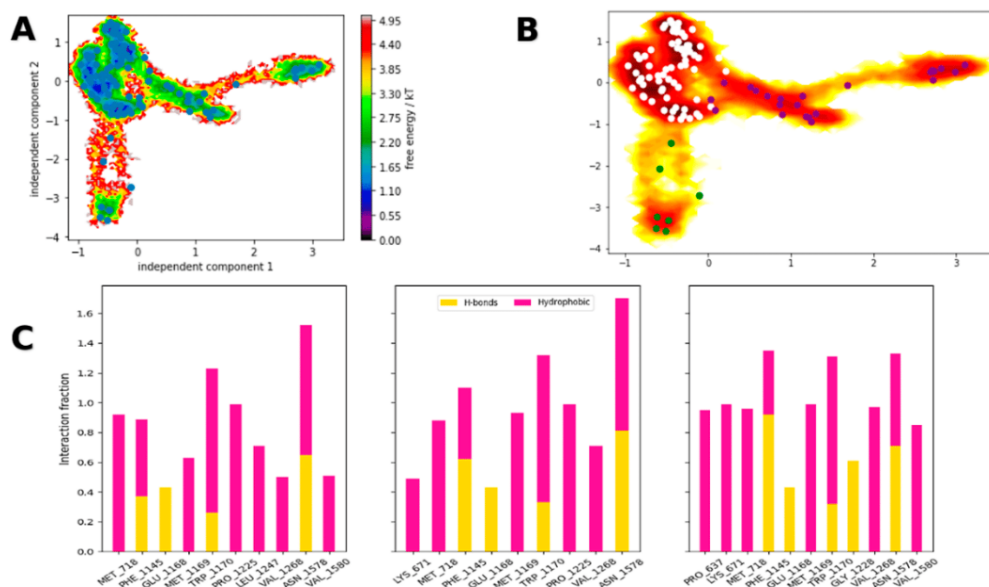


Figure 5. Free energy landscape of the glyburide–DNMT1 complex along the slowest independent components (A); microstate distribution as obtained with the Perron Cluster Analysis PCCA+ sampling (B); contact frequency obtained from the discretized microstates (C).

3. Discussion

3.1. Biochemical DNMT Assays

The initial goal of this study was to identify novel small molecule DNMT1 and DNMT3B inhibitors. In this work, we deprioritized testing with DNMT3A because it has been reported that this enzyme can act as both an oncogene and as a tumor suppressor gene, at least in lung cancer (it remains to be tested as to whether this paradox of DNMT3A applies to other cancers) [8]. Seven molecules with detectable inhibition of DNMT1 and two with DNMT3B were initially identified at single dose concentrations (Table 1). An approved drug for the treatment of type 2 diabetes (glyburide) and an approved anticancer drug known as HDAC inhibitor (panobinostat) were among the active compounds with DNMT1. Other active compounds against DNMT1 were a dietary component (theaflavin), a naphthalene sulfonamide (7936171) previously reported to be an inhibitor of dengue virus methyltransferase [52], and three small molecules from a DNMT focused library (CSC027480404, CSC026286840, and CSC027694519).

Overall, the tested compounds showed less activity with DNMT3B (Table 1). Only two compounds, the dietary component theaflavin, and the naphthalene sulfonamide, showed inhibition higher than 20% with DNMT3B at high concentration. Moreover, theaflavin showed the highest percentage of inhibition (33%), which was lower than the highest percentage of inhibition showed for DNMT1 (65%). None of the two compounds had an IC_{50} lower than 100 μ M with DNMT3B.

Dose–response assays for the most active compounds at a single dose (Table 2) revealed that, under the assay conditions used in this study, the two approved drugs glyburide and panobinostat had the lowest IC_{50} values (55.85 and 76.78 μ M, respectively). The other three active compounds (7936171, theaflavin, and CSC027694519) had IC_{50} values between 78.5 and 85 μ M. Overall, these IC_{50} values were relatively high. However, it should be noted that standard biochemical assays for DNMT enzymes have not been fully established and the results show a large variation between assay conditions. This point has been largely discussed in the literature for DNMT inhibitors [48,53]. In this work, we used SAH as a

positive control, since SAH is a very well known inhibitor of DNMT [54]. Even though the active compounds' potency values are not high, all compounds have chemical scaffolds different from the scaffolds reported for known DNMT inhibitors.

Glyburide (Figure 2) is a sulfonylurea. In 1984, it received approval by the United States FDA for the treatment of patients with diabetes mellitus type II. To the best of our knowledge, there are no reports of this compound as an inhibitor of DNMT1. However, several studies support that sulfonylureas are potential anticancer drug candidates for several cell lines such as colon cancer, human ovarian cancer, kidney cancer, melanoma, and lung cancer [55,56]. Therefore, we propose that glyburide could be further investigated and developed as an epi-drug with potential anticancer activity.

Panobinostat is an orally available pan-deacetylase inhibitor with broad antitumor activity [57]. It has shown inhibitory activity vs. different types of cancer, for example, malignant glioma cells with an IC_{50} range of 0.3 μ M to 0.23 μ M [58], carcinoma cells (IC_{50} between 0.4 nM and 1.3 nM) [59], and in human erythroleukemic cell lines (IC_{50} between 40 nM and 51 nM) [60]. One of the principal benefits of the panobinostat is that being an HDAC inhibitor will avoid the histone acetylation, which helps to maintain that the chromatin is closed and decrease the gene transcription involved with proliferation, differentiation, and progress of the cancer cells. Zopf et al. reported that panobinostat reduced DNMT1 (and DNMT3a) activities and expression in liver cancer cell lines [61]. In the study of Zopf et al., the authors concluded that inhibitors of HDACs can indirectly control DNA methylation. In this work, it was confirmed that panobinostat also inhibits the enzymatic activity of DNMT1 directly. Of note, there is a recent interest in developing dual inhibitors of DNMTs and HDACs, and results in this direction appear promising [17]. Indeed, Yuan et al. reported hydroxamic acid derivatives of a small molecule previously identified as a weak inhibitor of DNMT1 from virtual screening of a large chemical library [22]. In the work of Yuan et al., the hydroxamic acid derivatives of NSC 319745 showed inhibition of DNMT1, HDAC1, and HDAC6, plus cytotoxicity activity against human cancer cells [47]. In a second example in the clinic, the combined use of decitabine (a DNMT inhibitor) and belinostat (an HDAC inhibitor) increased the chemotherapy efficiency [62]. Moreover, Min et al. showed that an increase in transcription of DNMT1 is one of the mechanisms of resistance of anti-cancer drugs targeting HDACs, such as vorinostat. Consequently, it has been shown that co-targeting DNMT1 improved the antitumor efficacy of vorinostat and other HDAC inhibitors [63]. These contributions, including the findings of this work, are in line with the epigenetic multitargeting. The advantages of this approach have been demonstrated in both in vitro and in vivo disease models by the co-administration of an epigenetic agent with another drug [64].

The naphthalene sulfonamide **7936171** (also with ID ZINC 01078518) is a known inhibitor of the viral NS5 RNA methyltransferase with an IC_{50} of 64.2 μ M and an EC_{50} of 12 μ M [52]. The viral RNA domain involves the use of SAM as the methyl donor and generates SAH as the final product. Compound **7936171** has a "long" and different scaffold from the known DNMT1 inhibitors published thus far (Figure S1).

Theaflavin is a natural product polyphenol found in green and black tea and coffee (vide supra) with previously measured enzymatic inhibitory activity of DNMT3A [65]. However, there were no reports on its inhibitory potential of DNMT1 and DNMT3B. In this work, we measured for the first time its activity with both enzymes. Despite the fact its IC_{50} is high (85.3 μ M, under the assay conditions of this work), this dietary component could contribute to the modulation of DNMT1. Interestingly, it has been proposed that the modulation of normal levels of DNMT could be conveniently achieved through the dietary uptake of food chemicals (or other "safe" natural products). A prominent example of this hypothesis has been suggested for the polyphenol compound from green tea, EGCG (Figure 1), which has been proposed to inhibit DNMT1 and reactivate methylation-silenced genes in cancer [45].

The other two small molecules, **7936171** and **CSC027694519**, had no previous reports of inhibition of DNMTs. As discussed in the next section, both molecules have drug-

like characteristics and have novel scaffolds that can be further optimized for DNMT1 inhibition.

3.2. Computational Studies

Herein, we employed *in silico* methods to rationalize at the molecular level the experimental results of the most active molecules. Of the different mechanisms described to inactivate DNMT activity [66], we hypothesized that the herein identified inhibitors are SAM competitors. This hypothesis is based on the “long” scaffolds of the active compounds such as glyburide, panobinostat, **7936171**, and **CSC027694519**. The predicted binding modes suggested strategies for the structure-based optimization of the small molecules. For instance, comparing the predicted binding mode of **CSC027694519** with the co-crystallized position of SAH in PDB ID: 4WXX (Figure S2 in the Supplementary Materials) suggests that substitution with a polar group in the benzyl ring of **CSC027694519** could enhance the affinity, making polar or hydrogen bond contacts with amino acid residues such as Ser1146, Gly1150, Leu1151, and Val1580 (by comparison with the primary amine or carboxylate groups of SAH). Of note, **CSC027694519** already makes hydrogen bond interactions with the polar side chains of Asp1190 and Cys1191 (similar to the hydrogen bond contacts of SAH, Figure S2). Similarly, comparing the predicted binding mode of **7936171** with the co-crystallized position of SAH (Figure S2) suggests that the naphthalene ring of **7936171** could be replaced with polar heteroaromatic rings. Of course, the synthesis and testing of the analogs should be made, which is one of the major perspectives of this work. It remains to confirm the hypothesis of the putative binding site of these molecules that can be further tested experimentally with binding completion assays once more potent compounds are identified.

To gain some rationale behind the observed pIC_{50} of glyburide, we conducted molecular simulations to probe its putative binding mode in DNMT1. Using KTN models, we determined that **pose 6** had a higher contact stability. This pose had the urea moiety as the main anchor towards Glu1168, the main contact for charged amines, as described in previous works [67]. However, by means of adaptive sampling and Markov state models, we determined that three microstate distributions are possible for this binding mode. In these, a high prevalence of hydrophobic contacts with Met718, Trp1170, and Val1268 was observed. Moreover, a mixture of polar and hydrophobic interactions towards Phe1145, Trp1170, and Asn1578 was determined, with the latter showing the most concurrent profile along microstates, suggesting it as the presumptive anchor for glyburide. Hence, optimization around this scaffold could provide venues for ligand design without the need of charged groups.

On the basis of the Chapman–Kolmogorov test [68], researchers can only apply this preliminary model to small timescales. Thus, a more extensive validation is necessary to achieve a robust kinetic description of the possible binding/unbinding paths of glyburide; nonetheless, this goes beyond the current scope of this work. Furthermore, on the basis of ΔG values obtained from end-point calculations, we observed a better correlation with the experimental value, considering that similar predictions have been observed on other epi-modulators with IC_{50} values in the micromolar range [69]. This gives further support to the proposed binding mode and the experimental potency of glyburide.

4. Materials and Methods

We aimed to test compounds from different sources: approved drugs, synthetic compounds from a DNMT-focused library, and 1 compound from dietary origin. We experimentally tested 10 molecules with diverse chemical scaffolds and chemical structures different from reported DNMT inhibitors (Figure 2). Thus, we screened the following 2 approved drugs with “long” chemical scaffolds in biochemical assays as inhibitors of DNMTs (Figure 2A): the sulfonyleurea glybenclamide, approved for the treatment of diabetes (*vide infra*), and panobinostat, a non-selective and potent zinc-dependent HDAC inhibitor (the most potent deacetylase inhibitor on the market), approved in 2018 by the FDA for the treatment

of multiple myeloma. Of note, currently epigenetic multitargeting is focused on the inhibition of zinc-dependent HDACs as one of the action mechanisms [14]. Due to the known activity of polyphenols and previous evidence of theaflavin activity with DNMT3A [65], we hypothesized that theaflavin, a dietary component (Figure 2B) present in black tea, is an inhibitor of DNMT1 and DNMT3B. We also theorized that the naphthalene sulfonamide **7936171** (Figure 2C) with a long scaffold is an inhibitor of DNMT1 and DNMT3B [52]. Finally, we tested six molecules from a DNMT-focused library of synthetic molecules (Figure 2D), which are becoming attractive to experimentally screen molecules [6]. To select the 10 compounds in this work, we used commercial availability and reasonable price (the latter an important factor considering the current economic situation imposed by the coronavirus disease 2019 (COVID-19) pandemic) as additional criteria.

First, the 10 compounds were tested at a single concentration (100 μ M) in duplicate in a biochemical DNMT assay. SAH was used as a reference. The five most active molecules at one-single concentration were further tested in a dose–response manner to obtain the half-maximal inhibitory concentration (IC₅₀). Predicted binding modes for the five compounds were studied using molecular docking, implementing a novel re-scoring algorithm recently developed [70].

4.1. Compounds

All compounds were purchased from chemical vendors. Glyburide, panobinostat, and theaflavin were purchased from TargetMol. Molecules **7936171** and **6631802** were purchased from Chembridge Corporation. Molecules **CSC027480404**, **CSC026286840**, **CSC027694519**, and **CSC027083851** were acquired from ChemSpace. The compound purity confirmed by the chemical vendors was equal or higher than 90% (Table S1 in the Supplementary Materials).

4.2. Biochemical DNMT Assays

The inhibition of the enzymatic activity of DNMT1, DNMT3B, and DNMT3B/3L was tested using the HotSpotSM platform for methyltransferase assays available at Reaction Biology Corporation [71]. HotSpotSM is a low volume radioisotope-based assay that uses tritium-labeled AdoMet (³H-SAM) as a methyl donor. The test compounds diluted in dimethyl sulfoxide were added using acoustic technology (Echo550, Labcyte) into enzyme/substrate mixture in the nano-liter range. The corresponding reactions were commenced by the addition of ³H-SAM and incubated at 30°C. Total final methylations on the substrate (Poly dI-dC in DNMT1 assay, and Lambda DNA in DNMT3B; DNMT3B/3L assay) were identified by a filter binding method implemented in Reaction Biology. Data analysis was conducted with Graphed Prism software (La Jolla, CA, USA) for curve fits. Reactions were carried out at 1 μ M of SAM. In all assays, SAH was used as a standard positive control. The 10 compounds were tested first with DNMT1 and DNMT3B at one 100 μ M concentration in duplicate. The 5 most active compounds were tested in 10-dose IC₅₀ (effective concentration to inhibit DNMT1 activity by 50%) with a threefold serial dilution starting at 100 μ M. Theaflavin was also tested with DNMT3B and DNMT3B/L in 10-dose IC₅₀ with threefold serial dilution starting at 100 μ M. The authors have previously contracted the screening services of Reaction Biology Corporation to identify a novel inhibitor of DNMT1 [24].

4.3. Molecular Docking and Re-Scoring

The 5 compounds were tested in a dose–response manner, and that showed activity with DNMT1 (vide infra) were docked with this enzyme using the program Molecular Operating Environment (MOE), version 2018.08 [72]. The chemical structures of the 5 compounds were built with MOE. The docking was carried out with the crystal structure of the catalytic domain of DNMT1 obtained from the Protein Data Bank [73] (PDB ID: 4WXX [74]). This crystal structure is in complex with SAH and has a resolution of 2.62 Å. The structure of the protein was prepared with the “QuickPrep” tool of MOE v.2018 using

the parameters established by default, which helped us to remove the molecules of structural water and to add hydrogens atoms to the protein. In this process, the co-crystallized SAH was removed for the binding site to realize a direct docking. The docking was conducted using default parameters in MOE. Before docking the 5 newly tested compounds, we validated the docking protocol by re-docking the SAH obtaining a root mean square deviation of 1.3 angstroms (and a docking score of -8.96 kcal/mol).

Docking poses obtained with MOE were re-scored with the extended connectivity interactions features (ECIF) method. Briefly, ECIF is a recently reported set of descriptors to represent protein–ligand complexes. These descriptors are defined as a set of protein–ligand atom-type pair counts relying on a detailed description of the connectivity of the atoms involved. It has been shown that machine-learning scoring functions built on ECIF and gradient boosting trees consistently outperform the performance of scoring functions reported thus far regarding the obtention of binding scores in a linear correlation with experimental data, particularly when a distance cutoff criterion of 6 angstroms is used to derive the descriptors, and purely ligand-based descriptors are added (ECIF6::LD-GBT) [70]. From the poses obtained from MOE, the best-ranked one for each compound was prepared using X-Tool [75] and standardized using Standardizer, JChem 20.11.0, 2020, ChemAxon to perceive aromaticity in an interpretable way for the model. All poses were re-scored using the ECIF6::LD-GBT model [70].

4.4. Probing the Putative Binding Mode of Glyburide

To gain further insight on the presumptive interaction between glyburide and DNMT1, we conducted molecular dynamics (MD) simulations, mostly due to the observed pIC_{50} . To this end, the PDB ID: 3SWR was prepared with PDBfixer (v.1.6), completing missing atoms and residues. This structure was further optimized with YASARA2 forcefield (YASARA; v.20.8.1; [76]). Docking of glyburide was carried out using PLANTS (v.1.2; [77]) with a definition of the binding site in a sphere of 15 Å around the co-crystallized molecule of sinefungin. Twenty poses were obtained and analyzed choosing those with a similar arrangement to sinefungin or in its vicinity. The change of software at this stage was due to PLANTS performance on several benchmarks for binding mode prediction [78]. Protein–ligand complexes were parametrized with AMBERTools19, using the AMBER14 forcefield [79] for protein and GAFF2 for glyburide; the metal centers of DNMT1 were parametrized using the Cationic Dummy Atom model (CaDA; [80,81]). Initial structures were then buffered in an orthorhombic cell extended 1.4 nm beyond the protein in every direction, and all systems were simulated with explicit solvent using the TIP3P model. The systems were simulated using OpenMM (v.7.4.1; [82]), beginning with an energy minimization using steepest descent algorithm, with a convergence threshold of 10 kJ/mol/nm. This was followed by an equilibration of 1 ns, where all water molecules and hydrogen bonds were constrained; simulation timestep was set to 2 fs using the Langevin integrator. Selected poses were then simulated under NPT ensemble for 100 ns at 300 K and 1 atm of pressure using the Andersen thermostat and Monte Carlo Barostat. Trajectories were analyzed with MDtraj [83] and YASARA for protein–ligand contacts. Additionally, contact networks were constructed with Networkx [84] and analyzed using OpenKTN (in house library).

Adaptive Sampling and End-Point Calculations

With the aid of kinetic transition networks (KTN; [85]), we identified the most probable microstates along MD trajectories. We then followed this with a series of random walks starting from these conformations, using implicit solvent simulations. These were also carried out with OpenMM, using the Onufriev–Bashford–Case (OBC) model with GB^{OBCII} parameters [76]. The “walks” were also analyzed with MDtraj and OpenKTN to determine new starting conformations to improve sampling and transitions along microstates. These data were further analyzed with PyEMMA (v.2.5.7; [86]), accounting for a total simulation time of 1 μ s. We used trajectory featurization [87] to determine slow components and

metastability by means of time-lagged independent component analysis (TICA). With this information, preliminary Markov state models (MSMs) were constructed to gain insight on the kinetics of the system [88]. Finally, we used this information to conduct end-point calculations using the MM/PBSA method as implemented on APBS [89]).

5. Conclusions

DNMTs are a fundamental class of epigenetic regulatory enzymes. In this work, we tested 10 compounds with novel scaffolds as inhibitors of DNMT1 and DNMT3B in biochemical assays. Seven compounds showed at least 20% inhibition of DNMT1 at 100 μ M. Five molecules showed activity in dose-response inhibition assays with DNMT1 with IC_{50} between 55.8 and 85.3 μ M. Although the molecules' overall potency was not high under this work's assay conditions, four compounds were found to have novel chemical scaffolds not previously described as inhibitors of DNMT1 in biochemical assays. Notably, glyburide, an approved drug for diabetes type II treatment, showed the best potency of the compounds tested in this study and can be further investigated in its role in epigenetic mechanisms. Moreover, glyburide could be pursued for drug repurposing applications. It was also concluded that panobinostat is an inhibitor of the enzymatic activity of DNMT1. Remarkably, panobinostat, an epigenetic drug for treating different types of cancer, has shown inhibition of DNMT1 in liver cancer cell lines [61]. Therefore, it is concluded that panobinostat, a known HDAC inhibitor, can be further investigated as a multitarget epigenetic agent. In this work, we also concluded that theaflavin, a natural dietary product, inhibits DNMT1 and does not show significant inhibition of DNMT3B. Therefore, this work also complements theaflavin's DNMT activity profile that had reported activity with DNMT3A [65]. It was also concluded that compounds **7936171** and **CSC027694519** and the two compounds that showed 30% inhibition of DNMT1 at 100 μ M (**CSC027480404** and **CSC026286840**) can be starting points of optimization programs to improve their potency.

Molecular docking of the most active compounds helped propose binding modes with the catalytic binding site of DNMT1. Although MOE scores predicted SAH as the most potent inhibitor, the novel ECIF re-scoring scheme ECIF6::LD-GBT improved significantly the correlation of the docking scores calculated with MOE with the experimental pIC_{50} values considering the rest of the compounds. Therefore, docking with MOE for compound selection and re-scoring with ECIF for prioritizing experimental tests can guide the optimization programs of the DNMT1 inhibitors identified in this work, which is one of the main perspectives of this work. Another perspective is testing the DNA demethylation activity of the most active compounds. This can be achieved directly with **7936171** and **CSC027694519** or after the compounds are optimized.

Supplementary Materials: The following are available online at <https://www.mdpi.com/1424-8247/14/1/17/s1>, Figure S1: Chemical structures of the 10 most frequent (Bemis-Murcko) scaffolds of the active DNMT inhibitors available in ChEMBL 27. Table S1: Chemical vendors of the 10 compounds tested and their purity as supplied by the vendor. Figure S2: Comparison of the co-crystal position of SAH in the catalytic site of DNMT1 (PDB ID: 4WXX) with the predicted binding mode of (A) CSC027694519 and (B) 7936171. Figure S3: Docking poses of glyburide with DNMT1 obtained with the software PLANTS. Figure S4: Microstate network analysis of glyburide.

Author Contributions: Conceptualization, J.L.M.-F.; methodology, K.E.J.-M., F.D.P.-M., N.S.-C., A.P.-C. and D.P.-G.; formal analysis, K.E.J.-M., F.D.P.-M., N.S.-C., A.P.-C., D.P.-G. and J.L.M.-F.; resources, J.L.M.-F. and D.P.-G.; writing—original draft preparation, K.E.J.-M. and J.L.M.-F.; writing—review and editing, all authors; supervision, J.L.M.-F. and D.P.-G.; project administration, J.L.M.-F.; funding acquisition, J.L.M.-F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the *Consejo Nacional de Ciencia y Tecnología* (CONACyT), Mexico (grant 282785). K.E.J.-M. thanks the support from CONACyT. N.S.-C. and F.D.P.-M. are also thankful to CONACyT for the PhD scholarships no. 335997 and 660465/576637, respectively.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in the main text and the Supplementary Materials.

Acknowledgments: Valuable discussions with Oscar Méndez-Lucio, Jennifer Tamayo, Carlos Velazquez, and Miguel Herrera are highly appreciated.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Waddington, C.H. The epigenotype, endeavor, 1942, vol. 1 (pg. 18-20) reprinted in. *Int. J. Epidemiol.* **2012**, *41*, 10–13. [[CrossRef](#)] [[PubMed](#)]
2. Greally, J.M. A user's guide to the ambiguous word 'epigenetics'. *Nat. Rev. Mol. Cell Biol.* **2018**, *19*, 207–208. [[CrossRef](#)] [[PubMed](#)]
3. Wu, C.; Morris, J.R. Genes, genetics, and epigenetics: A correspondence. *Science* **2001**, *293*, 1103–1105. [[CrossRef](#)] [[PubMed](#)]
4. Cavalli, G.; Heard, E. Advances in epigenetics link genetics to the environment and disease. *Nature* **2019**, *571*, 489–499. [[CrossRef](#)]
5. Ganesan, A.; Arimondo, P.B.; Rots, M.G.; Jeronimo, C.; Berdasco, M. The timeline of epigenetic drug discovery: From reality to dreams. *Clin. Epigenet.* **2019**, *11*, 174. [[CrossRef](#)]
6. Sessions, Z.; Sánchez-Cruz, N.; Prieto-Martínez, F.D.; Alves, V.M.; Santos, H.P., Jr.; Muratov, E.; Tropsha, A.; Medina-Franco, J.L. Recent progress on cheminformatics approaches to epigenetic drug discovery. *Drug Discov. Today* **2020**, *25*, 2268–2276. [[CrossRef](#)]
7. Lyko, F. The DNA methyltransferase family: A versatile toolkit for epigenetic regulation. *Nat. Rev. Genet.* **2017**, *19*, 81. [[CrossRef](#)]
8. Zhang, J.; Yang, C.; Wu, C.; Cui, W.; Wang, L. DNA methyltransferases in cancer: Biology, paradox, aberrations, and targeted therapy. *Cancers* **2020**, *12*, 2123. [[CrossRef](#)]
9. Estey, E.H. Epigenetics in clinical practice: The examples of azacitidine and decitabine in myelodysplasia and acute myeloid leukemia. *Leukemia* **2013**, *27*, 1803–1812. [[CrossRef](#)]
10. Yu, J.; Xie, T.; Wang, Z.; Wang, X.; Zeng, S.; Kang, Y.; Hou, T. DNA methyltransferases: Emerging targets for the discovery of inhibitors as potent anticancer drugs. *Drug Discov. Today* **2019**, *24*, 2323–2331. [[CrossRef](#)]
11. Arguelles, A.O.; Meruvu, S.; Bowman, J.D.; Choudhury, M. Are epigenetic drugs for diabetes and obesity at our door step? *Drug Discovery Today* **2016**, *21*, 499–509. [[CrossRef](#)] [[PubMed](#)]
12. Teitell, M.; Richardson, B. Dna methylation in the immune system. *Clin. Immunol.* **2003**, *109*, 2–5. [[CrossRef](#)]
13. Weng, Y.-L.; An, R.; Shin, J.; Song, H.; Ming, G.-L. DNA modifications and neurological disorders. *Neurotherapeutics* **2013**, *10*, 556–567. [[CrossRef](#)] [[PubMed](#)]
14. de Lera, A.R.; Ganesan, A. Two-hit wonders: The expanding universe of multitargeting epigenetic agents. *Curr. Opin. Chem. Biol.* **2020**, *57*, 135–154. [[CrossRef](#)]
15. Rabal, O.; San José-Enériz, E.; Agirre, X.; Sánchez-Arias, J.A.; Vilas-Zornoza, A.; Ugarte, A.; de Miguel, I.; Miranda, E.; Garate, L.; Fraga, M.; et al. Discovery of reversible DNA methyltransferase and lysine methyltransferase g9a inhibitors with antitumoral in vivo efficacy. *J. Med. Chem.* **2018**, *61*, 6518–6545. [[CrossRef](#)]
16. López-López, E.; Rabal, O.; Oyarzabal, J.; Medina-Franco, J.L. Towards the understanding of the activity of g9a inhibitors: An activity landscape and molecular modeling approach. *J. Comp. Aided Mol. Des.* **2020**, *34*, 659–669. [[CrossRef](#)]
17. Yuan, Z.; Chen, S.; Gao, C.; Dai, Q.; Zhang, C.; Sun, Q.; Lin, J.-S.; Guo, C.; Chen, Y.; Jiang, Y. Development of a versatile DNMT and HDAC inhibitor C02S modulating multiple cancer hallmarks for breast cancer therapy. *Bioorg. Chem.* **2019**, *87*, 200–208. [[CrossRef](#)]
18. Arce, C.; Segura-Pacheco, B.; Perez-Cardenas, E.; Taja-Chayeb, L.; Candelaria, M.; Dueñas-Gonzalez, A. Hydralazine target: From blood vessels to the epigenome. *J. Transl. Med.* **2006**, *4*, 10. [[CrossRef](#)]
19. Kim, H.; Kim, J.; Chie, E.; DaYoung, P.; Kim, I.; Kim, I. DNMT (DNA methyltransferase) inhibitors radiosensitize human cancer cells by suppressing DNA repair activity. *Radiat. Oncol.* **2012**, *7*, 39. [[CrossRef](#)]
20. Schcolnik-Cabrera, A.; Domínguez-Gómez, G.; Dueñas-González, A. Comparison of DNA demethylating and histone deacetylase inhibitors hydralazine-valproate versus vorinostat-decitabine in cutaneous t-cell lymphoma in HUT78 cells. *Am. J. Blood Res.* **2018**, *8*, 5–16.
21. Singh, N.; Dueñas-González, A.; Lyko, F.; Medina-Franco, J.L. Molecular modeling and dynamics studies of hydralazine with human DNA methyltransferase 1. *ChemMedChem* **2009**, *4*, 792–799. [[CrossRef](#)] [[PubMed](#)]
22. Kuck, D.; Singh, N.; Lyko, F.; Medina-Franco, J.L. Novel and selective DNA methyltransferase inhibitors: Docking-based virtual screening and experimental evaluation. *Bioorg. Med. Chem.* **2010**, *18*, 822–829. [[CrossRef](#)] [[PubMed](#)]
23. Kuck, D.; Caulfield, T.; Lyko, F.; Medina-Franco, J.L. Nanaomycin a selectively inhibits DNMT3b and reactivates silenced tumor suppressor genes in human cancer cells. *Mol. Cancer Ther.* **2010**, *9*, 3015–3023. [[CrossRef](#)] [[PubMed](#)]
24. Yoo, J.; Medina-Franco, J.L. Trimethylaurintricarboxylic acid inhibits human DNA methyltransferase 1: Insights from enzymatic and molecular modeling studies. *J. Mol. Model.* **2012**, *18*, 1583–1589. [[CrossRef](#)] [[PubMed](#)]
25. Yoo, J.; Kim, J.H.; Robertson, K.D.; Medina-Franco, J.L. Molecular modeling of inhibitors of human DNA methyltransferase with a crystal structure: Discovery of a novel DNMT1 inhibitor. *Adv. Protein Chem. Struct. Biol.* **2012**, *87*, 219–247. [[PubMed](#)]
26. Méndez-Lucio, O.; Tran, J.; Medina-Franco, J.L.; Meurice, N.; Muller, M. Towards drug repurposing in epigenetics: Olsalazine as a novel hypomethylating compound active in a cellular context. *ChemMedChem* **2014**, *9*, 560–565. [[CrossRef](#)]

27. Aldawsari, F.S.; Aguayo-Ortiz, R.; Kapilashrami, K.; Yoo, J.; Luo, M.; Medina-Franco, J.L.; Velázquez-Martínez, C.A. Resveratrol-salicylate derivatives as selective DNMT3 inhibitors and anticancer agents. *J. Enzyme Inhib. Med. Chem.* **2016**, *31*, 695–703. [[CrossRef](#)]
28. Davide, G.; Sandra, A.; Emily, B.; Mattia, C.; Marta, G.; Annalisa, C.; Livio, S.; Gianluca, M.; Chiara, C.; Eli, F.-d.G.; et al. Design and synthesis of N-benzoyl amino acid derivatives as DNA methylation inhibitors. *Chem. Biol. Drug Des.* **2016**, *88*, 664–676.
29. Palomino-Hernandez, O.; Jardinez-Vera, A.; Medina-Franco, J. Progress on the computational development of epigenetic modulators of dna methyltransferases 3A and 3B. *J. Mex. Chem. Soc.* **2017**, *61*, 266–272. [[CrossRef](#)]
30. Pechalrieu, D.; Dauzonne, D.; Arimondo, P.B.; Lopez, M. Synthesis of novel 3-halo-3-nitroflavones and their activities as DNA methyltransferase inhibitors in cancer cells. *Eur. J. Med. Chem.* **2020**, *186*, 111829. [[CrossRef](#)]
31. Shao, Z.; Xu, P.; Xu, W.; Li, L.; Liu, S.; Zhang, R.; Liu, Y.-C.; Zhang, C.; Chen, S.; Luo, C. Discovery of novel DNA methyltransferase 3a inhibitors via structure-based virtual screening and biological assays. *Bioorg. Med. Chem. Lett.* **2017**, *27*, 342–346. [[CrossRef](#)] [[PubMed](#)]
32. Krishna, S.; Shukla, S.; Lakra, A.D.; Meeran, S.M.; Siddiqi, M.I. Identification of potent inhibitors of DNA methyltransferase 1 (DNMT1) through a pharmacophore-based virtual screening approach. *J. Mol. Graph. Model.* **2017**, *75*, 174–188. [[CrossRef](#)] [[PubMed](#)]
33. Erdmann, A.; Arimondo, P.B.; Guianvarc’h, D. Structure-guided optimization of DNA methyltransferase inhibitors. In *Epi-Informatics*; Medina-Franco, J.L., Ed.; Academic Press: London, UK, 2016; pp. 53–74.
34. Joshi, M.; Rajpathak, S.N.; Narwade, S.C.; Deobagkar, D. Ensemble-based virtual screening and experimental validation of inhibitors targeting a novel site of human DNMT1. *Chem. Biol. Drug Des.* **2016**, *88*, 5–16. [[CrossRef](#)] [[PubMed](#)]
35. Kabro, A.; Lachance, H.; Marcoux-Archambault, I.; Perrier, V.; Dore, V.; Gros, C.; Masson, V.; Gregoire, J.M.; Ausseil, F.; Cheishvili, D.; et al. Preparation of phenylethylbenzamide derivatives as modulators of dnmt3 activity. *MedChemComm* **2013**, *4*, 1562–1570. [[CrossRef](#)]
36. Castellano, S.; Kuck, D.; Viviano, M.; Yoo, J.; López-Vallejo, F.; Conti, P.; Tamborini, L.; Pinto, A.; Medina-Franco, J.L.; Sbardella, G. Synthesis and biochemical evaluation of δ 2-isoxazoline derivatives as DNA methyltransferase 1 inhibitors. *J. Med. Chem.* **2011**, *54*, 7663–7677. [[CrossRef](#)]
37. Newton, A.S.; Faver, J.C.; Micevic, G.; Muthusamy, V.; Kudalkar, S.N.; Bertoletti, N.; Anderson, K.S.; Bosenberg, M.W.; Jorgensen, W.L. Structure-guided identification of DNMT3b inhibitors. *ACS Med. Chem. Lett.* **2020**, *11*, 971–976. [[CrossRef](#)]
38. Naveja, J.J.; Medina-Franco, J.L. Insights from pharmacological similarity of epigenetic targets in epipolypharmacology. *Drug Discov. Today* **2018**, *23*, 141–150. [[CrossRef](#)]
39. Medina-Franco, J.L.; López-Vallejo, F.; Kuck, D.; Lyko, F. Natural products as DNA methyltransferase inhibitors: A computer-aided discovery approach. *Mol. Divers.* **2011**, *15*, 293–304. [[CrossRef](#)]
40. Saldívar-González, F.I.; Gómez-García, A.; Chávez-Ponce de León, D.E.; Sánchez-Cruz, N.; Ruiz-Rios, J.; Pilon-Jiménez, B.A.; Medina-Franco, J.L. Inhibitors of DNA methyltransferases from natural sources: A computational perspective. *Front. Pharmacol.* **2018**, *9*, 1144. [[CrossRef](#)]
41. Akone, S.H.; Ntie-Kang, F.; Stuhldreier, F.; Ewonkem, M.B.; Noah, A.M.; Mouelle, S.E.M.; Müller, R. Natural products impacting DNA methyltransferases and histone deacetylases. *Front. Pharmacol.* **2020**, *11*, 992. [[CrossRef](#)]
42. Lee, W.J.; Shim, J.Y.; Zhu, B.T. Mechanisms for the inhibition of DNA methyltransferases by tea catechins and bioflavonoids. *Mol. Pharmacol.* **2005**, *68*, 1018–1030. [[CrossRef](#)] [[PubMed](#)]
43. Lee, W.J.; Zhu, B.T. Inhibition of DNA methylation by caffeic acid and chlorogenic acid, two common catechol-containing coffee polyphenols. *Carcinogenesis* **2006**, *27*, 269–277. [[CrossRef](#)] [[PubMed](#)]
44. Martínez-Mayorga, K.; Montes, C.P. The role of nutrition in epigenetics and recent advances of in silico studies. In *Epi-Informatics*; Medina-Franco, J.L., Ed.; Academic Press: London, UK, 2016; pp. 385–398.
45. Prieto-Martínez, F.; Peña-Castillo, A.; Méndez-Lucio, O.; Fernández-de Gortari, E.; Medina-Franco, J.L. Molecular modeling and chemoinformatics to advance the development of modulators of epigenetic targets: A focus on DNA methyltransferases. *Adv. Protein Chem. Struct. Biol.* **2016**, *105*, 1–26. [[PubMed](#)]
46. Medina-Franco, J.L.; Méndez-Lucio, O.; Yoo, J.; Dueñas, A. Discovery and development of DNA methyltransferase inhibitors using in silico approaches. *Drug Discov. Today* **2015**, *20*, 569–577. [[CrossRef](#)] [[PubMed](#)]
47. Yuan, Z.; Sun, Q.; Li, D.; Miao, S.; Chen, S.; Song, L.; Gao, C.; Chen, Y.; Tan, C.; Jiang, Y. Design, synthesis and anticancer potential of NSC-319745 hydroxamic acid derivatives as DNMT and HDAC inhibitors. *Eur. J. Med. Chem.* **2017**, *134*, 281–292. [[CrossRef](#)]
48. Erdmann, A.; Halby, L.; Fahy, J.; Arimondo, P.B. Targeting DNA methylation with small molecules: What’s next? *J. Med. Chem.* **2014**, *58*, 2569–2583. [[CrossRef](#)]
49. Gros, C.; Fleury, L.; Nahoum, V.; Faux, C.; Valente, S.; Labella, D.; Cantagrel, F.; Rilova, E.; Bouhlel, M.A.; David-Cordonnier, M.-H.; et al. New insights on the mechanism of quinoline-based DNA methyltransferase inhibitors. *J. Biol. Chem.* **2015**, *290*, 6293–6302. [[CrossRef](#)]
50. Anandakrishnan, R.; Drozdetski, A.; Walker, R.C.; Onufriev, A.V. Speed of conformational change: Comparing explicit and implicit solvent molecular dynamics simulations. *Biophys. J.* **2015**, *108*, 1153–1164. [[CrossRef](#)]
51. Röblitz, S.; Weber, M. Fuzzy spectral clustering by PCCA+: Application to Markov state models and data classification. *Adv. Data Anal. Classif.* **2013**, *7*, 147–179. [[CrossRef](#)]

52. Podvinec, M.; Lim, S.P.; Schmidt, T.; Scarsi, M.; Wen, D.; Sonntag, L.S.; Sanschagrín, P.; Shenkin, P.S.; Schwede, T. Novel inhibitors of dengue virus methyltransferase: Discovery by in vitro-driven virtual screening on a desktop computer grid. *J. Med. Chem.* **2010**, *53*, 1483–1495. [CrossRef]
53. Guianvarc'h, D.; Arimondo, P.B. Challenges in developing novel DNA methyltransferases inhibitors for cancer therapy. *Fut. Med. Chem.* **2014**, *6*, 1237–1240. [CrossRef] [PubMed]
54. Xu, P.; Hu, G.; Luo, C.; Liang, Z. DNA methyltransferase inhibitors: An updated patent review (2012–2015). *Expert Opin. Ther. Pat.* **2016**, *26*, 1017–1030. [CrossRef] [PubMed]
55. Chen, W.; Seefeldt, T.; Young, A.; Zhang, X.; Guan, X. Design, synthesis, and biological evaluation of n-acetyl-(p-chlorophenylcarbamoyl)cysteine and its analogs as a novel class of anticancer agents. *Bioorg. Med. Chem.* **2011**, *19*, 287–294. [CrossRef]
56. Wu, Y.; Liu, H.-B.; Shi, X.-F.; Song, Y. Conventional hypoglycaemic agents and the risk of lung cancer in patients with diabetes: A meta-analysis. *PLoS ONE* **2014**, *9*, e99577. [CrossRef]
57. Prince, H.M.; Bishton, M.J.; Johnstone, R.W. Panobinostat (Ib589): A potent pan-deacetylase inhibitor with promising activity against hematologic and solid tumors. *Future Oncol.* **2009**, *5*, 601–612. [CrossRef] [PubMed]
58. Choi, S.A.; Lee, C.; Kwak, P.A.; Park, C.-K.; Wang, K.-C.; Phi, J.H.; Lee, J.Y.; Chong, S.; Kim, S.-K. Histone deacetylase inhibitor panobinostat potentiates the anti-cancer effects of mesenchymal stem cell-based strail gene therapy against malignant glioma. *Cancer Lett.* **2019**, *442*, 161–169. [CrossRef]
59. Jung, M.; Kim, S.; Lee, J.-K.; Yoon, S.O.; Park, H.S.; Hong, S.W.; Park, W.-S.; Kim, J.E.; Kim, J.; Keam, B.; et al. Clinicopathological and preclinical findings of nut carcinoma: A multicenter study. *Oncologist* **2019**, *24*, e740–e748. [CrossRef] [PubMed]
60. Matsuda, Y.; Yamauchi, T.; Hosono, N.; Uzui, K.; Negoro, E.; Morinaga, K.; Nishi, R.; Yoshida, A.; Kimura, S.; Maekawa, T.; et al. Combination of panobinostat with ponatinib synergistically overcomes imatinib-resistant CML cells. *Cancer Sci.* **2016**, *107*, 1029–1038. [CrossRef] [PubMed]
61. Zopf, S.; Ocker, M.; Neureiter, D.; Alinger, B.; Gahr, S.; Neurath, M.F.; Di Fazio, P. Inhibition of DNA methyltransferase activity and expression by treatment with the pan-deacetylase inhibitor panobinostat in hepatocellular carcinoma cell lines. *BMC Cancer* **2012**, *12*, 386. [CrossRef]
62. Steele, N.; Finn, P.; Brown, R.; Plumb, J.A. Combined inhibition of DNA methylation and histone acetylation enhances gene re-expression and drug sensitivity in vivo. *Br. J. Cancer* **2009**, *100*, 758–763. [CrossRef]
63. Min, H.-Y.; Lee, S.-C.; Woo, J.K.; Jung, H.J.; Park, K.H.; Jeong, H.M.; Hyun, S.Y.; Cho, J.; Lee, W.; Park, J.E.; et al. Essential role of DNA methyltransferase 1-mediated transcription of insulin-like growth factor 2 in resistance to histone deacetylase inhibitors. *Clin. Cancer Res.* **2017**, *23*, 1299. [CrossRef] [PubMed]
64. Morel, D.; Jeffery, D.; Aspeslagh, S.; Almouzni, G.; Postel-Vinay, S. Combining epigenetic drugs with other therapies for solid tumours—Past lessons and future promise. *Nat. Rev. Clin. Oncol.* **2020**, *17*, 91–107. [CrossRef] [PubMed]
65. Rajavelu, A.; Tulyasheva, Z.; Jaiswal, R.; Jeltsch, A.; Kuhnert, N. The inhibition of the mammalian DNA methyltransferase 3a (DNMT3a) by dietary black tea and coffee polyphenols. *BMC Biochem.* **2011**, *12*, 16. [CrossRef] [PubMed]
66. Castillo-Aguilera, O.; Depreux, P.; Halby, L.; Arimondo, P.; Goossens, L. DNA methylation targeting: The DNMT/HMT crosstalk challenge. *Biomolecules* **2017**, *7*, 3. [CrossRef] [PubMed]
67. López-López, E.; Prieto-Martínez, F.D.; Medina-Franco, J.L. Activity landscape and molecular modeling to explore the SAR of dual epigenetic inhibitors: A focus on g9a and DNMT1. *Molecules* **2018**, *23*, 3282. [CrossRef]
68. Prinz, J.H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J.D.; Schütte, C.; Noé, F. Markov models of molecular kinetics: Generation and validation. *J. Chem. Phys.* **2011**, *134*, 174105. [CrossRef]
69. López-López, E.; Barrientos-Salcedo, C.; Prieto-Martínez, F.D.; Medina-Franco, J.L. In silico tools to study molecular targets of neglected diseases: Inhibition of TcSir2rp3, an epigenetic enzyme of *Trypanosoma cruzi*. *Adv. Protein Chem. Struct. Biol.* **2020**, *122*, 203–229.
70. Sánchez-Cruz, N.; Medina-Franco, J.L.; Mestres, J.; Barril, X. Extended connectivity interaction features: Improving binding affinity prediction through chemical description. *Bioinformatic* **2020**, in press. [CrossRef]
71. Reaction Biology Corporation. Available online: <http://www.reactionbiology.com> (accessed on 12 December 2020).
72. Molecular Operating Environment (MOE), Version 2018.08, Chemical Computing Group Inc.: Montreal, QC, Canada. Available online: <http://www.chemcomp.com> (accessed on 12 December 2020).
73. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The protein data bank. *Nucl. Acids Res.* **2000**, *28*, 235–242. [CrossRef]
74. Zhang, Z.-M.; Liu, S.; Lin, K.; Luo, Y.; Perry, J.J.; Wang, Y.; Song, J. Crystal structure of human DNA methyltransferase 1. *J. Mol. Biol.* **2015**, *427*, 2520–2531. [CrossRef]
75. Wang, R.; Lai, L.; Wang, S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comp. Aided Mol. Des.* **2002**, *16*, 11–26. [CrossRef] [PubMed]
76. Onufriev, A.; Bashford, D.; Case, D.A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **2004**, *55*, 383–394. [CrossRef] [PubMed]
77. Korb, O.; Stützel, T.; Exner, T.E. Empirical scoring functions for advanced protein-ligand docking with PLANTS. *J. Chem. Inf. Model.* **2009**, *49*, 84–96. [CrossRef]

78. Ren, X.; Shi, Y.S.; Zhang, Y.; Liu, B.; Zhang, L.H.; Peng, Y.B.; Zeng, R. Novel consensus docking strategy to improve ligand pose prediction. *J. Chem. Inf. Model.* **2018**, *58*, 1662–1668. [[CrossRef](#)] [[PubMed](#)]
79. Maier, J.A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.E.; Simmerling, C. Ff14sb: Improving the accuracy of protein side chain and backbone parameters from ff99sb. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713. [[CrossRef](#)]
80. Aqvist, J.; Warshel, A. Free energy relationships in metalloenzyme-catalyzed reactions. Calculations of the effects of metal ion substitutions in staphylococcal nuclease. *J. Am. Chem. Soc.* **1990**, *112*, 2860–2868. [[CrossRef](#)]
81. Liao, Q.; Pabis, A.; Strodel, B.; Kamerlin, S.C.L. Extending the nonbonded cationic dummy model to account for ion-induced dipole interactions. *J. Phys. Chem. Lett.* **2017**, *8*, 5408–5414. [[CrossRef](#)]
82. Eastman, P.; Swails, J.; Chodera, J.D.; McGibbon, R.T.; Zhao, Y.; Beauchamp, K.A.; Wang, L.P.; Simmonett, A.C.; Harrigan, M.P.; Stern, C.D.; et al. Openmm 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **2017**, *13*, e1005659. [[CrossRef](#)]
83. McGibbon, R.T.; Beauchamp, K.A.; Harrigan, M.P.; Klein, C.; Swails, J.M.; Hernández, C.X.; Schwantes, C.R.; Wang, L.P.; Lane, T.J.; Pande, V.S. Mdtraj: A modern open library for the analysis of molecular dynamics trajectories. *Biophys. J.* **2015**, *109*, 1528–1532. [[CrossRef](#)]
84. Hagberg, A.; Schult, D.; Swart, P. Exploring network structure, dynamics, and function using networkx. In *Proceedings of the 7th Python in Science Conference (SciPy 2008)*; Varoquaux, G., Vaught, T., Millman, J., Eds.; SciPy Conference: Pasadena, CA, USA, 2008; pp. 11–15.
85. Swinburne, T.D.; Wales, D.J. Defining, calculating, and converging observables of a kinetic transition network. *J. Chem. Theory Comput.* **2020**, *16*, 2661–2679. [[CrossRef](#)]
86. Scherer, M.K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noé, F. Pyemma 2: A software package for estimation, validation, and analysis of markov models. *J. Chem. Theory Comput.* **2015**, *11*, 5525–5542. [[CrossRef](#)] [[PubMed](#)]
87. Plattner, N.; Noé, F. Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. *Nat. Commun.* **2015**, *6*, 7653. [[CrossRef](#)] [[PubMed](#)]
88. Husic, B.E.; Pande, V.S. Markov state models: From an art to a science. *J. Am. Chem. Soc.* **2018**, *140*, 2386–2396. [[CrossRef](#)] [[PubMed](#)]
89. Jurrus, E.; Engel, D.; Star, K.; Monson, K.; Brandi, J.; Felberg, L.E.; Brookes, D.H.; Wilson, L.; Chen, J.; Liles, K.; et al. Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* **2018**, *27*, 112–128. [[CrossRef](#)] [[PubMed](#)]