

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

Instituto de Biotecnología.

"Desarrollo de un protocolo para la identificación de un nuevo fármaco anti-SARS CoV2 utilizando herramientas computacionales de diseño de fármacos."

TESIS

QUE PARA OPTAR POR EL GRADO DE: Maestro en Ciencias

PRESENTA: Lic. Jorge Samuel León Magdaleno

> Dr. Ramón Garduño Juárez Instituto de Ciencias Físicas

MIEMBROS DEL COMITÉ TUTOR Dr. Edmundo Castillo Rosales Instituto de biotecnología.

Dr. Enrique Alejandro Reynaud Garza Instituto de biotecnología.

> Dr. José Luis Medina Franco Facultad de Química.

Cuernavaca, Mor., noviembre, 2022



Universidad Nacional Autónoma de México



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.

Agradecimientos.

Al Consejo Nacional de Ciencia y Tecnología por otorgarme el apoyo económico para realizar mis estudios de maestría (**CVU 1083135**).

Al proyecto asociado de la Dirección General de Cómputo y Tecnologías de la Información LANCAD-UNAM-DGTIC-033 para el uso de la Supercomputadora HP Clúster Platform 3000SL (MIZTLI).

Al proyecto DGAPA (IT102822 "IDENTIFICACIÓN DE POSIBLES INHIBIDORES DE LA VIROPIRONA DEL SARS-CoV-2 EMPLEANDO MÉTODOS DE DISEÑO DE FÁRMACOS ASISTIDOS POR COMPUTADORA").

A mi familia por el apoyo brindado.

Al Dr. Ramón Garduño Juárez por su paciencia y apoyo durante el desarrollo de este proyecto.

Al doctor José Luis Medina Franco por su tiempo, sus invaluables recomendaciones y consejos.

Por último, a mi comité totoral conformado por el Dr. Enrique Reynaud Garza y el Dr. Edmundo Castillo Rosales, por todas sus aportaciones y críticas constructivas que me dieron pie a pulir este trabajo.

Tabla de contenido

GLOSARIO DE ABREVIATURAS	4
Resumen	5
Introducción	6
Antecedentes	7
Características estructurales y genómicas del SARS CoV-2.	7
Ciclo de vida viral	8
NSPs	
NSP1	
NSP2	
NSP3	
PLpro	
NSP4	
NSP5	
NSP6	
NSP7/NSP8/NSP12.	14
NSP 9	
NSP10	
NSP11	
NSP13	
NSP14	
NSP15	
NSP16	
Proteínas accesorias.	
ORF3a	
ORF7a	20
ORF8	21
ORF9b	21
Proteínas estructurales	
Proteína espiga (S)	
Proteína de la Nucleocápside (N)	23
Proteína de Membrana (M)	23
Proteína E	24
Estructura y función de las viroporinas.	25

Inhibidores actuales de la proteína E	27
Diseño de fármacos asistido por computadora	27
Regla de los 5 de Lipinski	29
Compuestos PAINS	30
Cálculo de propiedades ADMET	31
Acople Molecular	32
Acople Molecular por consenso	32
Moléculas Señuelo	33
Huellas digitales de las interacciones proteína-ligando.	33
Planteamiento del problema	35
JUSTIFICACIÓN	35
HIPÓTESIS	35
METAS	35
OBJETIVO PRINCIPAL	36
OBJETIVOS PARTICULARES	36
METODOLOGÍA	36
Obtención de la estructura de la viroporina	36
Obtención de señuelos	36
Obtención y preparación de los ligandos	36
Preparación del receptor.	38
Conversión de formatos	38
Obtención de huellas digitales de interacción proteína-ligando y mapas de interacciones en dos dimensiones.	38
Cálculo de la desviación estándar de la raíz cuadrática (RMSD)	39
Generación de las curvas de característica operativa del receptor (Curvas ROC).	39
Cálculo de propiedades ADMET	40
Simulaciones de acople molecular	40
Obtención de poses consenso.	41
Ecuación 1:	41
Tc (A, B)) A ∩ B / A ∪ B	42
Cálculo de propiedades ADMET	42
Resultados y discusión	44
Resultado del acople molecular de señuelos	44
Resultado del acople molecular ciego y análisis de interacciones de las moléculas activas convergentes.	45
Resultados curvas ROC	55
Resultado del cribado virtual.	57

Comparación con estudios actuales de inhibición	100
Evaluación del perfil farmacocinético de los nuevos inhibidores	100
Conclusiones	111
Perspectivas.	112
BIBLIOGRAFÍA	113

GLOSARIO DE ABREVIATURAS.

1D Una Dimensión.

2D Dos Dimensiones.

3D Tres Dimensiones.

ADMET Absorción, distribución, metabolismo, absorción, excreción y toxicidad.

ARNm ARN mensajero.

AUC Área bajo la curva.

Cap Capuchón.

COVID Enfermedad de Coronavirus.

CV Cribado Virtual.

DM Dinámica Molecular.

DL50 Dosis letal 50

DTM Dominio Transmembranal.

NA Análogo de nucleótidos.

NSP Proteínas no estructurales.

ORF Marco de Lectura Abierto.

PAINS Pan-Assay Interference Compounds

PDB Banco de Datos de Proteínas.

PDBID Código de identificación en el PDB.

PLANTS Protein-Ligand ANTSystem.

PLIFS Huellas digitales de las interacciones proteína-ligando.

Puentes H Puentes de hidrogeno.

QSAR Quantitative Structure Activity Relationships.

RMSD Desviación Estándar de la Raíz Cuadrática.

ROC Curva característica de funcionamiento del receptor.

RTC Complejo de Replicación - Transcripción.

SMILES Simplified Molecular-Input Line-Entry System.

SPORES Structure PrOtonation and Recognition System.

Resumen.

En 2019 en la ciudad de Wuhan provincia de Hubei, China, surgieron una serie de casos de enfermedad respiratoria, descrita como una neumonía atípica. Estudios de biología molecular determinaron que el agente etiológico responsable de producir la enfermedad era un virus, perteneciente a la familia de los Coronavirus, el cual se nombró SARS CoV2. El SARS CoV-2 es un virus pleomórfico con envoltura que posee una cadena de ARN de sentido positivo de 32 kb. Taxonómicamente pertenece a el reino Riboviria, al orden de los Nidovirus, al suborden Cornidoviridae, a la familia Coronaviridae, a la subfamilia Orthocoronaviridae, al género Beta coronavirus, al subgénero Sarbecovirus y a las especies de coronavirus relacionadas con síndrome agudo respiratorio. El SARS CoV-2 tiene un total de 11 genes con 15 marcos de lectura abierto (ORF, Open Reading Frames): ORF 1ab, ORF 2 (proteína espiga), ORF 3a, ORF 4 (proteína E), ORF 5 (proteína de la membrana), ORF 6, ORF 7a, ORF 7b, ORF 8, ORF9 (proteína de la nucleocápside) y ORF 10.). El primer paso es la interacción de la proteína S con el receptor ACE2 del huésped, lo que conlleva a la liberación del material genético dentro del citoplasma. La proteína de la envoltura, o proteína E, es una proteína con una longitud de 76-109 aminoácidos esencial para el virus SARS-CoV-2, ya que forma un canal catiónico homopentamérico que es importante para la patogenicidad del virus. Se ha encontrado que la actividad de la proteína es intracelular, específicamente se ha localizado en la superficie de la membrana del retículo endoplásmico (Nieto-Torres et al., 2011). La proteína E es permeable a Ca2+, actúa en la disipación del potencial de para lograr la gemación de la membrana y generar la envoltura del virus.

Esta enfermedad fue declarada una emergencia de salud pública el 30 de junio de 2020. Esta enfermedad ha golpeado al mundo cobrando millones de vidas, así como a los sectores educativos y económicos. Actualmente existe una reducida cantidad de fármacos aprobados para combatir la infección, entre ellos el más conocido el Remdesivir, que es costoso y de uso hospitalario. Recientemente se han aprobado otras tres moléculas; Nirmatrelvir con Ritonavi, el cual es eficaz solo si se toma dentro de los primeros cinco días de la infección, Bebtelovimab, de administración intravenosa, y Molnupiravir. Debido a lo mencionado la comunidad científica se ha visto en la necesidad de buscar nuevos inhibidores, pero desafortunadamente todos los esfuerzos se han concentrado en la proteína S y la Mpro. Las herramientas computacionales en el campo de diseño de fármacos han demostrado ser muy útiles ya que permiten reducir tiempo y recursos en la búsqueda de nuevas moléculas con actividad contra algún blanco biológico. Una de las metodologías más utilizadas para explorar la afinidad de las moléculas por una diana biológica es el cribado virtual basado en estructura, cuya herramienta principal es el uso del acople molecular. Existen diversos tipos de algoritmos de acople molecular como los basados en optimización geométrica o los basado en algoritmo genético, pero lamentablemente todos poseen puntos deficientes durante la predicción de los probables sitios de interacción de un ligando en una macromolécula. En este protocolo aplicamos el uso de múltiples algoritmos de acople molecular para aumentar la precisión de las predicciones realizadas por un solo algoritmo de acople molecular, para obtener una pose consenso. Esta metodología está apoyada con el uso del cálculo de propiedades ADMET, el cual nos permitió encontrar moléculas con buen perfil de fármaco, así como un buen perfil toxicológico.

Introducción.

Los coronavirus humanos han aparecido periódicamente en diferentes partes alrededor del mundo, y han sido responsables de ocasionar pandemias causantes de neumonías letales desde el comienzo del siglo XXI (Wu et al., 2020). La primera pandemia producida por un coronavirus humano (SARS CoV) inició en noviembre del 2002 en Foshan, China, y ésta en el 2003 se convirtió rápidamente en un problema de salud global que generó una tasa de mortalidad del 10% (8096 casos, 774 muertes) en el mundo entero. Una década después, apareció la segunda pandemia generada por un coronavirus humano, causada por el virus de síndrome respiratorio de medio oriente (MERS-CoV), originado en junio del 2012 en Jeddah, Arabia Sudita (Hu et al., 2015) con una tasa de letalidad del 35 % (1413 casos,502 muertes). Recientemente la tercera pandemia, se originó en diciembre de 2019 en la provincia de Wuhan en China, causada por el novedoso virus del síndrome agudo respiratorio 2 (SARS CoV-2), nombrado así por el alto porcentaje de identidad de su genoma con el de SARS CoV. La organización mundial de la salud designó cómo COVID-19 al síndrome respiratorio generado por este agente etiológico, que se convirtió en una emergencia de salud pública a nivel internacional el 30 de junio de 2020 (Burki TK., 2020). Durante el periodo que duró la pandemia ha habido pérdidas irreparables como la muerte de millones de personas, pero aparte de esto ha impactado la economía global. Algunos países inhabilitaron por completo los medios de transporte, impidiendo cualquier tipo de actividad por un largo periodo de tiempo durante la pandemia. El sector educativo fue uno más afectados, debido a la inactividad durante dos años, producto de las medidas de contingencia impuestas por los gobiernos locales. Ahora todas las economías del mundo enfrentaron una alta inflación, lo cual se traduce en un incremento al del desempleo y baja producción del sector industrial (Chakraborty & Maity, 2020).

Debido a la urgencia de contar con alternativas para tratar la COVID-19, es necesario desarrollar protocolos y herramientas que nos permitan, agilizar al mismo tiempo que ahorrar recursos, realizar el descubrimiento de nuevos fármacos y vacunas. Generalmente, el descubrimiento de fármacos es un proceso que toma alrededor de 15 años con costos de miles de millones de dólares para que un fármaco esté disponible en el mercado (DiMasi et al., 2016). En la mayoría de las investigaciones farmacéuticas las tasas de éxito son muy reducidas, tan solo 13%. En la mayoría de los casos, las moléculas candidatas fallan en etapas finales de los estudios farmacológicos, debido a las pobres características farmacocinéticas en absorción, distribución, excreción y toxicidad (Hou & Xu, 2004). El uso de técnicas de diseño de fármacos asistido por computadora aplicadas a

los estudios preliminares han ayudado a minimizar los costos y reducir el número de moléculas fallidas en las etapas finales de los estudios (C. Basak, 2013). La aplicación del diseño racional de ligandos es una parte integral de las técnicas computacionales de diseño de fármacos que nos permite entender cuál es la afinidad de un ligando por una proteína además del modo y tipos de interacción (Macalino et al., 2015).

Antecedentes.

Características estructurales y genómicas del SARS CoV-2.

El SARS CoV-2 es un virus pleomórfico con envoltura que posee una cadena de ARN de sentido positivo de 32 kb. Taxonómicamente pertenece a el reino *Riboviria*, al orden de los *Nidovirus*, al suborden *Cornidoviridae*, a la familia *Coronaviridae*, a la subfamilia *Orthocoronaviridae*, al género *Beta coronavirus*, al subgénero *Sarbecovirus* y a las especies de coronavirus relacionadas con síndrome agudo respiratorio (Liu., et al. 2014; Zhu et al., 2020). El SARS CoV-2 tiene un total de 11 genes con 15 marcos de lectura abierto (ORF, Open Reading Frames): ORF 1ab, ORF 2 (proteína espiga), ORF 3a, ORF 4 (proteína E), ORF 5 (proteína de la membrana), ORF 6, ORF 7a, ORF 7b, ORF 8, ORF9 (proteína de la nucleocápside) y ORF 10 (Yoshimoto FK, 2020).

El primer gen, ORF 1ab, expresa una poliproteína la cual contiene a las poliproteínas 1a y 1b que mediante cortes proteolíticos realizados por la 3CL^{Pro} (gen 1a) genera las 16 proteínas no estructurales (NSP, Non Structural Protein) (Yoshimoto FK, 2020). Las NSPs intervienen en las distintas fases del ciclo de replicación viral, ya que participan como elementos críticos del complejo de replicación y transcripción (RTC) así como en la evasión de la respuesta inmune del hospedero (Gasmalbari y Abbadi, 2020). En la figura 1 se muestra la estructura del SARS CoV-2 y la organización de su genoma.



Figura 1.- En esta figura se muestra la estructura del SARS CoV-2 y la organización de su genoma, especificando que proteína es codificada a partir de cada gen, además del PDBID de cada proteína en el PDB (adaptada de Lubin et al., 2021).

Ciclo de vida viral.

En el tracto respiratorio, el SARS CoV-2 invade preferentemente las células calciformes y las células ciliadas, lo cual se ha probado a través de estudios realizados en biopsias y en estudios post-mortem (Hui et al., 2020). El primer paso es la interacción de la proteína S con el receptor ACE2 del huésped, específicamente con el dominio de unión del receptor (RBD) (Wang et al., 2020). Posteriormente la proteína S sufre cambios conformacionales debido a la acción de proteasas como la TMPRSS2 (Hoffmann et al., 2020) lo que conlleva a la liberación del material genético dentro del citoplasma. Una vez dentro del citoplasma, el material genético es traducido para obtener las poliproteínas pp1a y pp1ab, que subsecuentemente serán cortadas en proteínas más pequeñas (NSPS, ORF, estructurales). Una vez formadas las proteínas estructurales y el material genético replicado, los nuevos virus son ensamblados y liberados por exocitosis (V'kovski et al., 2020). En la figura 2 se muestra el esquema del ciclo de vida del SARS CoV-2.



Figura 2.- Ciclo de vida del SARS CoV-2. El SARS CoV-2 interactúa con el receptor ACE2 y con proteasas como la TMPRSS2 que produce un cambio conformacional en la proteína S, para producir la internalización del material genético al citoplasma. Una vez el material genético es traducido a pp1a y pp1ab, las cuales son procesadas para producir las NPS que forman el RTC y las proteínas estructurales para llevar a cabo el ensamble de los nuevos virus (adaptada de Lebeau et al., 2020).

Actualmente las proteínas que han adquirido mayor protagonismo en las investigaciones son las proteínas 3CLpro y la proteína espiga. Esto es debido a la importancia de su actividad dentro del ciclo infectivo del SARS CoV-2. La proteína espiga, o S, es el primer punto de contacto entre el virus y la célula del huésped. Esta interacción se da a través del dominio de unión al receptor (RBD, Receptor Binding Domain) el cual interactúa con la enzima convertidora de angiotensina-2 (ACE2), el receptor funcional para el SARS-CoV-2 expresado en las células alveolares, renales e intestinales (Hoffman et al.,2020).

A continuación, se realizó una breve descripción de las proteínas del SARS-CoV2 con el propósito de mostrar un panorama general acerca de los posibles blancos biológicos que se pueden abordar y dar a conocer al lector de manera sucinta, el papel que llevan a cabo cada una de ellas. Las figuras fueron elaboradas usando PyMOL (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.)

NSPs.

NSP1.

Esta proteína es el producto del corte del extremo N-terminal de los precursores, pp1a y pp1ab, producido por la actividad proteolítica de la papain proteasa viral (PLpro). La NSP1 interfiere con la síntesis proteica de las células del hospedero uniéndose a la subunidad ribosomal 40S, además de promover la degradación del ARNm (Kamitani et al., 2009). La NSP1 de SARS-CoV-2 muestra un 84% de identidad en su secuencia de aminoácidos en comparación con la NSP1 de SARS-CoV, sugiriendo que éstas son estructuras similares. Actualmente hay varias estructuras de las NSP del SARS-CoV-2 disponibles en el PDB (<u>https://www.rcsb.org/</u>) como la mostrada en la Figura 3 (Clark et al., 2021)



Figura 3.- Estructura de la proteína NSP1 obtenida mediante rayos X. Esta estructura puede ser encontrada en el PDB con el código 7k7p.

NSP2.

La NSP2 es una proteína conservada en los virus SARS. Se ha encontrado que la NSP2 se une a dos proteínas del huésped: las prohibitinas 1 y 2 (PHB1 y PHB2). Estas proteínas están relacionadas con la progresión del ciclo celular, migración celular, diferenciación, apoptosis y biogénesis mitocondrial (Zheng et al., 2021). En la figura 4 se muestra la estructura de la NSP2 encontrada en el PDB con PDBID 7MSW (QCRG Structural Biology Consortium, 2021).



Figura 4.- Estructura de la NSP2. La proteína obtenida por difracción de rayos X (PDBID:7MSW).

NSP3.

La NSP3 es una proteína con múltiples dominios, siendo la más larga codificada a partir del genoma del SARS CoV-2 (1945 aminoácidos) (Lei et al., 2018). La función de esta proteína es formar un anclaje entre la membrana del retículo endoplásmico con otros NSP y las proteínas del huésped para formar el RTC (Imbert et al., 2008). La proteína es obtenida mediante cortes proteolíticos realizados por la 3CLpro en las poliproteínas 1a y 1ab (Harcourt et al., 2004). Actualmente existen varias estructuras de esta proteína depositadas en el PDB. En la Figura 5 (PDBID:6W02) se muestra la estructura de la NSP3 obtenida a partir de difracción de rayos X (Michalska et al., 2020).



Figura 5.- Estructura de la NSP3. La proteína obtenida por difracción de rayos X (PDBID:6W02).

PLpro.

La función de la PLpro es cortar las poliproteínas virales, precursoras de las NSPS. Los cortes se producen en tres sitios que dan origen a las NSP1, NSP2 y NSP3 (Harcourt et al., 2004). La secuencia de reconocida por PLpro es LXGGX. Existe un alto porcentaje de similitud entre la PLpro de SARS CoV y la de SARS CoV-2 (83% de identidad). Actualmente existe un reducido número de estructuras de PLpro en el PDB. La figura 6 muestra la

estructura de la PLpro obtenida por difracción de rayos X.



Figura 6.- Estructura de la PLpro obtenida por difracción de rayos X con PDBID 7D47

NSP4.

La proteína NSP4 funciona como enriquecedor de las interacciones formadas por la NSP3. Esto lo realiza anclándose a las membranas del retículo endoplásmico y la mitocondria (Davies et al., 2020). Actualmente la función y significancia biológica de la NSP4 aún se encuentran bajo estudio. La estructura de la NSP4 aún no se encuentra depositada en al PDB.

NSP5.

La proteína NSP5 (3CLpro o Mpro), es una proteína de 33 kDa que tiene actividad de proteasa y promueve la maduración de las NSPs 4-16 que son importantes para la replicación viral. Esta proteína posee un alto grado de conservación con relación al SARS-CoV y al MERS-CoV (Tomar et al.,2015). Las proteínas NSP5 de SARS CoV y SARS CoV-2 comparten un 96% de similitud en sus secuencias. La Mpro corta las poliproteínas 1a y 1 ab en 12 proteínas funcionales (después de la actividad de PLpro) (Ziebuhr et al., 2000). Recientemente se han probado con éxito algunos fármacos como boceprivir (Ma et al., 2020) y telaprevir (Bafna et al., 2021). En la actualidad existen una gran cantidad de estructuras de la Mpro co-cristalizadas con el inhibidor boceprivir (PDBIDs: 6ZRU, 6WNP, 6ZRT, 7C7P). En la figura 7 se muestra la estructura de la Mpro obtenida a partir de difracción de rayos X (PDBID: 6YB7) (Owen et al., 2020).



Figura 7.- Estructura de la Mpro obtenida por difracción de rayos X con PDBID 6YB7.

NSP6.

La proteína NSP6 está relacionada con la formación de autofagosomas que facilitan el ensamble de proteínas de replicación, limitando la formación de lisosomas evitando que los componentes virales sean degradados (Cottam et al.,2014). Aún no hay estructuras disponibles de NSP6.

NSP7/NSP8/NSP12.

La proteína NSP12 (RdRp) es una proteína de 103 KDa conformada por múltiples subunidades, que funge como polimerasa dependiente de ARN. A lo largo del proceso de replicación, esta proteína es auxiliada por las proteínas NSP7 y NSP8 para formar el complejo de replicación. La NSP7 y 8 ayudan a brindarle mayor estabilidad a la RdRp durante la replicación del material genético formando un anillo alrededor del ARN. La NSP12 del SARS CoV-2 es muy similar a la de SARS CoV, y comparten un 96% de identidad entre sus secuencias (Gao et al., 2020). Se han encontrado algunos inhibidores análogos de nucleótidos (NA), similares a los inhibidores RH-RT usados contra el VIH (Crotty et al., 2000, Agostini et al., 2019). El NA remdesivir (GS-5734) recibió una autorización de emergencia por la FDA para tratar la COVID. El remdesivir ha mostrado una gran afinidad por RdRp y una alta velocidad de incorporación, mayor que la velocidad de corte de las exonucleasas que remueven las bases incorrectas (Agostini et al., 2018). Actualmente existen varias estructuras de la RdRp depositadas en el PDB. En la figura 8 se muestra la estructura de RdRp formando un complejo con la NSP7 y NSP8 (PDBID 7BW4), esta estructura se obtuvo mediante microscopia electrónica (Peng et al., 2020).



Figura 8.- Estructura del complejo de replicación formado por la RdRp (Azul), la NSP7 (verde) y la NSP8 (rojo y amarillo) (PDBID 7BW4).

NSP 9.

La NSP9 es una proteína de unión a ARN de hebra sencilla implicada en la virulencia del virus. La NSP9 del SARS CoV 2 comparte un 97 % de identidad con la secuencia de la proteína de SARS CoV. Esta proteína consiste en una serie de horquillas hacia fuera de su núcleo formado por barriles beta, los cuales están formados por aminoácidos cargados positivamente que interactúan con el ARN. En la figura 9 se muestra la estructura de la NSP 9 obtenida por difracción de rayos X (Littler et al., 2020).



Figura 9.- Estructura de la proteína NSP9 (PDBID 6W9Q).

NSP10.

La NSP 10 es una proteína pequeña de 139 aminoácidos con un solo dominio. La NSP 10 de SARS CoV-2 comparte un 99% de similitud con su homóloga de SARS CoV. NSP 10 funciona como una proteína cebadora para formar el complejo de metilación de la caperuza 5' de ARNm junto con la NSP14 y la NSP16 (Romano et al., 2020). La NSP10 ayuda a estabilizar y acelerar la actividad de las NSP14 y NSP16. La estructura de la NSP10 (PDBID 6ZCT) consiste en una lámina beta antiparalela y un dominio alfa-hélice (Figura 10) (Krafcikova et al., 2020).



Figura 10.- Estructura de la NSP10 obtenida mediante difracción de rayos X (PDBID 6ZCT).

NSP11.

La NSP11 es un péptido de 13-23 aminoácidos cuya función y estructura aún no han sido determinadas, aunque influye en el marco de lectura y en el procesamiento de la NSP10 y la NSP12 (Gadhave et al.,2020).

NSP13.

La NSP13 es una proteína con función de helicasa y de nucleósido trifosfatasa (NTPasa) (Chen et al., 2020). La estructura de la NSP13 del SARS CoV-2 es muy similar a la de SARS CoV. La estructura de la proteína está conformada por 5 subunidades que de manera coordinada actúan para llevar a cabo la actividad de helicasa. Actualmente existen algunas estructuras de esta proteína en el PDB. En la figura 11 se muestra la estructura de la NSP13 con PDBID 6ZSL obtenida a partir de difracción de rayos X (Newman et al., 2021). La NSP13 ha sido clasificada como un blanco para el desarrollo de antivirales. Entre los diversos inhibidores reportados en la literatura, el SSYA10-001 es uno de los más prometedores candidatos el cual interrumpe la actividad de helicasa de la proteína (Adedeji et al., 2014).



Figura 11.- Estructura de la NSP13 obtenida mediante difracción de rayos X (PDBID 6ZSL).

NSP14.

La NSP 14 del SARS CoV-2 tiene una función dual, teniendo un N-terminal con actividad de exonucleasa para corregir las bases agregadas incorrectamente, y un extremo C-terminal implicado en la transferencia de un grupo metilo a él N-7 de la guanina para formar la caperuza del ARNm (Y. Ma et al., 2015). La estructura del SARS CoV-2 ha sido determinada, y comparte un 99.1% de similitud con la de SARS CoV. Debido al papel crucial de exonucleasa de la NSP14, ésta interfiere con el papel de los antivirales NA, por lo que se ha propuesto desarrollar antivirales que inhiban esta proteína para crear un efecto sinérgico (Morse et al., 2020). En la figura 12 se puede apreciar la estructura de NSP14 de SARS CoV-1 obtenida por difracción de rayos X con PDBID 7QIF (Newman, J.A, et al., 2022).



Figura 12.- Estructura de la NSP14 obtenida mediante difracción de rayos X (PDBID 7QIF).

NSP15.

La NSP15 es una endoribonucleasa que corta el ARN específicamente en el extremo 3' de los uridilatos, según una investigación (Kim et al., 2020), la endoribonucleasa se compone de un ensamble hexamérico formado por 6 monómeros de la NSP 15. La actividad de esta proteína ayuda a el virus a evadir la respuesta inmune del huésped, previniendo la detección del ARN viral de doble cadena. La NSP15 corta las poliuridinas de la hebra de sentido negativo del ARN viral, las cuales son patrones asociados con patógenos, reconocidos por la MDA-5 (melanoma differentiation-associated gene 5) (Hackbart et al., 2020). Estudios de mutagénesis en la proteína han mostrado un impacto sobre la replicación llegando hasta la atenuación de la infección, por esta razón se ha considerado un blanco estratégico para el desarrollo de antivirales (Deng et al., 2017). Recientemente se han realizado estudios en los cuales el fármaco anticancerígeno Tipiracil, inhibe la función de la NSP15. La estructura de la NSP15 en complejo con la tipiracil ha sido determinada (figura 13) (Kim et al., 2020).



Figura 13.- Estructura de la NSP15 en complejo con el inhibidor tipiracil obtenida mediante difracción de rayos X (PDBID 6WXC).

NSP16.

La NSp16 es una m7GpppA (Cap-1)-specific, S-adenosil metionina 2'-O-metiltransferasa (SAM) que cataliza la formación de la caperuza en la posición 5' del ARN (Bouvet et al., 2010). La metilación del ARNm en su extremo 5'lo protege de su degradación por parte de la 5'-exoribonuleasa del huésped, lo cual contribuye a los mecanismos de evasión de la respuesta inmune (Almazán et al., 2006). A pesar de ser un atractivo blanco para el desarrollo de terapias antivirales, aún no hay alguno disponible. Actualmente existen estructuras de esta proteína depositadas en el PDB, tal como 6WKS, un heterodímero formado por la NSP16 y la NSP 10 (Viswanathan et al., 2020). En la figura 14 se muestra la estructura de la NSP16 obtenida del PDBID 6WKS.



Figura 14.- Estructura de la NSP16 (azul) en complejo con la NSP10 (verde), obtenida mediante difracción de rayos X (PDBID 6WXC).

Proteínas accesorias.

ORF3a.

La proteína ORF3a es una proteína transmembranal que forma canales iónicos en la membrana del retículo endoplásmico de la célula hospedera. Ésta comparte aproximadamente un 72% de identidad con la secuencia de su homónima en el SARS CoV. La proteína ha sido implicada en funciones como: la inducción de la apoptosis, la patogenicidad y la liberación de virus ensamblados (Ren et al., 2020). Recientemente se ha dilucidado la estructura del ORF3a (PDBID 6XDC). Se ha reportado que el canal tiene una gran preferencia por Ca²⁺ y K⁺. La inhibición de la proteína resulta en una disminución de la liberación de virus durante el ciclo de vida del SARS CoV-2, y por esta razón se ha propuesto como un potencial blanco para el desarrollo de nuevos fármacos antivirales (Kern et al., 2021). En la figura 15 se muestra la estructura de la ORF3a obtenida mediante criomicroscopía electrónica.



Figura 15.- Estructura de la ORF3a obtenida mediante microscopia electrónica (PDBID 6XDC). La región transmembranal se encuentra en medio de las figuras marcadas como membrana

ORF7a.

La proteína ORF7a es una proteína transmembranal, también conocida como U122, o proteína X4. Esta proteína posee un 85 % de identidad con la ORF7a de SARS CoV. Estudios sugieren que esta proteína permanece retenida en las células infectadas. A pesar de que está involucrada en las interacciones entre el virus y la célula huésped, muchas de sus funciones aún son desconocidas, hasta ahora se supone que participa en el tráfico molecular entre el retículo endoplásmico y el aparato de Golgi. Actualmente existen algunas estructuras de la proteína de SARS CoV-2 en el PDB (7Cl3) (Zhou Z., et al, 2021). En la figura 16 se muestra la estructura de la ORF7a, esta estructura corresponde al



Figura 16.- Estructura del ectodominio ORF7a de SARS CoV obtenida mediante microscopia electrónica (PDBID 6W37).

ectodominio o dominio tipo inmonuglobulina de la proteína, que forma interacciones con las células del sistema inmunológico y ayuda a evadir la respuesta inmunedel hospedero.

ORF8.

La ORF8 es una proteína accesoria de 121 residuos, que posee un N-terminal con una secuencia de aminoácidos que la señalizan para ser transportada al retículo endoplásmico, y un núcleo formado por una lámina beta que forma un complejo tipo inmunoglobulina. La similitud entre las secuencias de ORF8 entre SARS CoV-2 y SARS CoV es de apenas el 20%. A cerca de su función se han especulado diversas funciones, donde la más consistente parece ser la disminución de la función del MHC-I (complejo mayor de histocompatibilidad I), promoviendo la evasión del sistema inmune (Zhang et al., 2020). Actualmente la función biológica exacta aún no está clara por lo que necesitan realizarse más investigaciones. En el PDB hay depositadas algunas estructuras de esta proteína, en la figura 17 se muestra una estructura de la ORF8a en forma dimérica (PDBID7 JTL) (Flower et al., 2020).



Figura 17.- Estructura de la ORF8 obtenida mediante difracción de rayos X (PDBID 7JTL). **ORF9b.**

La ORF9b es una proteína accesoria de 97 aminoácidos. La función de la proteína es suprimir la respuesta del IFN-I (Interferón tipo I) (Jiang et al., 2020). El IFN I tiene un papel central en la respuesta inmune para combatir infecciones virales. La ORF9b de SARS CoV-2 muestra un 70.45% de identidad respecto a la ORF9b de SARS CoV. La proteína se encuentra cristalizada en el PDB con el PDBID 6Z4U, en donde se muestra que es una proteína dimerica compuesta principalmente por dominios hebra beta (Weeks et al., 2020). La estructura de la proteína ORF9b se muestra en la figura 18.



Figura 18.- Estructura de la ORF9b obtenida mediante difracción de rayos X (PDBID 6Z4U).

Proteínas estructurales.

Proteína espiga (S).

La envoltura en forma de corona del SARS CoV-2 contiene una gran cantidad de proyecciones en su superficie llamadas glicoproteínas largas de la superficie o proteínas espiga; son las responsables de reconocer los receptores de la membrana de la célula del huésped y su posterior fusión. La proteína S de SARS CoV-2 comparte un 77% de identidad con la proteína S de SARS CoV, esta proteína se une al receptor ACE2 para facilitar la entrada del virión al citoplasma de la célula hospedera (Wang et al., 2020). La proteína S es un trímero que está compuesto por tres subunidades principales, la S1 (residuos 14-685), la S2 (residuos 686-1273) y el dominio de unión a receptor (RBD) (residuos 319-541), con una longitud de 1273 aminoácidos. La subunidad S1 tiene un papel importante en el reconocimiento y unión del virus al receptor ACE2, mientras que la subunidad S2 media la fusión de la membrana del virus con la del huésped (Walls et al., 2020). Algunos estudios han encontrado que la proteína S puede ser inhibida mediante el diseño de anticuerpos específicos como CR3022 o el 47D11, por lo cual es un blanco potencial para el desarrollo de vacunas (C. Wang et al., 2020). Actualmente existen varias estructuras de la proteína S depositadas en el PDB. En la figura 19 se muestra la proteína S, una estructura obtenida por crio-microscopía electrónica (PDBID 6VYB) (Walls et al., 2020).



Figura 19.- Estructura de la proteína S obtenida mediante difracción de rayos X (PDBID 6VYB). En esta imagen se pueden visualizar las tres cadenas que componen la proteína: A (púrpura), B (Azul) y C (Cian).

Proteína de la Nucleocápside (N).

La proteína de la Nucleocápside (N) es una importante proteína estructural que se encarga del empaquetamiento del ARN viral dentro de una ribonucleocápside helicoidal (Chang et al., 2013). Esta proteína es producida en grandes cantidades durante la infección y es altamente inmunogénica por lo que es un blanco potencial para el desarrollo de vacunas (Shang et al., 2005). La proteína N tiene dos regiones altamente conservadas, el N-terminal o dominio de unión a ARN y el C-terminal o dominio de dimerización. Estos dos dominios se encuentran separados por una región intrínsecamente desordenada altamente fosforilada rica en serinas y argininas (motivo SR) (Ye et al., 2020). La estructura completa de la proteína N no ha sido dilucidada aún.

Proteína de Membrana (M).

La proteína de membrana, o proteína M, es una glicoproteína transmembranal de 222 aminoácidos y es la más abundante en la estructura del SARS CoV-2. La proteína M está organizada en tres dominios principales: N-terminal seguido de tres alfa hélices (TMH1-TMH3) y por último el C-terminal. La proteína M interactúa con otras de la misma clase, y con el resto de las proteínas estructurales, lo cual es esencial para el ensamblado de nuevos los nuevos viriones. Los residuos requeridos para interactuar con otras proteínas M se encuentran en el dominio de alfa hélice, mientras los residuos para formar interacciones con el resto de las proteínas se encuentran en el C-terminal (Ujike & Taguchi, 2015, Tseng et al., 2013). La estructura completa de la proteína M aún no ha sido determinada, pero se encuentra la estructura del virus HKU-5 que posee una similitud > 90 % con respecto al virus del SARS CoV-2. La estructura se encuentra en el PDB con un identificador 7Y9B, la



Figura 20.- Estructura de la proteína M obtenida mediante difracción de rayos X (PDBID 7Y9B

Proteína E.

La proteína de la envoltura, o proteína E, es una proteína con una longitud de 76-109 aminoácidos, de 12 kD, esencial para el virus SARS-CoV-2, ya que forma un canal catiónico homopentamérico simétrico que es importante para la patogenicidad del virus. La estructura de esta proteína en estado cerrado ha sido determinada mediante espectroscopia de RMN de estado sólido (Schoeman, D., et al., 2019). Está conformada por un dominio transmembranal (DTM) de cinco hélices (25 aminoacidos por cada DTM), que rodea un poro estrecho deshidratado, selectivo a cationes, a diferencia de otras proteínas homologas en especies virales distintas. Se ha encontrado que la actividad de la proteína es intracelular, específicamente se ha localizado en la superficie de la membrana del retículo endoplásmico (Nieto-Torres et al., 2011). Estudios de mutagénesis en la viroporina E, muestran que causan una gran reducción en la infectividad y letalidad (Nieto-Torres, J., et al., 2015). El péptido posee una carga neta de 0 Se han realizado diversos estudios in silico para determinar la conductividad iónica de esta estructura, además de encontrar el sitio activo de interacción farmacológica con las moléculas amantadina y la hexametilenamilorida, así como la generación de una conformación de canal abierto a través de simulaciones de dinámica molecular (Cao, Y., et al., 2020).

Se han identificado compuestos que bloquean viroporinas en otras especies virales, tal como la amantadina que bloquea la conducción del canal de la proteína M2 del virus de la influenza o la hexametilenamilorida que bloquea la viroporina E del virus SARS COV.(Scott

& Griffin, 2015). En la figura 21 se muestran las dos estructuras disponibles en el PDB de la viroporina E, la PDBID 5X29 de SARS CoV (Surya et al., 2018) y el segmento transmembranal de la viroporina de SARS CoV-2 (Mandala et al., 2020), ambas en estado cerrado. Ambas estructuras a pesar de su similitud se encuentran con una geometría distinta debido a que la estructura 5x29 fue obtenido mediante un sistema de expresión en E. coli, mientras que la estructura de 7K3G se encuentra en micelas de fosafitidilcolina, además de que esta estructura solo posee la región transmembranal de la proteína.



Figura 21.-. Actualmente se encuentran depositadas dos estructuras de la viroporina relacionadas con coronavirus. En el lado izquierdo (5x29) tenemos la viroporina del SARS CoV, mientras en el lado derecho (7K3G), tenemos la estructura del dominio transmembranal de la viroporina del SARS CoV-2.

Estructura y función de las viroporinas.

El ciclo celular de los virus está estrechamente relacionado con las membranas lipídicas. Durante el ciclo de infectividad un virus debe cruzar la membrana celular para posteriormente utilizar las membranas celulares para replicarse. Los virus utilizan, al igual que las células, el principio de compartimentalización lo cual le permite optimizar procesos. Una de las estrategias utilizadas por los virus, para aprovechar esta formación de compartimientos, es generar gradientes electroquímicos para acoplar la energía generada a las reacciones de la replicación viral. Para facilitar la formación de estos gradientes a partir del genoma viral, sintetizan proteínas que al polimerizarse en las membranas de los compartimentos intracelulares forman poros. Desde su descubrimiento en la década de los 70's, estas proteínas han sido catalogadas en la literatura como viroporinas (Fischer, W. B., et al., 2012). Las viroporinas están codificadas por diversas sepas virales de interés clínico tal como el virus de la hepatitis C (Griffin, S. D., et al., 2009), VIH-1 (Hout et al., 2004), coronavirus y el poliovirus, entre otros.

Las viroporinas son proteínas que forman poros acuosos y que se insertan en la membrana a través de un proceso de oligomerización. Las viroporinas se han clasificado en dos grupos y cuatro subgrupos de acuerdo con su topología. La clase IA contiene un dominio en su Nterminal que encuentra orientado hacia el lumen del compartimiento, mientras que la clase IB contiene un dominio corto y su N-terminal está dirigido hacia el citosol. Por otro lado, las clases II A y II B son segmentos transmembranales que se encuentran unidos por un asa, en donde los grupos N-terminal y C-terminal se dirigen hacia el lumen del compartimiento intracelular en el caso de IIA y en IIB se dirigen al citosol Figura 22 (Nieva, J., et al. 2012).

Las viroporinas comparten características en común tales como; a) son proteínas pequeñas de entre 50-120 aminoácidos y b) están compuestas por una gran proporción de aminoácidos hidrofóbicos. Algunas funciones de las viroporinas incluyen; la acidificación de compartimientos intracelulares (Martin & Heleniust, 1991), despolarización de la membrana del retículo endoplásmico para promover la gemación de las membranas (Agarkova et al., 2008), y la generación de fugas de las especies iónicas que promueven la activación de la respuesta apoptótica (Madan et al., 2007).



Figura 22 Clasificación de las viroporinas de acuerdo con su topología, en las clases IA y IB lo C- terminal y Nterminal se encuentran dirigidos en diferentes compartimentos, mientras que en el caso de la clase II de las viroporinas se dirigen hacia el mismo sitio (adaptada de Nieva, J., et al. 2012).

Las viroporinas se consideran blancos potenciales para el desarrollo de nuevas terapias antivirales debido a que su inhibición detiene la replicación viral, además de su diferencia estructural con respecto a los canales humanos. Algunos ejemplos de antivirales que tienen como blanco las viroporinas son la amantadina y la rimantadina, que inhiben la conducción del canal M2 del virus de la influenza (Wang, C., et al., 1993); la hexametilenamilorida que inhibe a la proteína Vpu del VIH-1 (Römer, W., et al., 2004), a la p7 del virus de la hepatitis y a la proteína E del SARS-CoV Wilson, L., et al., 2006).

Inhibidores actuales de la proteína E.

Los inhibidores conocidos actualmente son; los inhibidores clásicos de viroporinas amantadina, hexametilenamilorida (HMA) y rimantaina (Rim); los flavonoles quercetina (Que) y kaempferol (Kae); las flavonas apigenina (Api) y nobiletina (Nob); las isoflavonas genisteína (Gen) y naringerina(Nar); la flavonona epigalato de galocatequina (EGCG) así como 6-gingerol (6-Gin) y 8-gingerol (8-Gin), y el polifenol resveratrol (Res). Los compuestos fueron probados en células HEK-293 transfectadas (Human Embryo Kideny cells), en sé dónde empleó la técnica de patch-clamp para detectar corrientes iónicas (Breitinger et al., 2021). Durante el protocolo se encontró que al aplicar estas moléculas se obtuvo una inhibición de la corriente generada por la proteína E. En la figura 23 se muestran los inhibidores de la proteína E encontrados hasta el momento.



Figura 23 En esta figura se muestran los inhibidores clásicos de viroporinas y los inhibidores encontrados recientemente por Breitinger y colaboradores (Breitinger et al., 2021).

Diseño de fármacos asistido por computadora.

Los métodos de diseño de fármacos asistidos por computadora o CADD (Computer Aided Drug Design) son una herramienta que permite aumentar la productividad y efectividad en el proceso de descubrimiento de fármacos, permitiéndonos llevar a cabo el descubrimiento, el diseño y la optimización de compuestos para obtener las propiedades estructurales y fisicoquímicas, como cantidad de enlaces rotables, superficie polar, cantidad de grupos formadores de puente de hidrogeno y LogP, lo que impacta directamente sobre la absorción

y por lo tanto metabolismo de las moléculas (Doman et al., 2002). Uno de los métodos más revolucionarios y utilizados en el CADD es el cribado virtual (CV, virtual screening) que permite predecir la interacción entre una base de datos con miles o millones de moléculas, y uno o varios receptores, generando como resultado una lista de compuestos con alta afinidad por algún receptor. Diversas campañas en las que se han comparado los resultados entre CV y metodologías experimentales demuestran que el uso de CV disminuye la cantidad de falsos negativos producidos durante una prueba experimental, aumentando la tasa de aciertos (Li, Q., 2020).

El CV basado en estructura es un método en el cual se utiliza la información biológica de un blanco, el cual puede ser cualquier molécula biológica con actividad conocida, aunque la gran mayoría de estos estudios se realizan en proteínas. Este tipo de estudios está constituido por tres elementos fundamentales; 1) Las simulaciones de acople molecular, 2) las coordenadas atómicas en 3D de la estructura blanco y 3) una biblioteca de compuestos. Las herramientas para realizar el acople molecular toman en cuenta la estructura tridimensional de la proteína, y por medio de una estrategia de búsqueda se genera la mejor pose de interacción entre el ligando y el receptor, la cual es evaluada por una función de puntaje que nos indica la energía libre de interacción, la cual se define como la diferencia de la energía de una molécula entre un estado de interacción y uno completamente sin interacciones (Hata et al., 2021). La técnica de acople molecular se basa en la correspondencia geométrica y en principios fisicoquímicos. Un algoritmo típico de acople molecular utiliza una estrategia de búsqueda y le asigna un puntaje basado en una función de evaluación que mide una energía empírica de interacción, la cual refleja la afinidad de la molécula por el blanco biológico, implícitamente indicando de qué manera están interactuando ambas moléculas. El objetivo principal del CV es reducir el enorme espacio químico virtual disponible en las bases de datos químicos, a un número pequeño manejable con mayor potencia y selectividad para inhibir proteínas, y así poder obtener un candidato líder en el proceso de desarrollo de fármacos (Li, Q. 2020.).

Las bases de datos donde se depositan nuevos compuestos y sus estructuras como PubChem, proporcionan una amplia posibilidad de explorar opciones para encontrar nuevas moléculas que interactúen con diversos blancos biológicos. Además de las ya mencionadas, existen otros depósitos de moléculas, basados en los compuestos naturales de la medicina tradicional de diversas regiones, algunas de estas bases de datos son: NPACT (Mangal, M., et al., 2012), IMPPAT (Mohanraj, K., et al., 2018), Ambinter-Green-Pharma natural compound library (GPNCL), entre otros (Sorokina, M., et al., 2020).

Regla de los 5 de Lipinski.

En enero de 2001 Christopher Lipinski desarrolló un método para estimar la solubilidad y la permeabilidad de las moléculas, dado que, durante estudios de cribado molecular, no se podían determinar estos parámetros. Al realizarse la determinación de la IC50 muchas de estas moléculas exhibían una pobre solubilidad, lo cual dificulta la absorción de moléculas por vía oral. Para mejorar esta determinación se desarrollaron una serie de herramientas que ayudarían a predecir algunas propiedades fisicoquímicas de los compuestos, y que han ayudado a comprender la correlación entre dichas propiedades y la probabilidad de un compuesto a pasar con éxito o fracasar en las fases de ensayos clínicos I y II. Con el antecedente de que los compuestos que eran pobremente absorbidos serían los declinados en las siguientes fases, esto ayudó a encontrar cuales eran las características fisicoquímicas de los compuestos que sí contaban con buena biodisponibilidad oral. Para realizar este estudio se utilizó la quimioteca del Índice Mundial de Fármacos (WDI) para realizar el filtrado. Después de remover aquellas moléculas que eran de gran tamaño (nucleótidos, péptidos, polímeros) se obtuvo una lista de 2300 compuestos para su análisis que posteriormente fue nombrada como la quimioteca USAN. Las propiedades de estas moléculas fueron medidas y fue de aquí de donde se obtienen las propiedades moleculares requeridas para que una molécula o ligando pequeño tuviera buena biodisponibilidad oral (Lipinski et al., 2001). Las propiedades calculadas fueron: Peso Molecular, en donde previamente se había encontrado una correlación entre el alto peso molecular y la disminución de la permeabilidad a través del intestino y la barrera hematoencefálica (Navia, M., Chaturvedi, P. 1996). En la lista creada por Lipinski (USAN) se encontró que aquellos compuestos con la mejor biodisponibilidad oral eran aquellos que tenían un peso molecular < 500; Lipofobicidad, que se describe como la habilidad de una molécula de particionarse en octanol o agua, considerada una propiedad fisicoquímica altamente relevante para evaluar la absorción de una molécula. La lipofobicidad es expresada frecuentemente como el logaritmo del cociente de la partición de un fármaco dentro de un disolvente orgánico o fase acuosa (LogP) (Testa et al., 1996). En su investigación Lipinski encontró que la mayoría de los compuestos que habían pasado exitosamente las fases clínicas tenían un LogP< 5.; En adición al peso molecular y al LogP, se encontró también una relación entre la cantidad de hidrógenos formadores de puentes H y grupos aceptores para la formación de puentes H. Lipinski encontró que al realizar un conteo de puente N-H O-H era suficiente para describir esta característica, y en donde las moléculas con la mejor permeabilidad o biodisponibilidadoral tenían grupos donadores de H< 5 y grupos aceptores para la formación de puentes H< 10 (Lipinski et al., 2001).

Compuestos PAINS.

En 2014 Jonathan B. Baell realizó un cribado virtual con más de 100,000 compuestos, clasificándolos de acuerdo con el lugar donde habían sido adquiridos. Se incluyeron aquellos compuestos que no violaban las regla de los 5 de Lipinski y se eliminaron compuestos con más del 80% de similitud entre sí. Al realizar el cribado se encontró muchas moléculas con aparente afinidad y actividad por diferentes dianas biológicas, pero al reproducir los experimentos la actividad no fue detectada o se obtuvieron resultados no interpretables. El equipo de investigación notó que compuestos similares a los que ellos habían encontrado como problemáticos, presentaban comportamientos similares en otras investigaciones realizadas. Al condensar la información de todos estos quimotipos problemáticos se obtuvo como resultado la colección de compuestos de interferencia de ensayos o fragmentos PAINS (Pan-Assay Interference Compounds) (Figura. 24) En general los PAINS son compuestos que erróneamente son presentados como posibles «drogas» que pueden inhibir o activar proteínas diana. Algunos ejemplos de actividad de interferencia en los ensayos producidos por los compuestos PAINS son los ataques nucleófilos (reactividad), fotoreactividad, quelación de metales, actividad redox, formación de micelas o la atribución de propiedades fotocromáticas (coloración), los cuales pueden interferir con ensayos de adsorción de luz o fluorescencia. Es importante considerar estos fragmentos moleculares, y tener en cuenta que la quimioteca utilizada durante una investigación de cribado virtual ha sido debidamente curada y que los compuestos resultantes no presentaran problemas durante protocolos experimentales en laboratorios húmedos.



Figura 24.-En esta imagen se muestran los principales PAINS y el tipo de interferencia que producen durante el desarrollo de un ensayo (modoficado de Baell, 2013).

Cálculo de propiedades ADMET.

Las campañas de descubrimiento de fármacos vienen acompañadas de pasos críticos, en los que la determinación de las mejores propiedades fisicoquímicas en etapas tempranas es crucial para evitar que muchos o todos los compuestos sean rechazados en etapas avanzadas, propiedades farmacocinéticas y problemas de toxicidad son algunas de las causas que producen el fracaso de un compuesto en las fases clínicas de la investigación. (Landis MS. et al,2018). Para esto es necesario la aplicación de métodos que nos permita obtener en fases tempranas la información relacionada con la absorción, distribución, metabolismo, excreción y toxicidad. Estos resultados nos ayudan a minimizar el tiempo y costo de pruebas en la identificación de fuertes candidatos a fármacos. Las propiedades fisicoquímicas de una molécula tienen un importante efecto en su farmacocinética y conducta dentro del cuerpo. Algunas de las propiedades que se deben considerar es la solubilidad, el pKa, que se refiere al estado de protonación de los compuestos, este juega un importante papel en la solubilidad, permeabilidad y absorción; la permeabilidad, una característica físico química importante que tiene un efecto directo sobre la absorción. En recientes años se han hecho esfuerzos para diseñar algoritmos que permitan un cálculo preciso de la permeabilidad a través del tracto gastrointestinal y de la barrera hematoencefálica. Algunas herramientas aplicadas a este campo han sido el uso de modelos relación estructura actividad con el uso de múltiples descriptores químicos, regresión linear multivariada y las máquinas de vectores (N. N. Wang et al., 2016). Un punto importante es definir prematuramente los problemas relacionado con la toxicidad de una molécula. Ha sido reportado que aproximadamente 20-40% de los fármacos son eliminados en las fases de estudios preclínicos debido a las limitantes toxicológicas (Lu et al., 2017). Algunas herramientas como ProTox II tienen la capacidad de realizar predicciones de la actividad toxica de ciertas moléculas sobre puntos toxicológicos terminales como la carcinogenicidad, mutagenicidad, citotoxicidad, inmunotoxicidad entre otras. Los calculo se basan principalmente en incorporar información de moléculas toxicas y realizar un estudio de similitud, basado en un método de modelo de farmacóforo (Banerjee et al., 2018). Un farmacóforo es un arreglo de las características electronicas y estéricas de una molécula necesario para asegurar la formación optina de interacciones con un blanco biológico especifico (Seidel et al., 2020).

Acople Molecular.

El acople molecular es una técnica que ayuda a predecir la energía y sitios de interacción entre un complejo ligando receptor. El acople se puede realizar en dos pasos interrelacionados; 1) el muestreo de las conformaciones de un ligando en un sitio activo, 2) que serán evaluadas mediante una función de puntajes. Existen distintos algoritmos de búsqueda utilizados en diversos programas para calcular las interacciones entre un ligando y su receptor (Meng et al., 2011).

Existen diversos programas como Autodock Vina que utiliza un algoritmo de optimización de gradiente. Este consiste en la generación de conformaciones de un ligando que debe encontrar una geometría con baja energía a partir de una preexistente. Autodock Vina aplica los principios de mecánica molecular newtoniana en su algoritmo (Trott, O., & Olson, A. J., 2010). Otro programa popular de acople molecular es PLANTS en el que se utiliza un algoritmo de optimización de colonia de hormigas. Este algoritmo busca que la posición de los hidrógenos del ligando y el receptor sea optima. El algoritmo de PLANTS gira los enlaces rotables hasta encontrar la mejor solución entre la interacción de los átomos de hidrogeno (Stützle, Thomas, Exner, Thomas E. 2007). Por su cuenta Autodock 4, utiliza un algoritmo genético en el que genera múltiples conformaciones de ligandos, las evalúa dentro del receptor y solo conserva aquellas más "aptas" o con menor energía de interacción (Morris et al., 2009).

Acople Molecular por consenso.

Recientemente se ha desarrollado un nuevo enfoque del uso de las herramientas llamado acople molecular consenso, el cual se refiere al uso simultáneo de diferentes algoritmos de acople molecular. Éste se desarrolló al evaluar herramientas individualmente y calcular que las tasas de éxito para algunos algoritmos como Autodock Vina o Autodock tenían tasas de predicción bajas (64% y 58% respectivamente) en comparación con el porcentaje logrado con el uso simultáneo de diversos algoritmos (82%). En estudios se ha encontrado una relación entre el RMSD de las poses generadas por distintos algoritmos de acople molecular por consenso, y la reducción de la tasa de falsos positivos y falsos negativos al tratar de reproducir la pose de un ligando en complejo con un blanco biológico obtenido por métodos experimentales como la difracción de rayos X (Houston & Walkinshaw, 2013, Scardino et al., 2021). El RMSD o raíz de la desviación cuadrática, es la medida del promedio de las distancias entre los átomos de dos estructuras (con la misma cantidad de átomos) superpuestas y se define con la ecuación:

$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \delta_i^2}$$

Donde:

 $\delta^{2 e}$ es la diferencia entre las posiciones de dos átomos con el mismo índice elevada al cuadrado, y N es el número de átomos contados en el sistema.

El RMSD es comúnmente utilizado para estimar los grados de similitud estructural entre dos conformaciones de la misma molécula. Los valores del RMSD pueden ir desde 0 hasta infinito, en donde 0 significa mayor similitud molecular (Carugo & Pongor, 2008).

Moléculas Señuelo.

Las moléculas señuelo en los estudios de acople molecular se refiere a un grupo de ligandos que probablemente no tengan afinidad por un receptor. Estas moléculas pueden ser obtenidas a partir de moléculas activas conocidas para una diana biológica de interés utilizando servidores como DUDE-E (<u>http://dude.docking.org/).</u> El uso de señuelos en estudios de acople molecular es importante porque permite analizar como un algoritmo evalúa y discrimina una molécula activa de otra que no lo es (Graves et al., 2005).

Huellas digitales de las interacciones proteína-ligando.

Las huellas digitales de las interacciones proteína-ligando (PLIFS) es un método que convierte la información 3D contenida en un complejo en información 1D en forma de una cadena de código binario. En la construcción de cada bit contenido en las PLIFS, se utiliza un lenguaje Booleano en el que se asigna un 1 cuando existe una interacción entre un par de átomos, y un 0 en el caso contrario (Figura 25). Las huellas digitales de las interacciones proteína ligando son útiles ya que nos permite visualizar y analizar de una manera más sencillas los resultados de estudios como el acople y la dinámica moleculares.


Figura 25.- Esquema donde se muestra la manera en la que se crean los bits de las huellas digitales proteínaligando. Las huellas digitales facilitan la visualización y el análisis de las interacciones cuando se presentan varias moléculas al mismo tiempo (Deng et al., 2003).

Planteamiento del problema.

El descubrimiento de nuevos fármacos es un proceso que se realiza en un promedio de 20 años de investigación. Actualmente con el surgimiento de enfermedades emergentes con una alta tasa de morbilidad y mortalidad tal como la COVID-19, es necesario economizar recursos y tiempo para encontrar potenciales candidatos para un tratamiento. Actualmente la FDA a aprobado solo unos pocos medicamentos para el tratamiento de esta enfermedad, VEKLURY (remdesivir), el cual solo está recomendado para uso pediátrico, Nirmatrelvir con Ritonavi, el cual es eficaz solo si se toma dentro de los primeros cinco días de la infección, Bebtelovimab, de administración intravenosa, y Molnupiravir. (https://www.fda.gov/media/137566/download).

JUSTIFICACIÓN.

Debido a la rápida propagación y aumento en el número de muertes causadas por la COVID-19, es necesario utilizar herramientas que nos permitan encontrar rápidamente, y de manera precisa, candidatos potenciales para el desarrollo de fármacos o encontrar moléculas que ejerzan una actividad terapéutica para tratar la COVID-19. Los métodos computacionales han probado que pueden predecir de manera muy precisa los blancos farmacológicos, las moléculas y los sitios de interacción reduciendo de manera importante el trabajo de los laboratorios experimentales. Actualmente gracias a la disponibilidad de las coordenadas atómicas de la proteína E, la accesibilidad a bases de datos moleculares, y las asequibles herramientas de CADD podemos realizar la búsqueda de nuevos inhibidores.

HIPÓTESIS.

Existe una reducida cantidad de fármacos candidatos para tratar la COVID-19, y debido a la gran cantidad de moléculas disponibles, así como la gran diversidad de proteínas blanco del SARS CoV-2, dificultan la tarea de encontrar posibles soluciones. Sí exploramos quimiotecas novedosas en blancos moleculares poco estudiados como la proteína de la envoltura pequeña (E) de SARS-CoV-2, para la búsqueda de nuevas moléculas haciendo uso de técnicas computacionales de diseño de fármacos, es posible encontrar compuestos candidatos con actividad anti-SARS CoV-2.

METAS.

- I. Desarrollar un modelo de cribado virtual que permita encontrar potenciales moléculas inhibidoras de la viroporina E con precisión.
- II. Encontrar un conjunto de moléculas con alta afinidad y buenas propiedades ADMET que funcione como un inhibidor de la proteína E.

OBJETIVO PRINCIPAL.

I. Encontrar moléculas que posean un buen perfil fisicoquímico, toxicológico y farmacocinético, que formen interacciones energéticamente favorables con los residuos del dominio del poro de la proteína E para desarrollar fármacos bloqueadores de la corriente generada por la proteína E en el retículo endoplásmico, responsable de la generación de la membrana de los nuevos virus a partir de segmentos de la membrana del retículo endoplásmico del huésped.}

OBJETIVOS PARTICULARES.

- I. Plantear un protocolo para el filtrado de compuestos utilizando la técnica de acople molecular.
- II. Calcular las propiedades ADMET de las moléculas con mayor afinidad por el receptor obtenidas del acople molecular.
- III. Encontrar los patrones de interacción de las moléculas acopladas mediante huellas digitales de interacciones proteína-ligando (PLIFS) y determinar la similitud entre ellas.
- IV. Encontrar un consenso entre los resultados de los diferentes algoritmos al calcular el RMSD entre las poses de los compuestos acople s consenso entre los resultados de los diferentes algoritmos.
- V. Obtener los sitios más probables de interacción ligando receptor. Determinar la afinidad aparente de los ligandos por su receptor.

METODOLOGÍA.

Obtención de la estructura de la viroporina.

El archivo de las coordenadas atómicas de la estructura de la proteína E se obtuvo del material suplementario del trabajo publicado por Cao Y. (Cao Y., 2020), debido a la falta de estructuras disponibles en el PDB. La decisión de elegir esta estructura y no realizar el modelado de una nueva, reside en que, en los resultados presentados en esta investigación, muestran una estructura en estado semiabierto en la cual a través de técnicas de electrofisiología computacional se registró el flujo de iones.

Obtención de señuelos.

Los señuelos utilizados en el estudio fueron generados con el servidor DUDE-E (Mysinger et al., 2012). Se ingresaron los códigos SMILE de cada molécula para generar 50 moléculas señuelo por cada molécula activa.

Obtención y preparación de los ligandos.

Las moléculas que se utilizaron para realizar los experimentos de acople molecular se obtuvieron de PubChem (Kim et al., 2020) y de Asinex (www.asinex.com). Las moléculas

recabadas de PubChem corresponden a los flavonoides reportados como inhibidores de la proteína E, confirmados a través de estudios de registros electrofisiológicos. Las moléculas descargadas de Asinex pertenecen a la biblioteca de antivirales que se usaron en el estudio para encontrar nuevos inhibidores. La quimioteca de Asinex contiene un total de 6927 moléculas antivirales.

Los archivos descargados de las bases de datos poseían un formato SDF. Para obtener los archivos PDB a partir de dichos ficheros se utilizó Gypsum DL (Ropp et al., 2019). En el paso siguiente se utilizaron las herramientas encargadas de asignar las cargas, y realizar la conversión al formato correspondiente para cada programa de acople molecular. Para el caso de Autodock 4 y Autodock Vina se utilizó Prepare Ligand de Autodock Tools. Por otro lado, para preparar los archivos de entrada de PLANTS hidrogeno (Stützle, Thomas, Exner, Thomas E. 2007) se empleó SPORES (ten Brink & Exner, 2009). Las moléculas de la quimioteca de ASINEX así como los flavonoides se prepararon utilizando Gypsum DL con el cual se obtuvieron las estructuras protonadas de las proteínas en formato PDB. El programa está formado a partir de las bibliotecas de RDKIT para Python. Se Optimizaron las moléculas utilizando un campo de fuerza universal (UFF), y se protonaron a un pH de 7.4. Una vez que obtuvieron los archivos PDB de cada molécula, éstos se guardaron en carpetas destinadas para cada uno de los programas de acople molecular. Las moléculas utilizadas para realizar simulaciones de acople molecular con Autodock 4 y Autodock Vina fueron preparadas con Prepare Ligand de Autodock Tools. A los ligandos se les asignaron cargas tipo Gasteiger y se guardaron en formato PDBQT.

Los ligandos dirigidos a las simulaciones de acople molecular utilizando PLANTS fueron preparadas con SPORES. A los átomos se les asignó una carga haciendo uso del campo de fuerza Tripos, y posteriormente se almacenaron los archivos en formato MOL2.

Preparación del receptor.

Se usó OpenBabel (O'Boyle et al., 2011) para añadir los hidrógenos a un pH de 7.4. Posteriormente el receptor se preparó utilizando dos herramientas, para poder obtener los archivos con el formato correspondiente para cada programa de acople molecular. Para obtener el receptor con el formato adecuado para Autodock 4 y Autodock Vina se utilizó Prepare Receptor de Autodock tools. En el caso de PLANTS, se empleó el módulo de Chimera llamado Dock Prep (Pettersen et al., 2004). El receptor se preparó en dos pasos 1) adición de hidrógenos a un pH de 7.4 con Open Babel y 2) asignación de cargas y formato con Prepare Receptor y Dock Prep.

El receptor utilizado para las simulaciones de acople molecular con Autodock Vina y Autodock 4 fue preparado con Prepare Receptor. Los pasos de la preparación incluyeron la asignación de cargas Gasteriger y la eliminación de los hidrógenos no polares. Este receptor fue guardado en formato PDBQT.

El receptor utilizado para las simulaciones de acople molecular con PLANTS fue preparado con el módulo Dock Prep de Chimera. En este caso solo se le asignaron cargas Gasteiger y se guardó en formato MOL2.

Conversión de formatos.

La conversión de formatos requeridas en algunos pasos de la investigación se realizó utilizando Open Babel (O'Boyle et al., 2011).

Obtención de huellas digitales de interacción proteína-ligando y mapas de interacciones en dos dimensiones.

Las huellas digitales de las interacciones proteína-ligando y los mapas de interacciones en dos dimensiones fueron producidas utilizando ProLif (Bouysset & Fiorucci, 2021).

El análisis de las interacciones de los resultados obtenidos de los estudios de acople

molecular se realizó empleando ProLif (Bouysset & Fiorucci, 2021).. En primer lugar, se generaron las huellas digitales de las interacciones proteína-ligando de cada complejo. Para esto fue necesario utilizar SPORES para convertir todas las estructuras a formato MOL2 para obtener estructuras con todos los hidrógenos. Este paso fue necesario porque las estructuras en formato PDBQT resultantes de Autodock Vina y Autodock 4 solo conservan los hidrógenos polares. Una vez que se obtuvieron las estructuras en MOL2 se convirtieron a SDF utilizando OpenBabe (O'Boyle et al., 2011).l.

Las poses resultantes del cribado convertidas en formato SDF fueron utilizadas en ProLif para generar las huellas digitales de las interacciones proteína-ligando (PLIFS). Las figuras de los complejos fueron elaboradas usando PyMOL (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.)

Cálculo de la desviación estándar de la raíz cuadrática (RMSD).

El cálculo de RMSD de las poses obtenidas de cada ligando, se realizó utilizándolas herramientas RDKIT (<u>http://www.rdkit.org</u>). y MATPLOTLIB de Python. Los datos se ordenaron en forma de mapas de calor y dendógramas para poder visualizar el RMSD entre cada una de las poses y poder determinar si este parámetro era menor a 3 Å.

El cálculo de la RMSD se realizó utilizando RDKIT (http://www.rdkit.org).

Generación de las curvas de característica operativa del receptor (Curvas ROC). Los resultados obtenidos de las simulaciones de acople molecular en las que se utilizaron las moléculas "activas" y los señuelos permitieron describir la precisión de la discriminación de la predicción realizada por los algoritmos de acople molecular. El objetivo del cálculo de las curvas ROC fue establecer un punto de corte que determinó la sensibilidad y especificidad más alta de los puntajes de los programas de acople molecular asignado a cada ligando. Este punto es donde existe la mayor cantidad de verdaderos positivos (moléculas activas calificadas con un bajo puntaje) y la menor cantidad de falsos positivos (moléculas inactivas evaluadas con altos puntajes). Al aplicar este criterio al estudio de cribado virtual se consideraron solo aquellas moléculas que obtuvieron un puntaje menor que el del punto de corte establecido. Esto redujo la cantidad de falsos positivos obtenidos del cálculo de acople molecular.

A cada una de las curvas ROC se calculó el área bajo la curva (AUC) lo cual permite evaluar la funcionalidad del modelo. Los valores de AUC van de 0.5-1.0 en donde la herramienta con la mayor precisión es la que tiene un AUC=1 y la peor es AUC=0.5.

Las curvas ROC fueron generadas utilizando el servidor EASY-ROC (Goksuluk et al.,

2016) en donde se creó una lista para cada programa que contenía los valores de la mejor pose de cada molécula activa y cada molécula señuelo. A cada molécula "activa" se le asignó el índice 1 para indicar que eran los elementos control y el índice 0 a cada molécula señuelo para señalarlas como controles negativos.

Cálculo de propiedades ADMET.

Se realizó el cálculo de las propiedades ADMET utilizando los servidores Swiss-ADMET (<u>http://www.swissadme.ch/</u>) (Daina et al., 2017) y Pro-Tox-II (https://toxnew.charite.de/protox_II/) (Banerjee et al., 2018), ya que son herramientas de código abierto, además de ser ampliamente citadas en múltiples investigaciones. Al ser servidores web nos facilita la terea de resolver problemas de incompatibilidad con sistemas operativos.

Simulaciones de acople molecular.

Para realizar las simulaciones de acople molecular se utilizaron tres programas con algoritmos de búsqueda distintos. Los algoritmos Autodock Vina (Trott, O., & Olson, A. J., 2010), Autodock 4 (Morris et al., 2009) y PLANTS (Stützle, Thomas, Exner, Thomas E. 2007). La configuración de las simulaciones de acople molecular fueron receptor rígido y ligando flexible

Para realizar las simulaciones de acople molecular en Autodock Vina se empleó un programa con interfaz gráfica para apoyar la visualización de los sitios de interés para realizar el acople. En este caso se utilizó la interfaz gráfica de Autodock Tools.

Una vez cargada la proteína en el programa, es necesario determinar los sitios de la proteína que son de interés para evaluar sí alguna molécula puede unirse en ese espacio. La selección de dicho sitio de estudio se realiza introduciendo una caja que deberá contener los átomos de los residuos de interés. Se realizaron dos estudios con objetivos distintos. El primero fue un acople ciego, el cual se realizó colocando toda la proteína dentro de la caja de acople. Este estudio se realizó con la biblioteca de Flavonoides, para poder encontrar de qué manera interactúan los compuestos "activos" o "control" con la proteína, y así poder asignar un punto de referencia y compararlo con el acople de la biblioteca de antivirales. Para obtener las poses más consistentes, se realizaron 10 simulaciones de acople molecular a cada uno de los ligandos con los tres programas mencionados anteriormente.

Una vez acopladas las moléculas control, se realizó lo mismo con las moléculas señuelo obtenidas de DUDE-E. Los objetivos de este experimento fueron: 1) encontrar el sitio de convergencia de las moléculas, y 2) obtener y analizar como los algoritmos de búsqueda

de los programas de acople molecular calificaban, con sus respectivas funciones de puntajes, las moléculas activas de los señuelos, y poder determinar si tenían la capacidad de discriminar entre ambas.

Las poses que se seleccionaron para realizar el análisis fueron, la mejor pose de cada ligando generada en cada uno de los experimentos de acople molecular.

El segundo estudio realizado, fue una evaluación de la función de acople molecular en un espacio más reducido. Este se realizó en el espacio en donde convergieron los ligandos utilizando la biblioteca de compuestos antivirales.

Obtención de poses consenso.

La obtención de los compuestos y poses finales fue un proceso de múltiples pasos. El primer paso fue descartar todas aquellas moléculas que obtuvieron un puntaje de interacción por debajo de los puntos de corte calculados a partir de las curvas ROC. Este proceso ayudó a descartar alrededor de 1500 compuestos de cada algoritmo, ya que al ser evaluados con puntajes elevados existía una mayor probabilidad de que estos representaran falsos positivos.

El segundo paso fue encontrar aquellos compuestos que generaron consenso, es decir, que los tres algoritmos de búsqueda hayan coincidido en la misma molécula calificada como un compuesto activo. A partir de los puntajes obtenidos de cada algoritmo se calculó un puntaje consenso utilizando la ecuación 1:

Ecuación 1:

((Puntaje de molécula del algoritmo 1/puntaje más bajo obtenido por algoritmo 1) + (Puntaje de molécula del algoritmo 2/puntaje más bajo obtenido por algoritmo 2) + (Puntaje de molécula del algoritmo 3/puntaje más bajo obtenido por algoritmo 3))

número de algoritmos utilizados.

Después de realizar el cálculo consenso de las tres poses (una por cada algoritmo de búsqueda), se seleccionaron las diez moléculas con los puntajes más altos. Estas moléculas se sometieron a un análisis de RMSD y al análisis de similitud de huellas digitales de interacciones proteína-ligando. Para realizar el análisis de similitud de las PLIFS se utilizó el índice de Tanimoto:

Tc (A, B)) |A ∩ B|/|A ∪ B|

Donde $|A \cap B|$ representa el número de bits en común entre dos conjuntos y $|A \cup B|$ es el número de bits diferentes entre los conjuntos.

Cálculo de propiedades ADMET.

Para evaluar las propiedades ADMET de los compuestos se usaron los servidores web SwissADME y Pro-Tox-II. Swiss-ADME es un servidor que evalúa las propiedades fisicoquímicas de las moléculas de acuerdo con su estructura y genera un cálculo sobre la permeabilidad de la molécula a través de la barrera hematoencefálica, y de la biodisponibilidad del fármaco a través de una administración oral. El resultado es una gráfica de radar (Figura 41 y Figura 42) por cada molécula en donde se puede apreciar a simple vista el perfil tipo fármaco de una molécula. El área rosa representa el intervalo óptimo para cada propiedad (lipofobicidad: XLOGP de -0.7 a +5.0, tamaño: Peso molecular de 150 a 500 g/mol, polaridad: TPSA de 20 a 130 Å, solubilidad: log S no mayor a 6, saturación: fracción de carbonos con hibridación sp3 no menor a 0.25 y flexibilidad: no más de 9 enlaces rotables), y una gráfica de "huevo cocido" en donde la región blanca representa alta probabilidad de absorción pasiva a través del tracto gastrointestinal, y la amarilla representa alta probabilidad de cruzar la barrera hematoencefálica. En adición, los puntos azules predicen si la molécula puede ser expulsada del citoplasma celular activamente a través de la glicoproteína P y los rojos sí no es sustrato para la glicoproteína P.

Por otro lado, Pro-Tox-II es un servidor que utiliza el código SMILES como entrada de datos. Este código SMILES es comparado con la base de Toxicóforos del servidor a través de un modelo de QSAR para determinar la probabilidad de que la molécula evaluada posea una actividad toxica. Adicionalmente Pro-Tox-II nos proporciona una DL50 basándose en la similitud de la molécula con otras halladas en bases de datos

como PubChem. Los resultados de Pro-Tox-II son mostrados en una gráfica de radar que contiene una escala porcentual que indica el nivel de toxicidad de las moléculas. En esta grafica se aprecia una sombra naranja para los valores de las moléculas que son parte del modelo, y en azul el resultado del porcentaje de toxicidad de la molécula de estudio.

Resultados y discusión.

Resultado del acople molecular de señuelos.

El acople molecular de las moléculas señuelo se realizó en el sitio de mayor frecuencia de interacción de las moléculas activas. En este caso no se consideraron las interacciones formadas por las moléculas señuelo, ya que el objetivo fue evaluar la puntuación que cada algoritmo le asigna a una molécula "activa" y a un señuelo; y de esta manera poder evaluar si el algoritmo es capaz de discriminar las moléculas activas de las inactivas reduciendo la cantidad de falsos positivos durante el estudio de cribado virtual. Los puntajes considerados para cada molécula activa se muestran en la Figura 26, que corresponden a las mejores poses generadas por cada algoritmo de acople molecular.



Figura 26.- Los resultados del acople molecular ciego muestran una convergencia de las moléculas dentro del poro de la proteína, lo que sugiere que el sito más probable de interacción de la proteína se localice ahí.

Resultado del acople molecular ciego y análisis de interacciones de las moléculas activas convergentes.

Mediante una inspección visual se pudo determinar que el sitio en donde hubo una mayor convergencia de las moléculas activas encontradas por Breitinger y colaboradores (Breitinger et al., 2021), utilizadas como control positivo, fue en el poro del canal (figura 26). Esto nos condujo a descartar el resto de las poses que no interactuaban en la región del poro y dejar a un lado otros probables sitios de interacción, para así realizar la selección y el análisis de las interacciones formadas por cada una de las moléculas ahí localizadas.

Las poses de las moléculas que se utilizaron para9oioi el análisis de interacciones fueron seleccionadas utilizando dos criterios: 1) la energía de unión, y 2) el RMSD.

Los algoritmos de búsqueda utilizados para el estudio no convergieron para encontrar poses similares con RMSD menores a 3 Å. El RMSD encontrado entre cada pose fue de entre 5-10 Å. A pesar de la diferencia entre la geometría de las poses, los patrones de interacción fueron similares entre Autodock Vina, PLANTS y Autodock 4, para los estudios de acople molecular (Figura 26). En los mapas de interacción se puede apreciar que los residuos que interactúan con los ligandos son similares, en donde la presencia de los residuos es constante. En las interacciones se puede apreciar una alta influencia de la Phe 19 localizada en la parte media del dominio del poro para formar interacciones Pi-Pi, la Ala 15 para formar interacciones hidrofóbicas y la Thr 23 para generar interacciones tipo puente de hidrogeno. Otro punto importante por resaltar es que los algoritmos de Autodock Vina, PLANTS y Autodock 4, no lograron encontrar poses para algunas moléculas en la región del poro, por esa razón algunas moléculas solo están representadas por una sola pose generada por un algoritmo de búsqueda, tal como lo son la gliclazida (PLANTS) y el resveratrol (Autodock Vina). En la mayoría de los casos, al menos dos algoritmos lograron identificar a una molécula dentro del poro como molécula activa. El objetivo del acople ciego fue encontrar los puntajes que cada algoritmo asignó a cada molécula como "activa" y poder generar una referencia para los puntos de corte energéticos. En la figura 27 se muestra el complejo formado por las poses del compuesto 6-Gingerol en el dominio del poro de la proteína E, en la figura se muestran las interacciones formadas por las diferentes poses del compuesto y las cadenas laterales de la que forman el lumen del poro de la protepina en donde podemos obsevar que existe una mayor cantidad de interacciones hidrofóbicas, principalmente con los residuos Phe 16, Phe 19 y Val 15.



Figura 27.- En esta figura se muestran las mejores poses de la molécula 6-Gingerol, reportada como activa por Breitinger y colaboradores (Breitinger et al., 2021), obtenidas con cada programa de acople molecular (Autodock Vina, PLANTS, ATD4). En el lado izquierdo se pueden apreciar el modo de unión y la geometría adoptada por la molécula. En el lado derecho se esquematizan las interacciones formadas entre la proteína y el ligando en un mapa. En le mapa de interacciones se indica el nombre del residuo, el número de identificador dentro de la secuencia y la letra de la cadena a la que pertenece. Los cuadros de colores azul, magenta y amarillo indican el algoritmo que produjo la pose en el complejo.

De la misma manera se calcularon las interacciones del compuesto activo 8-Gingerol, para el cual se encontraron interacciones principalmente alifáticas y π . En la figura 28 se muestran las poses generadas por los tres algoritmos y que las poses interactúan principalmente con Phe 16 y Phe 19 formando interacciones tipo π del homopentaméro.



Figura 28.- En esta figura se muestran las mejores poses de la molécula 8-Gingerol, reportada como activa por Breitinger y colaboradores (Breitinger et al., 2021), obtenidas con cada programa de acople molecular (Autodock Vina, PLANTS, ATD4). En el lado izquierdo se pueden apreciar el modo de unión y la geometría adoptada por la molécula. En el lado derecho se esquematizan las interacciones formadas entre la proteína y el ligando en un mapa. En el mapa de interacciones se indica el nombre del residuo, el número de identificador dentro de la secuencia y la letra de la cadena a la que pertenece. Los cuadros de colores azul, magenta y amarillo indican el algoritmo que produjo la pose en el complejo.

Utilizando el mismo procedimiento se calcularon las interacciones de la molécula Apigenina, otra de las moléculas reportadas como activas en la investigación de Breitinger y colaboradores (Breitinger et al., 2021). Al igual que las moléculas revisadas anteriormente, los resultados de la simulación de acople molecular indican que el sitio más probable de interacción es la región del poro, en donde forma interacciones de tipo π y alifáticas principalmente. Las interacciones son formadas principalmente con los residuos Ala15, Phe 19 y Phe 16, que son residuos que se localizan en la región intermedia del poro. En la figura 29 se muestran las poses y las interacciones que forma con los residuos del dominio del poro de la proteína.



Figura 29.- En esta figura se muestran las mejores poses de la molécula Apigenina, reportada como activa por Breitinger y colaboradores (Breitinger et al., 2021), obtenidas con cada programa de acople molecular (Autodock Vina, PLANTS, ATD4). En el lado izquierdo se pueden apreciar el modo de unión y la geometría adoptada por la molécula. En el lado derecho se esquematizan las interacciones formadas entre la proteína y el ligando en un mapa. En el mapa de interacciones se indica el nombre del residuo, el número de identificador dentro de la secuencia y la letra de la cadena a la que pertenece. Los cuadros de colores azul, magenta y amarillo indican el algoritmo que produjo la pose en el complejo.

El siguiente es el caso de la molécula Kaempferol, que al igual que las anteriores fue reportada experimentalmente en la misma investigación. Las interacciones formadas por las distintas poses obtenidas durante las simulaciones de acople molecular muestran un patrón de interacciones similar al de las moléculas antes mencionadas. En la figura 30 se muestra el complejo de las poses calculadas para el compuesto Kaempferol, así como las interacciones que se formaron entre la molécula y la proteína E. Se observa que las poses del compuesto se encuentran localizadas en el dominio del poro y forman interacciones alifáticas y tipo principalmente con los residuos Phe 16, Phe 19 y Val 22.



Figura 30.- En esta figura se muestran las mejores poses de la molécula Kaempferol, reportada como activa por Breitinger y colaboradores (Breitinger et al., 2021), obtenidas con cada programa de acople molecular (Autodock Vina, PLANTS, ATD4). En el lado izquierdo se pueden apreciar el modo de unión y la geometría adoptada por la molécula. En el lado derecho se esquematizan las interacciones formadas entre la proteína y el ligando en un mapa. En el mapa de interacciones se indica el nombre del residuo, el número de identificador dentro de la secuencia y la letra de la cadena a la que pertenece. Los cuadros de colores azul, magenta y amarillo indican el algoritmo que produjo la pose en el complejo.

De la misma manera se calcularon las interacciones del compuesto activo EGCG, para el cual se encontraron interacciones principalmente alifáticas y π . En la figura 31 se muestran las poses generadas por los tres algoritmos y que las poses interactúan principalmente con Phe 16, Phe 19 y Leu 12 formando interacciones tipo π y alifáticas con los residuos del homopentaméro.



Figura 31.- En esta figura se muestran las mejores poses de la molécula Kaempferol, reportada como activa por Breitinger y colaboradores (Breitinger et al., 2021), obtenidas con cada programa de acople molecular (Autodock Vina, PLANTS, ATD4). En el lado izquierdo se pueden apreciar el modo de unión y la geometría adoptada por la molécula. En el lado derecho se esquematizan las interacciones formadas entre la proteína y el ligando en un mapa. En el mapa de interacciones se indica el nombre del residuo, el número de identificador dentro de la secuencia y la letra de la cadena a la que pertenece. Los cuadros de colores azul, magenta y amarillo indican el algoritmo que produjo la pose en el complejo.

La siguiente es la molécula Genisteina, que al igual que las anteriores fue reportada experimentalmente. Las interacciones formadas por las distintas poses obtenidas durante las simulaciones de acople molecular muestran un patrón de interacciones similar al de las moléculas antes mencionadas. En la figura 32 se muestra el complejo de las poses calculadas para el compuesto Genisteina, así como las interacciones que se formaron entre la molécula y la proteína E. Se observa que las poses del compuesto se encuentran localizadas en el dominio del poro y forman interacciones alifáticas y tipo principalmente con los residuos Phe 16, Phe 19 y Leu 12.



Figura 32.- En esta figura se muestran las mejores poses de la molécula Genisteina, reportada como activa por Breitinger y colaboradores (Breitinger et al., 2021), obtenidas con cada programa de acople molecular (Autodock Vina, PLANTS, ATD4). En el lado izquierdo se pueden apreciar el modo de unión y la geometría adoptada por la molécula. En el lado derecho se esquematizan las interacciones formadas entre la proteína y el ligando en un mapa. En el mapa de interacciones se indica el nombre del residuo, el número de identificador dentro de la secuencia y la letra de la cadena a la que pertenece. Los cuadros de colores azul, magenta y amarillo indican el algoritmo que produjo la pose en el complejo.

Repitiendo los pasos anteriores se calcularon las interacciones de la molécula Gliclazida, otra de las moléculas reportadas como activas en la investigación de Breitinger y colaboradores (Breitinger et al., 2021). Al igual que las moléculas anteriores, los resultados de la simulación de acople molecular encontraron que el sitio más probable de interacción es la región del poro, en donde forma interacciones de tipo π y alifáticas principalmente. Las interacciones son formadas principalmente con los residuos Ala 15, Phe 19 y Phe 16, que son residuos que se localizan en la región intermedia del poro. En la figura 33 se muestran las poses y las interacciones que forma con los residuos del dominio del poro de la proteína.



Figura 33.- En esta figura se muestran las mejores poses de la molécula Glilclazida, reportada como activa por Breitinger y colaboradores (Breitinger et al., 2021), obtenidas con cada programa de acople molecular (Autodock Vina, PLANTS, ATD4). En el lado izquierdo se pueden apreciar el modo de unión y la geometría adoptada por la molécula. En el lado derecho se esquematizan las interacciones formadas entre la proteína y el ligando en un mapa. En el mapa de interacciones se indica el nombre del residuo, el número de identificador dentro de la secuencia y la letra de la cadena a la que pertenece. Los cuadros de colores azul, magenta y amarillo indican el algoritmo que produjo la pose en el complejo.

De la misma manera se calcularon las interacciones del compuesto activo Hexametileno de amilorida, para el cual se encontraron interacciones principalmente alifáticas y π . En la figura 34 se muestran las poses generadas por los tres algoritmos y que las poses interactúan principalmente con. Las interacciones son formadas principalmente con los residuos Ala 15, Phe 19 y Phe 16, que son residuos que se localizan en la región intermedia del poro. En la figura 33 se muestran las poses y las interacciones que forma con los residuos del dominio del poro de la proteína.



Figura 34.- En esta figura se muestran las mejores poses de la molécula Hexametileno de amilorida, reportada como activa por Breitinger y colaboradores (Breitinger et al., 2021), obtenidas con cada programa de acople molecular (Autodock Vina, PLANTS, ATD4). En el lado izquierdo se pueden apreciar el modo de unión y la geometría adoptada por la molécula. En el lado derecho se esquematizan las interacciones formadas entre la proteína y el ligando en un mapa. En el mapa de interacciones se indica el nombre del residuo, el número de identificador dentro de la secuencia y la letra de la cadena a la que pertenece. Los cuadros de colores azul, magenta y amarillo indican el algoritmo que produjo la pose en el complejo.

La siguiente la molécula Naringerina, reportada experimentalmente en la investigación de Breitinger y colaboradores (Breitinger et al., 2021, de la misma manera que los ligandos anteriores. Las interacciones formadas por las distintas poses obtenidas durante las simulaciones de acople molecular muestran un patrón de interacciones similar al de las moléculas antes mencionadas. En la figura 35 se muestra el complejo de las poses calculadas para el ligando Naringerina, así como las interacciones que se formaron entre la molécula y la proteína E. Se observa que las poses del compuesto se encuentran localizadas en el dominio del poro y forman interacciones alifáticas y tipo principalmente con los residuos Phe 16, Phe 19 y Leu 12.



Figura 35.- En esta figura se muestran las mejores poses de la molécula Naringerina, reportada como activa por Breitinger y colaboradores (Breitinger et al., 2021), obtenidas con cada programa de acople molecular (Autodock Vina, PLANTS, ATD4). En el lado izquierdo se pueden apreciar el modo de unión y la geometría adoptada por la molécula. En el lado derecho se esquematizan las interacciones formadas entre la proteína y el ligando en un mapa. En el mapa de interacciones se indica el nombre del residuo, el número de identificador dentro de la secuencia y la letra de la cadena a la que pertenece. Los cuadros de colores azul, magenta y amarillo indican el algoritmo que produjo la pose en el complejo.

Aplicando el mismo flujo de trabajo, se calcularon las interacciones de la molécula Nobiletina, una de las moléculas reportadas como activas en la investigación de Breitinger y colaboradores (Breitinger et al., 2021). Los resultados de la simulación de acople molecular encontraron que el sitio con mayor afinidad es la región del poro, en donde forma interacciones de tipo π y alifáticas principalmente. Las interacciones son predominantemente hidrofóbicas, del tipo alifáticas y π , principalmente con los residuos Ala 15, Phe 19 y Phe 16, que son residuos que se localizan en la región intermedia del poro. En la figura 36 se muestran las poses y las interacciones que forma con los residuos del dominio del poro de la proteína.



Figura 36.- En esta figura se muestran las mejores poses de la molécula Naringerina, reportada como activa por Breitinger y colaboradores (Breitinger et al., 2021), obtenidas con cada programa de acople molecular (Autodock Vina, PLANTS, ATD4). En el lado izquierdo se pueden apreciar el modo de unión y la geometría adoptada por la molécula. En el lado derecho se esquematizan las interacciones formadas entre la proteína y el ligando en un mapa. En el mapa de interacciones se indica el nombre del residuo, el número de identificador dentro de la secuencia y la letra de la cadena a la que pertenece. Los cuadros de colores azul, magenta y amarillo indican el algoritmo que produjo la pose en el complejo.

Resultados curvas ROC.

Los datos de los puntajes de cada molécula "activa" y señuelo permitió calcular las curvas ROC para evaluar la funcionalidad de cada algoritmo de acople molecular. La Figura 37 contiene una curva para los modelos predichos por Autodock 4, Autodock Vina y PLANTS con sus respectivos valores de AUC. Estos modelos constan de la proteína y tres poses para cada señuelo y para cada molécula activa correspondientes a las moléculas reportadas por (Breitinger et al., 2021).



Figura 37.- En esta figura se presentan las curvas ROC de cada algoritmo, en donde los valores del área bajo la curva permiten apreciar cual es el algoritmo con mayor especificidad y sensibilidad, lo que se traduce en mayor precisión

Los resultados del área bajo la curva asignados a cada algoritmo reflejan la sensibilidad que poseen para diferenciar entre moléculas activas e inactivas, siendo todas >0.5, lo que significa que todas tienen la capacidad de discriminar los señuelos de los activos.

De acuerdo con los datos obtenidos de las curvas ROC, se encontraron los valores de los puntos de corte para cada algoritmo; PLANTS =-78.71 CHEMPLP, Autodock Vina= -6.7 kcal/mol y Autodock 4= -5.8 kcal/mol. Como se puede observar en la Figura 27 el área bajo la curva con mayor valor lo obtuvo Autodock 4, mientras que el peor lo obtuvo PLANTS, lo que quiere decir que las predicciones realizadas por Autodock 4 son más precisas.

Resultado del cribado virtual.

Las moléculas encontradas, como resultado del estudio de cribado virtual, satisficieron todos los requisitos antes mencionados. Las poses calculadas por los diferentes algoritmos de acople molecular son consistentes en los patrones de interacción y una RMSD entre cada una de las poses < 3 Å. El cálculo de los puntajes consenso a partir de la ecuación 1, nos arrojó un valor de 0.98 para la mejor evaluada, y un valor de 0.79 para la peor. Los 10 compuestos elegidos como los mejores se seleccionaron a partir de la combinación de un buen puntaje consenso y una RMSD < 3 Å. El puntaje consenso calculado para cada compuesto fue: 0.91±0.6 (BDB 26417644), 0.87±0.08 (BDB 26420214), 0.86±0.04 (BDB 26422058), 0.85±0.05 (BDC 23168933), 0.91±0.05 (BDC 2369243), 0.98±0.01 (BDC 23205600), 0.85 ±0.07(BDG 34062417) v 0.92±0.07 (BDG 34121287). En las Figuras siguientes se muestran las escenas del modo de interacción de cada ligando acompañado de su mapa de interacciones en 2D y el puntaje de manera individual asignado por cada programa de acople molecular utilizado durante el estudio. Cabe resaltar que al igual que las moléculas "activas" previamente acopladas, los aminoácidos que forman las interacciones son los mismos, en donde se ve una gran participación de la Phe 19 de cada segmento transmembranal. En la parte inferior de cada complejo ligando-proteína se muestran los patrones de interacción de cada pose. El objetivo de esta sección es mostrar la convergencia de los tres algoritmos en las poses de los diferentes compuestos a través de la similitud de sus patrones de interacción y el RMSD entre ellas.

En la figura 28 se muestra el modo de interacción de la molécula BDB 26417644 y la viroporina E. Las poses calculadas por los algoritmos de acople molecular Autodock Vina, Autodock 4, y PLANTS convergieron dentro del poro del canal con una orientación similar y un valor de RMSD bajo. En la figura 38 se pude apreciar la escena de la proteína con los ligandos en la región del poro y formando interacciones similares. En los mapas de interacción (Figura 39) podemos observar el tipo de interacciones formadas y lo aminoácidos responsables de la formación de estas. Se puede observar un gran número de interacciones hidrofóbicas con residuos alifáticos como Val 22 o Ala 15, y en segundo lugar una considerable presencia de interacciones tipo π , debido a la naturaleza hidrófoba de la estructura del poro y el núcleo de la estructura del compuesto. Más adelante en la imagen 40, que corresponde a los PLIFs calculados para cada una de las poses, se puede observar una considerable conservación de los patrones de interacción, siendo los residuos Phe 16 C y 19 de las cadenas C, D y E los más constantes entre las tres diferentes poses calculadas. Para comparar y calcular la similitud de los patrones de interacción calculados, se graficó un mapa de calor (figura 41), que muestra los índices de Tanimoto, que indica la similitud entre dos conjuntos de datos en cada unidad del mapa.

Estos índices ordenados en forma de matriz indican la relación entre cada patrón de interacciones. Como se puede apreciar el índice de similitud indica que los patrones con mayores coincidencias son las poses generadas por lo algoritmos de PLANTS y Autodock 4 con un valor de 0.88, sugiriendo una alta conservación de interacciones. Estos resultados son complementados con el mapa de calor que muestra el valor de la RMSD de los compuestos ordenados en pares (figura 38) Los valores de RMSD van desde 1.5 Å (Autodock Vina-Autodock 4) hasta 3 Å (PLANTS-Autodock Vina). Con los valores en altos de los índices de similitud y los bajos valores de RMSD, sugieren una alta precisión en el cálculo de las interacciones para esta molécula.



Figura 38.- En esta figura se muestra en la parte superior la virporina E en una vista lateral, en donde en el centro se puede observar las poses del ligando BDB 26417644. En la parte inferior se muestra un acercamiento a la región interna del poro en donde se encuentra el ligando, y todas las cadenas laterales envueltas en la formación de interacciones (Naranja). En esta imagen se encuentra anotado el número del residuo, el nombre en su código de tres letras y el identificador de la cadena del homopentaméro donde se encuentra localizado el residuo.



Figura 39.- En esta figura se muestran los mapas de interacciones del ligando BDB 26417644. Las etiquetas de interacciones (esquina inferior izquierda) indica el tipo de residuo e interacción presentes en el complejo ligandoproteína. Junto a cada mapa de interacción se encuentra anotado el algoritmo que calculó esa pose, el valor de la función de puntaje, y el nombre del compuesto. En esta imagen se encuentra anotado el número del residuo, el nombre en su código de tres letras y el identificador de la cadena del homopentaméro donde se encuentra localizado el residuo.



Huellas digitales de las intereacciones ligando-proteína del complejo BDB 26417644-Viroporina E.

Figura 40.- En esta figura se exponen los resultados de las huellas digitales de las interacciones ligando-proteína del ligando BDB 26417644 y los residuos de la viroporina E. Las etiquetas en la parte superior corresponden al residuo involucrado en la interacción con cada pose. Estas indican el tipo de residuo el número del residuo, el nombre en su código de tres letras y el identificador de la cadena del homopentaméro donde se encuentra localizado el residuo. Las etiquetas del lado izquierdo indican el nombre del algoritmo que calculó la pose del ligando BDB 26417644. En el centro se encuentran marcas que indican el tipo de interacción formada entre la pose del ligando BDB 26417644 y la proteína. La información de los tipos de interacciones se encuentra en la lista localizada en la parte derecha de la gráfica.



Figura 41.- En este mapa de calor se muestra la similitud entre las huellas digitales de interacción de las diferentes poses del ligando BDB 26417644. Las etiquetas contienen el nombre de los algoritmos y los cuadros al centro muestran la relación de similitud entre los patrones de interacciones calculadas por cada algoritmo. En la derecha se encuentra la escala de los colores que permiten identificar el valor de la relación entre cada par del mapa.



Figura 42.- Esta figura muestra un mapa de calor que nos ayuda a visualizar el valor del RMSD entre cada una de las poses calculadas durante las simulaciones de acople molecular del ligando BDB 26417644. La escala de colores tiene asignado un valor numérico, el cual se muestra en la escala de la derecha. El dendograma de la izquierda es para identificar con mayor facilidad los objetos que guardan mayor relación al tener valores de RMSD menores entre sus átomos.

Las poses del compuesto BDB 26410214, generadas por los algoritmos de

acople molecular utilizados en este estudio, convergieron al igual que todas las moléculas anteriores, dentro del poro, ocupando el mismo sitio (figura 43). En los mapas de interacciones (figura 44) se observa, que las interacciones y residuos que imperan son hidrofóbicos, del tipo alifático y aromáticos. Al igual que las moléculas previamente mostradas, los residuos predominantes corresponden a la Phe 19 de las cadenas D y E, con las cuales se forman interacciones tipo π . Por otro lado, las interacciones hidrofóbicas corresponden a los residuos Leu 12 E, Ala 15 E y el residuo Thr 23 de la cadena E principalmente. En la figura 45 se muestran los PLIFs calculados en donde se presentan el tipo y numero de residuo, y el tipo de interacción. La similitud de estos patrones se expone en el mapa de calor de la figura 46, en donde los índices de similitud van desde 0.75 entre las poses calculadas por PLANTS y Autodock 4, a 0.93 entre las poses generadas por Autodock 4 y Autodock Vina. De forma complementaria en la figura 29 E, se muestran los valores de RMSD entre las diferentes poses, donde los valores se encuentran entre el intervalo de 0.8 Å (PLANTS-Autodock Vina) a 2.0 Å (PLANTS- Autodock Vina), con lo cual podemos suponer que la predicción del modo de interacción de esta molécula posee gran precisión.



Figura 43.- En esta figura se muestran los mapas de interacciones del ligando BDB 26410214. Las etiquetas de interacciones (esquina inferior izquierda) indica el tipo de residuo e interacción presentes en el complejo ligandoproteína. Junto a cada mapa de interacción se encuentra anotado el algoritmo que calculó esa pose, el valor de la función de puntaje, y el nombre del compuesto. En esta imagen se encuentra anotado el número del residuo, el nombre en su código de tres letras y el identificador de la cadena del homopentaméro donde se encuentra localizado el residuo.



Figura 44.- En esta figura se muestran los mapas de interacciones del ligando BDB 26410214. Las etiquetas de interacciones (esquina inferior izquierda) indica el tipo de residuo e interacción presentes en el complejo ligandoproteína. Junto a cada mapa de interacción se encuentra anotado el algoritmo que calculó esa pose, el valor de la función de puntaje, y el nombre del compuesto. En esta imagen se encuentra anotado el número del residuo, el nombre en su código de tres letras y el identificador de la cadena del homopentaméro donde se encuentra localizado el residuo.



Huellas digitales de las intereacciones ligando-proteína del complejo BDB 26420214-Viroporina E.

Figura 45.- En esta figura se exponen los resultados de las huellas digitales de las interacciones ligando-proteína

del ligando BDB 26410214 y los residuos de la viroporina E. Las etiquetas en la parte superior corresponden al residuo involucrado en la interacción con cada pose. Estas indican el tipo de residuo el número del residuo, el nombre en su código de tres letras y el identificador de la cadena del homopentaméro donde se encuentra localizado el residuo. Las etiquetas del lado izquierdo indican el nombre del algoritmo que calculó la pose del ligando BDB 26410214. En el centro se encuentran marcas que indican el tipo de interacción formada entre la pose del ligando BDB 26410214 y la proteína. La información de los tipos de interacciones se encuentra en la lista localizada en la parte derecha de la gráfica.



Figura 46.- En este mapa de calor se muestra la similitud entre las huellas digitales de interacción de las diferentes poses del ligando BDB 26410214. Las etiquetas contienen el nombre de los algoritmos y los cuadros al centro muestran la relación de similitud entre los patrones de interacciones calculadas por cada algoritmo. En la derecha se encuentra la escala de los colores que permiten identificar el valor de la relación entre cada par del mapa.



RMSD de las poses del compuesto BDB 26420214 en complejo con la Viroporina E.

Figura 47.- Esta figura muestra un mapa de calor que nos ayuda a visualizar el valor del RMSD entre cada una de las poses calculadas durante las simulaciones de acople molecular del ligando BDB 26410214. La escala de colores tiene asignado un valor numérico, el cual se muestra en la escala de la derecha. El dendograma de la izquierda es para identificar con mayor facilidad los objetos que guardan mayor relación al tener valores de RMSD menores entre sus átomos.

En la figura 48 se muestra el complejo formado por la viroporina E y el compuesto BDB 26422058 (figura 48). En esta figura se observa una orientación similar de las poses del compuesto, que se encuentran ocupando la misma cavidad. El análisis de las interacciones muestra un patrón similar con coincidencia de varias cadenas laterales (figuras 49 y 50), principalmente la Phe 16C, la Phe 19C, la Val 18D, la Phe 19D y la Phe 19E. A pesar de las coincidencias se observa que la similitud de los PLIFs (figura 51) es menor a 0.6, lo que sugiere que en este caso de estudio los resultados poseen una menor precisión en la predicción del modo de interacción de dicha molécula. Esto se puede corroborar con el análisis de la RMSD de las poses (figura 52), en la que los valores son ≥ 3.0 Å.



Figura 48.- En esta figura se muestran los mapas de interacciones del ligando BDB 26422058. Las etiquetas de interacciones (esquina inferior izquierda) indica el tipo de residuo e interacción presentes en el complejo ligandoproteína. Junto a cada mapa de interacción se encuentra anotado el algoritmo que calculó esa pose, el valor de la función de puntaje, y el nombre del compuesto. En esta imagen se encuentra anotado el número del residuo, el nombre en su código de tres letras y el identificador de la cadena del homopentaméro donde se encuentra localizado el residuo.



Figura 49.- En esta figura se muestran los mapas de interacciones del ligando BDB 26422058. Las etiquetas de interacciones (esquina inferior izquierda) indica el tipo de residuo e interacción presentes en el complejo ligandoproteína. Junto a cada mapa de interacción se encuentra anotado el algoritmo que calculó esa pose, el valor de la función de puntaje, y el nombre del compuesto. En esta imagen se encuentra anotado el número del residuo, el nombre en su código de tres letras y el identificador de la cadena del homopentaméro donde se encuentra localizado el residuo.



Huellas digitales de las intereacciones ligando-proteína del complejo BDB 26422058-Viroporina E.

Figura 50.- En esta figura se exponen los resultados de las huellas digitales de las interacciones ligando-proteína del ligando BDB 26422058 y los residuos de la viroporina E. Las etiquetas en la parte superior corresponden al residuo involucrado en la interacción con cada pose. Estas indican el tipo de residuo el número del residuo, el nombre en su código de tres letras y el identificador de la cadena del homopentaméro donde se encuentra localizado el residuo. Las etiquetas del lado izquierdo indican el nombre del algoritmo que calculó la pose del ligando BDB 26422058. En el centro se encuentran marcas que indican el tipo de interacción formada entre la pose del ligando BDB 26422058 y la proteína. La información de los tipos de interacciones se encuentra en la lista localizada en la parte derecha de la gráfica.



Figura 51.- En este mapa de calor se muestra la similitud entre las huellas digitales de interacción de las diferentes poses del ligando BDB 26422058. Las etiquetas contienen el nombre de los algoritmos y los cuadros al centro muestran la relación de similitud entre los patrones de interacciones calculadas por cada algoritmo. En la derecha se encuentra la escala de los colores que permiten identificar el valor de la relación entre cada par del mapa.



RMSD de las poses del compuesto BDB 26422058 en complejo con la Viroporina E.

Figura 52.- Esta figura muestra un mapa de calor que nos ayuda a visualizar el valor del RMSD entre cada una de las poses calculadas durante las simulaciones de acople molecular del ligando BDB 26422058. La escala de colores tiene asignado un valor numérico, el cual se muestra en la escala de la derecha. El dendograma de la izquierda es para identificar con mayor facilidad los objetos que guardan mayor relación al tener valores de RMSD menores entre sus
átomos.

El complejo formado por las poses de la molécula BDB 23168933 esquematizado en la figura 53, podemos ver mediante una inspección visual que hay una alta convergencia de las poses generadas por los algoritmos de acople molecular (figura 53). Mediante los mapas de interacción (figura 54) y la tabla de PLIFs (figura 55), notamos que los patrones de interacción están altamente conservados entre las tres poses, formando interacciones predominantemente de tipo hidrofóbicas con los residuos Ala 15A, Val 18A, Phe 19A, Val 22A, Phe 16C, Phe 19D, Val 22D, Leucina 12E, Ala 15E, Phe 16E, Phe 19E, Leu 20E y Thr 23E. La alta similitud de los patrones de interacción se puede corroborar con el mapa de calor de la figura 56), en donde se muestra una similitud de entre 0.79 (Autodock 4 y Autodock Vina) y 0.89 (PLANTS y Autodock Vina), además de un valor de RMSD \leq 1.5 Å entre las poses (figura 57). Estos valores sugieren una gran precisión en la predicción de las interacciones de ligando y la proteína.



Figura 53.- En esta figura se muestran los mapas de interacciones del ligando BDB 23168933. Las etiquetas de interacciones (esquina inferior izquierda) indica el tipo de residuo e interacción presentes en el complejo ligandoproteína. Junto a cada mapa de interacción se encuentra anotado el algoritmo que calculó esa pose, el valor de la función de puntaje, y el nombre del compuesto. En esta imagen se encuentra anotado el número del residuo, el nombre en su código de tres letras y el identificador de la cadena del homopentaméro donde se encuentra localizado el residuo.



Figura 54.- En esta figura se muestran los mapas de interacciones del ligando BDB 23168933. Las etiquetas de interacciones (esquina inferior izquierda) indica el tipo de residuo e interacción presentes en el complejo ligandoproteína. Junto a cada mapa de interacción se encuentra anotado el algoritmo que calculó esa pose, el valor de la función de puntaje, y el nombre del compuesto. En esta imagen se encuentra anotado el número del residuo, el nombre en su código de tres letras y el identificador de la cadena del homopentaméro donde se encuentra localizado el residuo.





Figura 55.- En esta figura se exponen los resultados de las huellas digitales de las interacciones ligando-proteína del ligando BDB 23168933 y los residuos de la viroporina E. Las etiquetas en la parte superior corresponden al residuo involucrado en la interacción con cada pose. Estas indican el tipo de residuo el número del residuo, el nombre en su código de tres letras y el identificador de la cadena del homopentaméro donde se encuentra localizado el residuo. Las etiquetas del lado izquierdo indican el nombre del algoritmo que calculó la pose del ligando BDB 23168933. En el centro se encuentran marcas que indican el tipo de interacción formada entre la pose del ligando BDB 23168933 y la proteína. La información de los tipos de interacciones se encuentra en la lista localizada en la parte derecha de la gráfica.



Figura 56.- En este mapa de calor se muestra la similitud entre las huellas digitales de interacción de las diferentes poses del ligando BDB 23168933. Las etiquetas contienen el nombre de los algoritmos y los cuadros al centro muestran la relación de similitud entre los patrones de interacciones calculadas por cada algoritmo. En la derecha se encuentra la escala de los colores que permiten identificar el valor de la relación entre cada par del mapa.



Consensus docking pose/RMSD clustering

RMSD de las poses del compuesto BDB 23168933 en complejo con la Viroporina E.

Figura 57.- Esta figura muestra un mapa de calor que nos ayuda a visualizar el valor del RMSD entre cada una de las poses calculadas durante las simulaciones de acople molecular del ligando BDB 23168933. La escala de colores tiene asignado un valor numérico, el cual se muestra en la escala de la derecha. El dendograma de la izquierda es para identificar con mayor

facilidad los objetos que guardan mayor relación al tener valores de RMSD menores entre sus átomos.

De la misma manera que los casos anteriores, en la figura 58 se muestra el complejo formado por las poses del compuesto BDC 23169243 (figura 58) en donde mediante una rápida inspección visual podemos percatarnos de la alta convergencia de las poses. En el análisis de interacciones (figura 59 y 60) se observa que existe un patrón de interacciones casi idéntico entre las tres poses, esto se puede apreciar fácilmente en el esquema que muestra los PLIFs (figura 61), en donde hay una prevalencia de interacciones con los residuos Phe 19A, Phe 16C, Phe 19C, Val 18D, Phe 19D, Phe 16E y Phe 19E, tal como se observó en los casos anteriores. Con lo anterior se puede observar la consistencia de los resultados en el mapa de calor de la figura 62 que muestra la similitud de los PLIFs, obtenidos de un Análisis de similitud de Animito. Por último, en la figura 63 se puede notar bajos niveles de RMSD que van desde los 1.5 Å (Autodock Vina y PLANTS) hasta 0.9 Å (PLANTS y Autodock 4). Los valores de similitud de los PLIFS y los bajos valores de la RMSD sugieren una alta precisión de la predicción de los resultados de las metares.



Figura 58.- En esta figura se muestran los mapas de interacciones del ligando BDB 23168933. Las etiquetas de interacciones (esquina inferior izquierda) indica el tipo de residuo e interacción presentes en el complejo ligandoproteína. Junto a cada mapa de interacción se encuentra anotado el algoritmo que calculó esa pose, el valor de la función de puntaje, y el nombre del compuesto. En esta imagen se encuentra anotado el número del residuo, el nombre en su código de tres letras y el identificador de la cadena del homopentaméro donde se encuentra localizado el residuo.



Figura 59.- En esta figura se muestran los mapas de interacciones del ligando BDB 23168933. Las etiquetas de interacciones (esquina inferior izquierda) indica el tipo de residuo e interacción presentes en el complejo ligandoproteína. Junto a cada mapa de interacción se encuentra anotado el algoritmo que calculó esa pose, el valor de la función de puntaje, y el nombre del compuesto. En esta imagen se encuentra anotado el número del residuo, el nombre en su código de tres letras y el identificador de la cadena del homopentaméro donde se encuentra localizado el residuo.



Huellas digitales de las intereacciones ligando-proteína del complejo BDC 23169243-Viroporina E.

Figura 60.- En esta figura se exponen los resultados de las huellas digitales de las interacciones ligando-proteína del ligando BDB 23168933 y los residuos de la viroporina E. Las etiquetas en la parte superior corresponden al residuo involucrado en la interacción con cada pose. Estas indican el tipo de residuo el número del residuo, el nombre en su código de tres letras y el identificador de la cadena del homopentaméro donde se encuentra localizado el residuo. Las etiquetas del lado izquierdo indican el nombre del algoritmo que calculó la pose del ligando BDB 23168933. En el centro se encuentran marcas que indican el tipo de interacción formada entre la pose del ligando BDB 23168933 y la proteína. La información de los tipos de interacciones se encuentra en la lista localizada en la parte derecha de la gráfica.



Figura 61.- En este mapa de calor se muestra la similitud entre las huellas digitales de interacción de las diferentes poses del ligando BDB 23168933. Las etiquetas contienen el nombre de los algoritmos y los cuadros al centro muestran la relación de similitud entre los patrones de interacciones calculadas por cada algoritmo. En la derecha se encuentra la escala de los colores que permiten identificar el valor de la relación entre cada par del mapa.



RMSD de las poses del compuesto BDC 23169243 en complejo con la Viroporina E.

Figura 62.- Esta figura muestra un mapa de calor que nos ayuda a visualizar el valor del RMSD entre cada una de las poses calculadas durante las simulaciones de acople molecular del ligando BDB 23168933. La escala de colores tiene asignado un valor numérico, el cual se muestra en la escala de la derecha. El dendograma de la izquierda es para identificar con mayor facilidad los objetos que guardan mayor relación al tener valores de RMSD menores entre sus átomos.

En la figura 63 se muestra ver el complejo formado por la viroporina E y las poses de compuesto BDC 23205600, con una rápida inspección visual se observa que las poses se encuentran ocupando coordenadas similares (figura 63). A través del mapa de interacciones (figura 64) y la tabla de PLIFs (figura 65) se aprecia un patrón de interacciones casi idéntico, en donde las poses se encuentran formando interacciones con los residuos Val 22A, Phe 16B, Leucina 12C, Ala 15C, Phe 16 C, Phe 19C, Ala 15D, Val 18D, Phe 19D, Phe 16E, Leu 20 E y Thr 23E. Mediante el mapa de calor de la figura 66 se aprecia con mayor facilidad la similitud entre los patrones de interacción, en donde Autodock Vina y Autodock 4 comparten una similitud de 1, mientras que PLANTS posee un valor de 0.78. Por último, en la figura 67, se observa un RMSD de 1.5 Å entre las poses generadas por Autodock Vina y Autodock 4, y por otro lado una RMSD de 2.3 Å entre las poses de PLANTS y Autodock 4.



Figura 63.- En esta figura se muestran los mapas de interacciones del ligando BDC 23205600. Las etiquetas de interacciones (esquina inferior izquierda) indica el tipo de residuo e interacción presentes en el complejo ligandoproteína. Junto a cada mapa de interacción se encuentra anotado el algoritmo que calculó esa pose, el valor de la función de puntaje, y el nombre del compuesto. En esta imagen se encuentra anotado el número del residuo, el nombre en su código de tres letras y el identificador de la cadena del homopentaméro donde se encuentra localizado el residuo.



Figura 64.- En esta figura se muestran los mapas de interacciones del ligando BDC 23205600. Las etiquetas de interacciones (esquina inferior izquierda) indica el tipo de residuo e interacción presentes en el complejo ligando-proteína. Junto a cada mapa de interacción se encuentra anotado el algoritmo que calculó esa pose, el valor de la función de puntaje, y el nombre del compuesto. En esta imagen se encuentra anotado el número del residuo, el nombre en su código de tres letras y el identificador de la cadena del homopentaméro donde se encuentra localizado el residuo.



Huellas digitales de las intereacciones ligando-proteína del complejo BDC 23205600-Viroporina E.

Figura 65.- En esta figura se exponen los resultados de las huellas digitales de las interacciones ligando-proteína del ligando BDC 23205600 y los residuos de la viroporina E. Las etiquetas en la parte superior corresponden al residuo involucrado en la interacción con cada pose. Estas indican el tipo de residuo el número del residuo, el nombre en su código de tres letras y el identificador de la cadena del homopentaméro donde se encuentra localizado el residuo. Las etiquetas del lado izquierdo indican el nombre del algoritmo que calculó la pose del ligando BDC 23205600. En el centro se encuentran marcas que indican el tipo de interacción formada entre la pose del ligando BDC 23205600 y la proteína. La información de los tipos de interacciones se encuentra en la lista localizada en la parte derecha de la gráfica.



Figura 66.- En este mapa de calor se muestra la similitud entre las huellas digitales de interacción de las diferentes poses del ligando BDC 23205600. Las etiquetas contienen el nombre de los algoritmos y los cuadros al centro muestran la relación de similitud entre los patrones de interacciones calculadas por cada algoritmo. En la derecha se encuentra la escala de los colores que permiten identificar el valor de la relación entre cada par del mapa.



RMSD de las poses del compuesto BDC 23205600 en complejo con la Viroporina E.

Figura 67.- Esta figura muestra un mapa de calor que nos ayuda a visualizar el valor del RMSD entre cada una de las poses calculadas durante las simulaciones de acople molecular del ligando BDC 23205600. La escala de colores tiene asignado un valor numérico, el cual se muestra en la escala de la derecha. El dendograma de la izquierda es para identificar con mayor facilidad los objetos que guardan mayor relación al tener valores de RMSD menores entre sus átomos.

El siguiente complejo corresponde a la molécula BDC 232025697, que en la figura 68 se muestra la disposición de las poses generadas a partir de las simulaciones de acople molecular (figura 68). En los mapas de interacción se observa una predominancia de residuos e interacciones hidrofóbicas (figura 69), esto se observa claramente en la tabla de PLIFs, la cual muestra una alta coincidencia de los patrones de interacción, con una alta incidencia de interacciones formadas por la Phe 16 y 19 de las cadenas B, D y E (figura 70). En la figura 71 se muestra el mapa de calor que contiene los índices de similitud entre los patrones de interacción formados por cada pose. Los índices de similitud de las interacciones van desde 0.65 (PLANTS-Autodock 4, Autodock Vina-Autodock 4) hasta 0.79 (PLANTS-Autodock Vina), lo cual indica un índice de conservación alto. Para complementar el análisis de interacciones en la figura 72 se muestra el mapa de calor que contiene los átomos de las poses, cuyos valores van desde 1.5 Å entre las poses generadas por PLANTS y Autodock 4, hasta 2.5 Å entre las poses generadas por Autodock Vina y Autodock 4.



Figura 68.- En esta figura se muestran los mapas de interacciones del ligando BDC 232025697. Las etiquetas de interacciones (esquina inferior izquierda) indica el tipo de residuo e interacción presentes en el complejo ligando-proteína. Junto a cada mapa de interacción se encuentra anotado el algoritmo que calculó esa pose, el valor de la función de puntaje, y el nombre del compuesto. En esta imagen se encuentra anotado el número del residuo, el nombre en su código de tres letras y el identificador de la cadena del homopentaméro donde se encuentra localizado el residuo.



Figura 69.- En esta figura se muestran los mapas de interacciones del ligando BDC 232025697. Las etiquetas de interacciones (esquina inferior izquierda) indica el tipo de residuo e interacción presentes en el complejo ligando-proteína. Junto a cada mapa de interacción se encuentra anotado el algoritmo que calculó esa pose, el valor de la función de puntaje, y el nombre del compuesto. En esta imagen se encuentra anotado el número del residuo, el nombre en su código de tres letras y el identificador de la cadena del homopentaméro donde se encuentra localizado el residuo.



Huellas digitales de las intereacciones ligando-proteína del complejo BDC 23205697-Viroporina E.

Figura 70.- En esta figura se exponen los resultados de las huellas digitales de las interacciones ligando-proteína del ligando BDC 232025697 y los residuos de la viroporina E. Las etiquetas en la parte superior corresponden al residuo involucrado en la interacción con cada pose. Estas indican el tipo de residuo el número del residuo, el nombre en su código de tres letras y el identificador de la cadena del homopentaméro donde se encuentra localizado el residuo. Las etiquetas del lado izquierdo indican el nombre del algoritmo que calculó la pose del ligando BDC 232025697. En el centro se encuentran marcas que indican el tipo de interacción formada entre la pose del ligando BDC 232025697 y la proteína. La información de los tipos de interacciones se encuentra en la lista localizada en la parte derecha de la gráfica.



Figura 71.- En este mapa de calor se muestra la similitud entre las huellas digitales de interacción de las diferentes poses del ligando BDC 232025697. Las etiquetas contienen el nombre de los algoritmos y los cuadros al centro muestran la relación de similitud entre los patrones de interacciones calculadas por cada algoritmo. En la derecha se encuentra la escala de los colores que permiten identificar el valor de la relación entre cada par del mapa.



RMSD de las poses del compuesto BDC 23205697 en complejo con la Viroporina E.

Figura 72.- Esta figura muestra un mapa de calor que nos ayuda a visualizar el valor del RMSD entre cada una de las poses calculadas durante las simulaciones de acople molecular del ligando BDC 232025697. La escala de colores tiene asignado un valor numérico, el cual se muestra en la escala de la derecha. El dendograma de la izquierda es para identificar con mayor facilidad los objetos que guardan mayor relación al tener valores de RMSD menores entre sus átomos.

Las poses del compuesto BDG 34062417, generadas por los algoritmos de acople molecular utilizados en este estudio, convergieron al igual que todas las moléculas anteriores, dentro del poro, ocupando el mismo sitio (figura 73). En los mapas de interacciones (figura 74) se observa, que las interacciones y residuos que imperan son hidrofóbicos, del tipo alifático y aromáticos. Al igual que las moléculas anteriores, los residuos predominantes corresponden a la Phe 16 y 19 de las cadenas B, C y D, con las cuales se forman interacciones tipo π . Por otro lado, las interacciones hidrofóbicas corresponden a los residuos Ala 12C, Leu 15C y las Val 18 y 22 de la cadena D, principalmente. En la figura 75 se pueden apreciar los PLIFs calculados en donde se presentan el tipo y numero de residuo, y el tipo de interacción. La similitud de estos patrones se expone en el mapa de calor de la figura 76, en donde los índices de similitud van desde 0.76 entre las poses calculadas por PLANTS y Autodock 4, a 0.88 entre las poses generadas por Autodock 4 y Autodock Vina. De forma complementaria en la figura 77, se muestran los valores de RMSD entre las diferentes poses, donde los valores se encuentran entre el intervalo de 0.65 Å (PLANTS-Autodock 4) a 0.9 Å (PLANTS-Autodock Vina), con lo cual podemos suponer que la predicción del modo de interacción de esta molécula presume de gran precisión.



Figura 73.- En esta figura se muestran los mapas de interacciones del ligando BDG 34062417. Las etiquetas de interacciones (esquina inferior izquierda) indica el tipo de residuo e interacción presentes en el complejo ligandoproteína. Junto a cada mapa de interacción se encuentra anotado el algoritmo que calculó esa pose, el valor de

la función de puntaje, y el nombre del compuesto. En esta imagen se encuentra anotado el número del residuo, el nombre en su código de tres letras y el identificador de la cadena del homopentaméro donde se encuentra localizado el residuo.



Figura 74.- En esta figura se muestran los mapas de interacciones del ligando BDG 34062417. Las etiquetas de interacciones (esquina inferior izquierda) indica el tipo de residuo e interacción presentes en el complejo ligandoproteína. Junto a cada mapa de interacción se encuentra anotado el algoritmo que calculó esa pose, el valor de la función de puntaje, y el nombre del compuesto. En esta imagen se encuentra anotado el número del residuo, el nombre en su código de tres letras y el identificador de la cadena del homopentaméro donde se encuentra localizado el residuo.



Huellas digitales de las intereacciones ligando-proteína del complejo BDG 34062417 -Viroporina E.

Figura 75.- En esta figura se exponen los resultados de las huellas digitales de las interacciones ligando-proteína del ligando BDG 34062417 y los residuos de la viroporina E. Las etiquetas en la parte superior corresponden al residuo involucrado en la interacción con cada pose. Estas indican el tipo de residuo el número del residuo, el nombre en su código de tres letras y el identificador de la cadena del homopentaméro donde se encuentra localizado el residuo. Las etiquetas del lado izquierdo indican el nombre del algoritmo que calculó la pose del ligando BDG 34062417. En el centro se encuentran marcas que indican el tipo de interacción formada entre la pose del ligando BDG 34062417 y la proteína. La información de los tipos de interacciones se encuentra en la lista localizada en la parte derecha de la gráfica.



Figura 76.- En este mapa de calor se muestra la similitud entre las huellas digitales de interacción de las diferentes poses del ligando BDG 34062417. Las etiquetas contienen el nombre de los algoritmos y los cuadros al centro muestran la relación de similitud entre los patrones de interacciones calculadas por cada algoritmo. En la derecha se encuentra la escala de los colores que permiten identificar el valor de la relación entre cada par del mapa.



Consensus docking pose/RMSD clustering



Figura 77.- Esta figura muestra un mapa de calor que nos ayuda a visualizar el valor del RMSD entre cada una de las poses calculadas durante las simulaciones de acople molecular del ligando BDG 34062417. La escala de colores tiene asignado un valor numérico, el cual se muestra en la escala de la derecha. El dendograma de la izquierda es para identificar con mayor facilidad los objetos que guardan mayor relación al tener valores de RMSD menores entre sus átomos.

El siguiente complejo pertenece a él formado por la viroporina E y el compuesto BDG 34121287. En la imagen 78 se muestran los resultados obtenidos del acople molecular de esta molécula. Sobre esta molécula se observa que las tres poses obtenidas a partir de las simulaciones de acople molecular se orientaron de una forma casi idéntica (figura 78). Para esclarecer la similitud de las predicciones generadas, observamos que en los mapas de interacciones (figura 79), hay una similitud en el tipo de interacciones y en los residuos que forman las interacciones con cada una de las poses. Se puede apreciar en la tabla de PLIFs (figura 80) una persistencia de las interacciones π formadas con los residuos Phe 16 y 19 de las cadenas C y E, además de interacciones del tipo hidrofóbicas con los residuos Leucina 12C, así como con la Ala 15 de las cadenas C y E. Los resultados del análisis de similitud de los patrones de interacción se muestran en el mapa de calor en la figura 81, en donde los índices de similitud son \geq 0.9, estos resultados aunados a el análisis de RMSD (figura 82) con valores \leq 1.0 Å, sugieren una alta precisión en el cálculo del modo de unión de la molécula con el receptor.



Figura 78.- En esta figura se muestran los mapas de interacciones del ligando BDG 34062417. Las etiquetas de interacciones (esquina inferior izquierda) indica el tipo de residuo e interacción presentes en el complejo ligandoproteína. Junto a cada mapa de interacción se encuentra anotado el algoritmo que calculó esa pose, el valor de la función de puntaje, y el nombre del compuesto. En esta imagen se encuentra anotado el número del residuo, el nombre en su código de tres letras y el identificador de la cadena del homopentaméro donde se encuentra localizado el residuo.



Figura 79.- En esta figura se muestran los mapas de interacciones del ligando BDG 34062417. Las etiquetas de interacciones (esquina inferior izquierda) indica el tipo de residuo e interacción presentes en el complejo ligandoproteína. Junto a cada mapa de interacción se encuentra anotado el algoritmo que calculó esa pose, el valor de la función de puntaje, y el nombre del compuesto. En esta imagen se encuentra anotado el número del residuo, el nombre en su código de tres letras y el identificador de la cadena del homopentaméro donde se encuentra localizado el residuo.



Huellas digitales de las intereacciones ligando-proteína del complejo BDG 34121287 -Viroporina E.

Figura 80.- En esta figura se exponen los resultados de las huellas digitales de las interacciones ligando-proteína del ligando BDG 34062417 y los residuos de la viroporina E. Las etiquetas en la parte superior corresponden al residuo involucrado en la interacción con cada pose. Estas indican el tipo de residuo el número del residuo, el nombre en su código de tres letras y el identificador de la cadena del homopentaméro donde se encuentra localizado el residuo. Las etiquetas del lado izquierdo indican el nombre del algoritmo que calculó la pose del ligando BDG 34062417. En el centro se encuentran marcas que indican el tipo de interacción formada entre la pose del ligando BDG 34062417 y la proteína. La información de los tipos de interacciones se encuentra en la lista localizada en la parte derecha de la gráfica.



Figura 81.- En este mapa de calor se muestra la similitud entre las huellas digitales de interacción de las diferentes poses del ligando BDG 34062417. Las etiquetas contienen el nombre de los algoritmos y los cuadros al centro muestran la relación de similitud entre los patrones de interacciones calculadas por cada algoritmo. En la derecha se encuentra la escala de los colores que permiten identificar el valor de la relación entre cada par del mapa.



(15,35,125,14R)-11,11-dimethyl-2,10-dioxapentacyclo[12.10.0.0^{3,12}.0^{4,9}.0^{15,20}]tetracosa-4,6,8,15(20),16,18-hexaene-6-carboxylic acid **Consensus docking pose/RMSD clustering**

Figura 82.- Esta figura muestra un mapa de calor que nos ayuda a visualizar el valor del RMSD entre cada una de las poses calculadas durante las simulaciones de acople molecular del ligando BDG 34062417. La escala de colores tiene asignado un valor numérico, el cual se muestra en la escala de la derecha. El dendograma de la izquierda es para identificar con mayor facilidad los objetos que guardan mayor relación al tener valores de RMSD menores entre sus átomos.

Por último, tenemos al complejo formado por el compuesto BDB 26419775 y la proteína E (figura 37). En este caso, en la imagen 37 A se observa que las poses encontradas como más favorables, poseen una geometría similar, lo que conduce a las moléculas a formar patrones de interacción similares. En los mapas de interacción (figura 37 B) se observa una considerable similitud de las cadenas laterales y el tipo de interacción que forman. En la tabla de PLIFs (figura 37 C) se observa una prevalencia de los residuos responsables de formar las interacciones tal como la Phe 16 y 19 de las cadenas A, C, D y E, que participan en el establecimiento de interacciones tipo π . Para analizar la similitud de los patrones, se graficó el mapa de calor (figura 37 D) con los índices de similitud, en donde los valores van desde 0.68 (PLANTS-Autodock 4), hasta 0.83 (Autodock Vina-Autodock 4), lo que sugiere que las poses con un índice de 0.83 están altamente relacionadas. Para complementar este análisis se calculó el RMSD de las poses, en donde los valores son mostrados en el mapa de calor de la figura 37 E. En el mapa de calor podemos apreciar que los valores de la RMSD son \leq 1.5 Å, en donde las poses con menor RMSD son las predichas por Autodock Vina y Autodock 4.



Figura 83.- En esta figura se muestran los mapas de interacciones del ligando BDB 26419775. Las etiquetas de interacciones (esquina inferior izquierda) indica el tipo de residuo e interacción presentes en el complejo ligandoproteína. Junto a cada mapa de interacción se encuentra anotado el algoritmo que calculó esa pose, el valor de la función de puntaje, y el nombre del compuesto. En esta imagen se encuentra anotado el número del residuo, el nombre en su código de tres letras y el identificador de la cadena del homopentaméro donde se encuentra localizado el residuo.



Figura 84.- En esta figura se muestran los mapas de interacciones del ligando BDB 26419775. Las etiquetas de interacciones (esquina inferior izquierda) indica el tipo de residuo e interacción presentes en el complejo ligandoproteína. Junto a cada mapa de interacción se encuentra anotado el algoritmo que calculó esa pose, el valor de la función de puntaje, y el nombre del compuesto. En esta imagen se encuentra anotado el número del residuo, el nombre en su código de tres letras y el identificador de la cadena del homopentaméro donde se encuentra localizado el residuo.



Huellas digitales de las intereacciones ligando-proteína del complejo BDB 26419775-Viroporina E.

Figura 85.- En esta figura se exponen los resultados de las huellas digitales de las interacciones ligando-proteína del ligando BDB 26419775 y los residuos de la viroporina E. Las etiquetas en la parte superior corresponden al residuo involucrado en la interacción con cada pose. Estas indican el tipo de residuo el número del residuo, el nombre en su código de tres letras y el identificador de la cadena del homopentaméro donde se encuentra localizado el residuo. Las etiquetas del lado izquierdo indican el nombre del algoritmo que calculó la pose del ligando BDB 26419775. En el centro se encuentran marcas que indican el tipo de interacción formada entre la pose del ligando BDB 26419775 y la proteína. La información de los tipos de interacciones se encuentra en la lista localizada en la parte derecha de la gráfica.



Figura 86.- En este mapa de calor se muestra la similitud entre las huellas digitales de interacción de las diferentes poses del ligando BDB 26419775. Las etiquetas contienen el nombre de los algoritmos y los cuadros al centro muestran la relación de similitud entre los patrones de interacciones calculadas por cada algoritmo. En la derecha se encuentra la escala de los colores que permiten identificar el valor de la relación entre cada par del mapa.



RMSD de las poses del compuesto BDB 26419775 en complejo con la Viroporina E.

Figura 87.- Esta figura muestra un mapa de calor que nos ayuda a visualizar el valor del RMSD entre cada una de las poses calculadas durante las simulaciones de acople molecular del ligando BDB 26419775. La escala de colores tiene asignado un valor numérico, el cual se muestra en la escala de la derecha. El dendograma de la izquierda es para identificar con mayor facilidad los objetos que guardan mayor relación al tener valores de RMSD menores entre sus átomos.

Para resumir todos los resultados obtenidos se seleccionó la pose más representativa de cada complejo, con base en los valores de los índices de Tanimoto de los PLIFs, y su RMSD, es decir la pose que guardara mayor relación con los dos restantes fue seleccionada como la óptima. Para el compuesto BDC_232005600, se seleccionó la pose generada por Autodock 4, para la molécula BDC_23205697 se seleccionó la pose calculada por Autodock Vina, para le ligando BDG_34062417 se escogió la pose generada por Autodock 4, BDB_26417644 fue la pose acoplada por Autodock Vina, la pose seleccionada para el compuesto BDG34121287 fue la obtenida por PLANTS, para BDB 265419775 fue la pose generad por Autodock Vina, para BDB_26420214 se seleccionó la generada por Autodock 4, en el caso de BDB_26422058 se optó por la pose generada por Autodock Vina, para BDB_23168933 se seleccionó la generada por PLANTS y por último para la molécula BDC_23169243 se tomó la calculada por el algoritmo de Autodock Vina.

Se calcularon los PLIFs de todas las poses representativas de cada uno de los nuevos inhibidores, para poder apreciar la frecuencia de la participación de cada uno de los residuos en la formación de interacciones (figura 88). En la tabla de PLIFs que contiene los resultados condensados, podemos observar que hay una gran participación de la Phe 19 de las cadenas C, D y E, con lo que podemos proponer que este residuo en cada una de las cadenas de pentámero es un sitio calve para la interacción de fármacos.

	4425.0	Val. 28.4	PHE19	1452	LEU12	PHE, O	8.97	HE19.P	LEU12	24425	PHE, C	Du. 6.C	7619.C	N425.0	Val. 28.	PHEN	10	0.20	LEU12E	ALAISE	PHE	3.0.E	74E19 -	1EU205	THREE	2.	
BDC_23205600						-			-	-		-		-	-	-							•••	-	-		
BDC_23205697			-			-	=			-	-					-	-				-	-		ŧ.			
BDG_34062417						-					-	4	• •	-		-	-										Hydrophob
BDG_34121287											-																HBDonor
BDB_26417644						-				-	-				-	-	-	-	-	-					-		PiStacking
BDB 26419775									-	-			• •		-			-									Cationic
BDB_26420214		-											-		-	-			-	-							
BDB_26422058																	-	-						6			
BDB_23168933	-	-														-			-	-					-		
BDC_23169243											-				-		-		-		-			-	-		



Figura 88.- En esta figura se muestran los PLIFs de las poses representativas de cada compuesto.

Comparación con estudios actuales de inhibición.

A través de una revisión en la literatura se ha encontrado que en estudios de cribado virtual de compuestos derivados de Tretrionina, como el realizado por Dey, D., Borkotoky, S., & Banerjee, M. (2020)., se han reportado como sitios de unión a ligandos la Leucina 21 (Cadena C, D), Leucina 24 (Cadena C), Ala 25 (Cadena B,C), a diferencia de lo reportado en este estudio, en donde los residuos que predominan como sitios de unión a ligandos son la Phe 19 (Cadena B,C,D,E,), Ala 15 (Cadena A,C,D,E) y Phe 16 (Cadena B,C,E), esto derivado del estado conformacional de la proteína utilizada en la mayoría de estudios de cribado virtual, ya que los modelos utilizados se construyen a partir de la estructura cristalizada de la Proteína E del SARS CoV (5x29) la cual se encuentra en estado cerrado. El cambio de estado de cerrado a abierto requiere el giro de los monómeros que conforman la proteína en forma de tornillo por lo que los residuos expuestos al dominio del poro varían de un estado conformacional a otro. Por otro lado, existen estudios de acople molecular que buscan inhibidores de la polimerización de los monómeros para formar el complejo pentámerico en la membrana del retículo endoplásmico y el aparato de Golgi., como el realizado por Bhowmik, D y colaboradores (Bhowmik et al., 2020), en el que se reportan sitios de unión a fármacos en los residuos de los segmentos dirigidos hacia el citoplasma, tal como la Tirosina 35 y Tirosina 50. Dentro de los objetivos de este estudio de acople molecular se consideraron solo los residuos localizados dentro del dominio del poro, debido a que la finalidad fue encontrar inhibidores de la corriente generada por la proteína E.

Evaluación del perfil farmacocinético de los nuevos inhibidores.

Como último paso se realizó una evaluación de las propiedades ADMET (absorción, distribución, metabolismo, absorción, excreción y toxicidad) de los compuestos resultantes del CV. Como punto de referencia se seleccionó el remdesivir para hacer una comparación de las propiedades de los fármacos actualmente aprobados para tratar la COVID-19 contra las propiedades de las moléculas encontradas en este estudio. En la figura 89 se muestran las propiedades ADMET del remdesivir, las cuales son muy pobres ya que el compuesto no tiene la capacidad de ser absorbido oralmente ni puede atravesar la barrera hematoencefálica.



Grafica de huevo cocido. La región blanca representa alta probabilidad de absorción pasiva a través del tracto gastrointestinal, y la amarilla representa alta probabilidad de cruzar la barrera hematoencefálica. En adición los puntos azules predicen si la molécula puede ser sacado activamente a través de la glicoproteína P y los rojos sí no es sustrato para la glicoproteína P.

Figura 89.- Las propiedades del remdesivir como fármaco son muy pobres, esto se refleja en la mala distribución y la dificultad de su administración que es parenteral. En la gráfica de huevo cocido no se puede apreciar el remdesivir ya que se encuentra fuera de los intervalos

En la Figura 90 se muestra el perfil de toxicidad del remdesivir, en el cual no se encontró actividad de ningún tipo.



Figura 90.- Perfil de toxicidad del remdesivir- En la gráfica de radar se puede apreciar las diferentes dianas en las cuales las moléculas pueden presentar toxicidad. La sombra naranja indica los valores esperados para moléculas toxicológicamente activas, mientras que el circulo azul en el centro representa la actividad toxica de nuestra molécula de estudio.

Los valores del radar se encuentran establecidos con base en la regla de los 5 de Lipinski. Como se puede observar en la figura 89 el compuesto Remdesivir tiene varias violaciones a los valores permitidos dentro de la regla de los 5 de Lipinski, entre las que se encuentran un gran número de enlaces rotables (Flex), una superficie polar amplia (Polar) y un alto peso molecular (Size), esto se puede notar fácilmente ya que la línea roja que representa los valores de las propiedades fisicoquímicas de la mólecula salen del radar rosa. La violación de uno o más de estos parámetros disminuye la absorción y biodisponibilidad de un fármaco administrado de forma oral.









BDB 26419775

BDG 34062417



BDB 26417644









LIPO







BDC 23205600

BDC 23169243



LIPO

BDC 23168933











El radar de biodisponibilidad permite apreciar a simple vista el perfil tipo fármaco de una molécula. El área rosa representa el intervalo óptimo para cada propiedad; (lipofobicidad: XLOGP de –0.7 a +5.0, tamaño: Peso molecular de 150 a 500 g/mol, polaridad: TPSA de 20 a 130 Å, solubilidad : log S no mayor a 6, saturación: fracción de carbonos con hibridación sp3 no menor a 0.25 y flexibilidad: no más de 9 enlaces rotables).

Figura 91.- El radar de biodisponibilidad nos permite observar que todos los compuestos obtenidos del cribado virtual se encuentran dentro de los valores óptimos de un fármaco con un buen perfil de absorción oral y con capacidad de permear la barrera hematoencefálica.

En contraste con lo encontrado con remdesivir, las moléculas obtenidas en este estudio de cribado virtual poseen un perfil toxicológico seguro, con DL50 elevadas además de presentar características fisicoquímicas que les permiten tener altos índices de absorción a través de una administración oral. En la Figura 91 se muestran los perfiles ADME individuales, donde se puede apreciar que todas las propiedades Lipinski de las moléculas se encuentran dentro de valores óptimos, lo que sugiere que son fármacos que poseen un buen perfil para una administración oral, lo que haría más fácil la administración y transporte del fármaco a diferencia de uno de administración parenteral. La gráfica de huevo cocido de la Figura 92 indica que el 60 % de los nuevos inhibidores pueden atravesar la barrera hematoencefálica, pero solo dos de ellos no son sustrato para la Glicoproteína P, lo que asegura que no serán expulsado del citoplasma con facilidad.





Adicionalmente a él análisis de las propiedades farmacocinéticas, se analizaron las propiedades toxicológicas de cada una de las moléculas obtenidas del cribado virtual. Los resultados mostraron que ningún compuesto posee actividad toxica contra algunas dianas biológicas de mayor relevancia tal como son los casos de; hepatotoxicidad, carcinogenicidad, Inmunotoxicidad, mutagenicidad, citotoxicidad, afinidad por el receptor de aril-hidrocarburo, el receptor de andrógenos, al receptor de estrógenos, al factor nuclear de respuesta a elementos antioxidantes (nrt2/ARE), al factor de respuesta al choque térmico (HSE), o con algún tipo de actividad que interfiera con el potencial de membrana mitocondrial, así como cualquier tipo de actividad sobre la proteína p53. Con base en los resultados de las propiedades toxicológicas podemos sugerir que las BDC_232005600, BDC_23205697 BDG_34062417, BDB_26417644, moléculas BDG34121287, BDB 265419775, BDB 26420214, BDB 26422058, BDB 23168933 y BDC_23169243 son compuestos candidatos para su utilización en experimentos in-vitro o incluso in-vivo. Como ya se explicó anteriormente el radar contiene las dianas ya mencionadas y remarcadas con una silueta anaranjada, que representa los valores de las moléculas con las que se entrenó el modelo que si poseen actividad. La manera de observar si el compuesto tiene afinidad por alguna de las dianas biológicas, es con un trazo azul sobre el eje que corresponde a cada blanco biológico. Así con esta
información, se puede apreciar claramente que no hay ningún tipo de actividad tóxica. Cada uno de los radares está acompañado de la estructura de la molécula a la izquierda del radar, y debajo se encuentra el valor de la DL50 relativa a cada fármaco. Los valores de las DL50 de los inhibidores obtenidos en este estudio son en su mayoría > 500 mg/kg, lo que significa que tienen un perfil igual de seguro que el Ibuprofeno, que posee una LD50 de 636 mg/kg en rata, de acuerdo con datos depositados en Drug Bank (<u>https://www.drugbank.ca/drugs/DB01050</u>). En la figura 93, se muestran las gráficas de radar con los resultados de la predicción de propiedades toxicológicas de cada una de las moléculas obtenidas del cribado virtual.









Receptor de androgénos. Dominio de unión del receptor de androgénos. Aromatasa. Receptor de estrogénos. Receptor de estrogénos.

Receptor de

hidrocaruburos de arilo.

El factor nuclear eritroide .



Receptor de androgénos.

Dominio de unión del receptor de androgénos.

110

Receptores activados por proliferadores

de peroxisomas.

Dominio de unión del receptor de estrogénos.

Receptor de estrogénos.

Aromatasa.



Figura 93.- En esta figura se exponen los resultados del análisis de toxicidad de las moléculas antivirales BDC_232005600 (A), BDC_23205697 (B), BDG_34062417 (C), BDB_26417644 (D), BDG34121287 (E), BDB 265419775 (F), BDB_26420214 (G), BDB_26422058 (H), BDB_23168933 (I) y por último para la molécula BDC_23169243 (J). Los radares muestran del lado derecho la estructura del compuesto acompañado de su LD50 en rata. En el centro el radar muestra blancos biológicos sensibles a toxicidad. En los esquemas se observa que ninguna molécula muestra actividad toxica de ningún tipo.

Con el conjunto de datos obtenido del cálculo de las propiedades ADME más la predicción de la toxicidad de cada molécula proponemos dichas moléculas como potenciales inhibidores, candidatos a fármacos los cuáles tienen una alta probabilidad de proporcionar buenos resultados en experimentos in vitro.

Conclusiones.

Mediante el uso de una estructura de la viroporina del SARS CoV 2, obtenida de estudios computacionales de electrofisiología, que es un potencial blanco para el desarrollo de nuevos tratamientos de la COVID, se realizó un estudio de acoplamiento molecular, utilizando diferentes algoritmos para aumentar la precisión de los resultados. Entre los resultados obtenidos de este estudio:

Se encontró un potencial sitio de unión a fármacos (el residuo Phe 19 de las cadenas A, B, C,

D y E) dentro del dominio del poro de la proteína.Los algoritmos de acople molecular convergieron en los resultados obtenidos, generando poses de las moléculas con patrones de gran similitud y un bajo RMSD. Esto nos permite sugerir que son potenciales bloqueadores de la corriente entrante de la proteína E.

 Las propiedades ADMET de las moléculas son óptimas, por lo que estos ligandos pueden utilizarse en otro tipo de estudios, utilizando metodologías de laboratorio húmedo.

Perspectivas.

En futuros estudios se ha planteado la posibilidad de explorar diversas conformaciones del canal utilizando métodos de dinámica molecular dirigida, para determinar la conductividad del canal en cada uno de ellos, así como para medir la estabilidad del ligando dentro del poro de la proteína. Para medir la conductividad del canal se propone el uso de un protocolo en el cual se inserta la proteína en una bicapa lipídica, este sistema se coloca en medio de dos membranas para formar dos compartimientos en los cuales se añaden concentraciones fisiológicas de CaCl2 con distintas concentraciones para generar un gradiente. La diferencia de gradientes genera un flujo de corriente y posteriormente se mide la cantidad de iones que cruzan el poro del canal por unidad de tiempo. Este protocolo puede ser desarrollado utilizando herramientas como gromacs.

Finalmente sería recomendable realizar un estudio de mutagénesis en el residuo Phe 19 de cada uno de los monómeros para evaluar la aportación energética de las interacciones formadas por dicho residuo.

BIBLIOGRAFÍA.

Adedeji, A. O., Singh, K., Kassim, A., Coleman, C. M., Elliott, R., Weiss, S. R., Frieman, M. B., & Sarafianos, S. G. (2014). Evaluation of SSYA10-001 as a Replication Inhibitor of Severe Acute Respiratory Syndrome, Mouse Hepatitis, and Middle East Respiratory Syndrome Coronaviruses. Antimicrobial Agents and Chemotherapy, 58(8), 4894–4898. https://doi.org/10.1128/aac.02994-14

Agarkova, I., Dunigan, D., Gurnon, J., Greiner, T., Barres, J., Thiel, G., & Van Etten, J. L. (2008). Chlorovirus-Mediated Membrane Depolarization of Chlorella Alters Secondary Active Transport of Solutes. Journal of Virology, 82(24), 12181–12190. https://doi.org/10.1128/jvi.01687-08

Agostini, M. L., Andres, E. L., Sims, A. C., Graham, R. L., Sheahan, T. P., Lu, X., Smith, E. C., Case, J. B., Feng, J. Y., Jordan, R., Ray, A. S., Cihlar, T., Siegel, D., Mackman, R. L., Clarke, M. O., Baric, R. S., & Denison, M. R. (2018). Coronavirus Susceptibility to the Antiviral Remdesivir (GS-5734) Is Mediated by the Viral Polymerase and the Proofreading Exoribonuclease. mBio, 9(2). https://doi.org/10.1128/mbio.00221-18

Agostini, M. L., Pruijssers, A. J., Chappell, J. D., Gribble, J., Lu, X., Andres, E. L., Bluemling, G. R., Lockwood, M. A., Sheahan, T. P., Sims, A. C., Natchus, M. G., Saindane, M., Kolykhalov, A. A., Painter, G. R., Baric, R. S., & Denison, M. R. (2019). Small-Molecule Antiviral β - d -- N 4 -Hydroxycytidine Inhibits a Proofreading-Intact Coronavirus with a High Genetic Barrier to Resistance. Journal of Virology, 93(24). https://doi.org/10.1128/jvi.01348-19

Almazán, F., DeDiego, M. L., GaláN, C., Escors, D., ÁLvarez, E., Ortego, J., Sola, I., ZuñIga, S., Alonso, S., Moreno, J. L., Nogales, A., Capiscol, C., & Enjuanes, L. (2006). Construction of a Severe Acute Respiratory Syndrome Coronavirus Infectious cDNA Clone and a Replicon To Study Coronavirus RNA Synthesis. Journal of Virology, 80(21), 10900–10906. https://doi.org/10.1128/jvi.00385-06

Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R., & Hilgenfeld, R. (2003). Coronavirus Main Proteinase (3CL pro) Structure: Basis for Design of Anti-SARS Drugs. Science, 300(5626), 1763–1767. https://doi.org/10.1126/science.1085658

Baell, J., & Walters, M. A. (2014). Chemistry: Chemical con artists foil drug discovery. *Nature*, *513*(7519), 481-483

Bafna, K., White, K., Harish, B., Rosales, R., Ramelot, T. A., Acton, T. B., Moreno, E., Kehrer, T., Miorin, L., Royer, C. A., García-Sastre, A., Krug, R. M., & Montelione, G. T. (2021). Hepatitis C virus drugs that inhibit SARS-CoV-2 papain-like protease synergize with remdesivir to suppress viral replication in cell culture. Cell Reports, 35(7), 109133. https://doi.org/10.1016/j.celrep.2021.109133

Banerjee, P., Eckert, A. O., Schrey, A. K., & Preissner, R. (2018). ProTox-II: a webserver for the prediction of toxicity of chemicals. Nucleic Acids Research, 46(W1), W257–W263. https://doi.org/10.1093/nar/gky318

Bhowmik, D., Nandi, R., Jagadeesan, R., Kumar, N., Prakash, A., & Kumar, D. (2020). Identification of potential inhibitors against SARS-CoV-2 by targeting proteins responsible for envelope formation and virion assembly using docking based virtual screening, and

pharmacokinetics approaches. Infection, Genetics and Evolution, 84, 104451. https://doi.org/10.1016/j.meegid.2020.104451

Bouvet, M., Debarnot, C., Imbert, I., Selisko, B., Snijder, E. J., Canard, B., & Decroly, E. (2010). In Vitro Reconstitution of SARS-Coronavirus mRNA Cap Methylation. PloS Pathogens, 6(4), e1000863. https://doi.org/10.1371/journal.ppat.1000863

Bouysset, C., & Fiorucci, S. (2021). ProLIF: a library to encode molecular interactions as fingerprints. Journal of Cheminformatics, 13(1). https://doi.org/10.1186/s13321-021-00548-6

Breitinger, U., Ali, N. K. M., Sticht, H., & Breitinger, H. G. (2021b). Inhibition of SARS CoV Envelope Protein by Flavonoids and Classical Viroporin Inhibitors. Frontiers in Microbiology, 12. <u>https://doi</u>.org/10.3389/fmicb.2021.692423

Burki TK. (2020). The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nature Microbiology, 5(4), 536–544. https://doi.org/10.1038/s41564-020-0695-z

C. Basak, S. (2013). Editorial (Recent Developments and Future Directions at Current Computer Aided Drug Design). Current Computer -- Aided Drug Design, 9(1), 1. https://doi.org/10.2174/157340913804998766

Cao, Y. (2020). Computational Study of the Ion and Water Permeation and Transport Mechanisms of the SARS-CoV-2 Pentameric E Protein Channel. Frontiers. https://www.frontiersin.org/articles/10.3389/fmolb.2020.565797/full

Chakraborty, I., & Maity, P. (2020). COVID-19 outbreak: Migration, effects on society, global environment and prevention. Science of The Total Environment, 728, 138882. https://doi.org/10.1016/j.scitotenv.2020.138882

Chang, C. K., Chen, C. M. M., Chiang, M. H., Hsu, Y. L., & Huang, T. H. (2013). Transient Oligomerization of the SARS-CoV N Protein – Implication for Virus Ribonucleoprotein Packaging. PloS ONE, 8(5), e65045. https://doi.org/10.1371/journal.pone.0065045

Chen, J., Malone, B., Llewellyn, E., Grasso, M., Shelton, P. M., Olinares, P. D. B., Maruthi, K., Eng, E. T., Vatandaslar, H., Chait, B. T., Kapoor, T. M., Darst, S. A., & Campbell, E. A. (2020). Structural Basis for Helicase-Polymerase Coupling in the SARS-CoV-2 Replication-Transcription Complex. Cell, 182(6), 1560–1573.e13. https://doi.org/10.1016/j.cell.2020.07.033

Clark, L. K., Green, T. J., & Petit, C. M. (2021). Structure of Nonstructural Protein 1 from SARS-CoV-2. Journal of Virology, 95(4). https://doi.org/10.1128/jvi.02019-20

Daina, A., Michielin, O., & Zoete, V. (2017). SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Scientific Reports, 7(1). https://doi.org/10.1038/srep42717

Davies, J. P., Almasy, K. M., McDonald, E. F., & Plate, L. (2020). Comparative multiplexed interactomics of SARS-CoV-2 and homologous coronavirus non-structural proteins identifies unique and shared host-cell dependencies. bioRxiv. https://doi.org/10.1101/2020.07.13.201517

Deng, X., Hackbart, M., Mettelman, R. C., O'Brien, A., Mielech, A. M., Yi, G., Kao, C. C., & Baker, S. C. (2017). Coronavirus nonstructural protein 15 mediates evasion of dsRNA

sensors and limits apoptosis in macrophages. Proceedings of the National Academy of Sciences, 114(21). https://doi.org/10.1073/pnas.1618310114

Deng, Z., Chuaqui, C., & Singh, J. (2003). Structural Interaction Fingerprint (SIFt): A Novel Method for Analyzing Three-Dimensional Protein–Ligand Binding Interactions. Journal of Medicinal Chemistry, 47(2), 337–344. https://doi.org/10.1021/jm030331x

Dey, D., Borkotoky, S., & Banerjee, M. (2020). In silico identification of Tretinoin as a SARS-CoV-2 envelope (E)I protein ion channel inhibitor. Computers in Biology and Medicine, 127, 104063. https://doi.org/10.1016/j.compbiomed.2020.104063

DiMasi, J. A., Grabowski, H. G., & Hansen, R. W. (2016). Innovation in the pharmaceutical industry: New estimates of R&D costs. Journal of Health Economics, 47, 20–33. https://doi.org/10.1016/j.jhealeco.2016.01.012

Doman, T. N., McGovern, S. L., Witherbee, B. J., Kasten, T. P., Kurumbail, R., Stallings, W. C., Connolly, D. T., & Shoichet, B. K. (2002). Molecular Docking and High-Throughput Screening for Novel Inhibitors of Protein Tyrosine Phosphatase-1B. Journal of Medicinal Chemistry, 45(11), 2213–2221. https://doi.org/10.1021/jm010548w

Flower, T. G., Buffalo, C. Z., Hooy, R. M., Allaire, M., Ren, X., & Hurley, J. H. (2020). Structure of SARS-CoV-2 ORF8, a rapidly evolving immune evasion protein. Proceedings of the National Academy of Sciences, 118(2). https://doi.org/10.1073/pnas.2021785118

Full length SARS-CoV-2 Nsp2. (2021). RCSB. https://doi.org/10.2210/pdb7msw/pdb

Gadhave, K., Kumar, P., Kumar, A., Bhardwaj, T., Garg, N., & Giri, R. (2021). Conformational dynamics of 13 amino acids long NSP11 of SARS-CoV-2 under membrane mimetics and different solvent conditions. Microbial Pathogenesis, 158, 105041.

Gao, Y., Yan, L., Huang, Y., Liu, F., Zhao, Y., Cao, L., Wang, T., Sun, Q., Ming, Z., Zhang, L., Ge, J., Zheng, L., Zhang, Y., Wang, H., Zhu, Y., Zhu, C., Hu, T., Hua, T., Zhang, B., . . . Rao, Z. (2020). Structure of the RNA-dependent RNA polymerase from COVID-19 virus. Science, 368(6492), 779–782. https://doi.org/10.1126/science.abb7498

Gasmalbari E., Abbadi O. S. Non-Structural Proteins of SARS-CoV-2 as potential sources for vaccine synthesis Infectious Diseases & Tropical Medicine 2020; 6: e667 https://doi.org/10.32113/idtm_202010_667

Goksuluk, D., Korkmaz, S., Zararsiz, G., & Karaagaoglu, A. (2016). easyROC: An Interactive Web-tool for ROC Curve Analysis Using R Language Environment. The R Journal, 8(2), 21<u>3. https://</u>doi.org/10.32614/rj-2016-042

Graves, A. P., Brenk, R., & Shoichet, B. K. (2005). Decoys for Docking. Journal of Medicinal Chemistry, 48(11), 3714–3728. https://doi.org/10.1021/jm0491187

Griffin, S. D. C. (2009). Plugging the holes in hepatitis C virus antiviral therapy. Proceedings of the National Academy of Sciences, 106(31), 12567–12568. https://doi.org/10.1073/pnas.0906760106

Hackbart, M., Deng, X., & Baker, S. C. (2020). Coronavirus endoribonuclease targets viral polyuridine sequences to evade activating host sensors. Proceedings of the National Academy of Sciences, 117(14), 8094–8103. https://doi.org/10.1073/pnas.1921485117

Harcourt, B. H., Jukneliene, D., Kanjanahaluethai, A., Bechill, J., Severson, K. M., Smith, C. M., Rota, P. A., & Baker, S. C. (2004). Identification of Severe Acute Respiratory Syndrome Coronavirus Replicase Products and Characterization of Papain-Like Protease Activity. Journal of Virology, 78(24), 13600–13612. https://doi.org/10.1128/jvi.78.24.13600-13612.2004

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. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell, 181(2), 271–280.e8. https://doi.org/10.1016/j.cell.2020.02.052

Hou, T., & Xu, X. (2004). Recent Development and Application of Virtual Screening in Drug Discovery: An Overview. Current Pharmaceutical Design, 10(9), 1011–1033. https://doi.org/10.2174/1381612043452721

Houston, D. R., & Walkinshaw, M. D. (2013). Consensus Docking: Improving the Reliability of Docking in a Virtual Screening Context. Journal of Chemical Information and Modeling, 53(2), 384–390. https://doi.org/10.1021/ci300399w

Houston, D. R., & Walkinshaw, M. D. (2013). Consensus Docking: Improving the Reliability of Docking in a Virtual Screening Context. Journal of Chemical Information and Modeling, 53(2), 384–390. https://doi.org/10.1021/ci300399w

Hout, D., Mulcahy, E., Pacyniak, E., Gomez, L., Gomez, M., & Stephens, E. (2004). Vpu: A Multifunctional Protein that Enhances the Pathogenesis of Human Immunodeficiency Virus Type 1. Current HIV Research, 2(3), 255–270. https://doi.org/10.2174/1570162043351246

https://doi.org/10.1016/j.micpath.2021.105041

Hu, B., Ge, X., Wang, L. F., & Shi, Z. (2015). Bat origin of human coronaviruses. Virology Journal, 12(1). https://doi.org/10.1186/s12985-015-0422-1

Hui, K. P. Y., Cheung, M. C., Perera, R. A. P. M., Ng, K. C., Bui, C. H. T., Ho, J. C. W., Ng, M. M. T., Kuok, D. I. T., Shih, K. C., Tsao, S. W., Poon, L. L. M., Peiris, M., Nicholls, J. M., & Chan, M. C. W. (2020). Tropism, replication competence, and innate immune responses of the coronavirus SARS-CoV-2 in human respiratory tract and conjunctiva: an analysis in ex-vivo and in-vitro cultures. The Lancet Respiratory Medicine, 8(7), 687–695. https://doi.org/10.1016/s2213-2600(20)30193-4

Imbert, I., Snijder, E. J., Dimitrova, M., Guillemot, J. C., Lécine, P., & Canard, B. (2008). The SARS-Coronavirus PLnc domain of nsp3 as a replication/transcription scaffolding protein. Virus Research, 133(2), 136–148. https://doi.org/10.1016/j.virusres.2007.11.017

Jiang, H. W., Zhang, H. N., Meng, Q. F., Xie, J., Li, Y., Chen, H., Zheng, Y. X., Wang, X. N., Qi, H., Zhang, J., Wang, P. H., Han, Z. G., & Tao, S. C. (2020). SARS-CoV-2 Orf9b suppresses type I interferon responses by targeting TOM70. Cellular & Molecular Immunology, 17(9), 998–1000. https://doi.org/10.1038/s41423-020-0514-8

Kamitani, W., Huang, C., Narayanan, K., Lokugamage, K. G., & Makino, S. (2009). A twopronged strategy to suppress host protein synthesis by SARS coronavirus Nsp1 protein. Nature Structural & Molecular Biology, 16(11), 1134–1140. https://doi.org/10.1038/nsmb.1680 Kern, D. M., Sorum, B., Mali, S. S., Hoel, C. M., Sridharan, S., Remis, J. P., Toso, D. B., Kotecha, A., Bautista, D. M., & Brohawn, S. G. (2021). Cryo-EM structure of SARS-CoV-2 ORF3a in lipid nanodiscs. Nature Structural & Molecular Biology, 28(7), 573–582. https://doi.org/10.1038/s41594-021-00619-0

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., Zaslavsky, L., Zhang, J., & Bolton, E. E. (2020). PubChem in 2021: new data content and improved web interfaces. Nucleic Acids Research, 49(D1), D1388–D1395. https://doi.org/10.1093/nar/gkaa971

Kim, Y., Wower, J., Maltseva, N., Chang, C., Jedrzejczak, R., Wilamowski, M., Kang, S., Nicolaescu, V., Randall, G., Michalska, K., & Joachimiak, A. (2020). Tipiracil binds to uridine site and inhibits Nsp15 endoribonuclease NendoU from SARS-CoV-2. BioRxv. https://doi.org/10.1101/2020.06.26.173872

Krafcikova, P., Silhan, J., Nencka, R., & Boura, E. (2020). Structural analysis of the SARS-CoV-2 methyltransferase complex involved in RNA cap creation bound to sinefungin. Nature Communications, 11(1). https://doi.org/10.1038/s41467-020-17495-9

Lei, J., Kusov, Y., & Hilgenfeld, R. (2018). Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. Antiviral Research, 149, 58–74. https://doi.org/10.1016/j.antiviral.2017.11.001

Li, Q. (2020). Virtual screening of small-molecule libraries. Small Molecule Drug Discovery, 103–125. https://doi.org/10.1016/b978-0-12-818349-6.00004-2

Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings 1PII of original article: S0169-409X(96)00423-1. The article was originally published in Advanced Drug Delivery Reviews 23 (1997) 3–25. 1. Advanced Drug Delivery Reviews, 46(1–3), 3–26. https://doi.org/10.1016/s0169-409x(00)00129-0

Littler, D. R., Gully, B. S., Colson, R. N., & Rossjohn, J. (2020). Crystal Structure of the SARS-CoV-2 Non-structural Protein 9, Nsp9. iScience, 23(7), 101258. https://doi.org/10.1016/j.isci.2020.101258

M. A., Drosten, C., & Pöhlmann, S. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell, 181(2), 271–280.e8. https://doi.org/10.1016/j.cell.2020.02.052

Ma, C., Sacco, M. D., Hurst, B., Townsend, J. A., Hu, Y., Szeto, T., Zhang, X., Tarbet, B., Marty, M. T., Chen, Y., & Wang, J. (2020). Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease. Cell Research, 30(8), 678–692. https://doi.org/10.1038/s41422-020-0356-z

Ma, Y., Wu, L., Shaw, N., Gao, Y., Wang, J., Sun, Y., Lou, Z., Yan, L., Zhang, R., & Rao, Z. (2015). Structural basis and functional analysis of the SARS coronavirus nsp14–nsp10 complex. Proceedings of the National Academy of Sciences, 112(30), 9436–9441. https://doi.org/10.1073/pnas.1508686112

Macalino, S. J. Y., Gosu, V., Hong, S., & Choi, S. (2015). Role of computer- aided drug design in modern drug discovery. Archives of Pharmacal Research, 38(9), 1686–1701. https://doi.org/10.1007/s12272-015-0640-5 Madan, V., Castelló, A., & Carrasco, L. (2007). Viroporins from RNA viruses induce caspase- dependent apoptosis. Cellular Microbiology, 0(0), 071027034427002-??? https://doi.org/10.1111/j.1462-5822.2007.01057.x

Mandala, V. S., McKay, M. J., Shcherbakov, A. A., Dregni, A. J., Kolocouris, A., & Hong, M. (2020). Structure and drug binding of the SARS-CoV-2 envelope protein transmembrane domain in lipid bilayers. Nature Structural & Molecular Biology, 27(12), 1202–1208. https://doi.org/10.1038/s41594-020-00536-8

Mandala, V. S., McKay, M. J., Shcherbakov, A. A., Dregni, A. J., Kolocouris, A., & Hong, M. (2020). Structure and drug binding of the SARS-CoV-2 envelope protein transmembrane domain in lipid bilayers. Nature Structural & Molecular Biology, 27(12), 1202–1208. https://doi.org/10.1038/s41594-020-00536-8

Mangal, M., Sagar, P., Singh, H., Raghava, G. P. S., & Agarwal, S. M. (2012). NPACT: Naturally Occurring Plant- based Anti-cancer Compound-Activity-Target database. Nucleic Acids Research, 41(D1), D1124-D1129.https://doi.org/10.1093/nar/gks1047

Martin, K., & Heleniust, A. (1991). Nuclear transport of influenza virus ribonucleoproteins: The viral matrix protein (M1) promotes export and inhibits import. Cell, 67(1), 117–130. https://doi.org/10.1016/0092-8674(91)90576-k

Michalska, K., Kim, Y., Jedrzejczak, R., Maltseva, N. I., Stols, L., Endres, M., & Joachimiak, A. (2020). Crystal structures of SARS-CoV-2 ADP-ribose phosphatase: from the apo form to ligand complexes. IUCrJ, 7(5), 814–824. https://doi.org/10.1107/s2052252520009653

Mohanraj, K., Karthikeyan, B. S., Vivek-Ananth, R. P., Chand, R. P. B., Aparna, S. R., Mangalapandi, P., & Samal, A. (2018). IMPPAT: A curated database of Indian Medicinal Plants, Phytochemistry and Therapeutics. Scientific Reports, 8(1). https://doi.org/10.1038/s41598-018-22631-z

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. Journal of Computational Chemistry, 30(16), 2785–279<u>1.</u> <u>https://</u>doi.org/10.1002/jcc.21256

Morse, J. S., Lalonde, T., Xu, S., & Liu, W. R. (2020). Learning from the Past: Possible Urgent Prevention and Treatment Options for Severe Acute Respiratory Infections Caused by 2019-nCoV. ChemBioChem, 21(5), 730–738. https://doi.org/10.1002/cbic.202000047

Mysinger, M. M., Carchia, M., Irwin, J. J., & Shoichet, B. K. (2012). Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. Journal of Medicinal Chemistry, 55(14), 6582–6594. https://doi.org/10.1021/jm300687e

Navia, M., Chaturvedi, P. (1996). Design principles for orally bioavailable drugs. Drug Discovery Today, 1(5), 179–189. https://doi.org/10.1016/1359-6446(96)10020-9

Navia, M., Chaturvedi, P. (1996). Design principles for orally bioavailable drugs. Drug Discovery Today, 1(5), 179–189. https://doi.org/10.1016/1359-6446(96)10020-9

Nelson, C. A., Pekosz, A., Lee, C. A., Diamond, M. S., & Fremont, D. H. (2005). Structure and Intracellular Targeting of the SARS-Coronavirus Orf7a Accessory Protein. Structure, 13(1), 75–85. https://doi.org/10.1016/j.str.2004.10.010

Newman, J. A., Douangamath, A., Yadzani, S., Yosaatmadja, Y., Aimon, A., Brandão-Neto, J., Dunnett, L., Gorrie-stone, T., Skyner, R., Fearon, D., Schapira, M., von Delft, F., & Gileadi, O. (2021). Structure, mechanism and crystallographic fragment screening of the SARS-CoV-2 NSP13 helicase. Nature Communications, 12(1). https://doi.org/10.1038/s41467-021-25166-6

Nieto-Torres, J. L., DeDiego, M. L., Álvarez, E., Jiménez-Guardeño, J. M., Regla-Nava, J. A., Llorente, M., Kremer, L., Shuo, S., & Enjuanes, L. (2011). Subcellular location and topology of severe acute respiratory syndrome coronavirus envelope protein. Virology, 415(2), 69–82. https://doi.org/10.1016/j.virol.2011.03.029

Nieva, J. L., Madan, V., & Carrasco, L. (2012). Viroporins: structure and biological functions. Nature Reviews Microbiology,10(8),563-574. https://doi.org/10.1038/nrmicro2820

O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., & Hutchison, G. R. (2011). Open Babel: An open chemical toolbox. Journal of Cheminformatics, 3(1). https://doi.org/10.1186/1758-2946-3-33

Owen, C., Lukacik, P., Strain-Damerell, C., Douangamath, A., Powell, A., Fearon, D., Brandao-Neto, J., Crawshaw, A., Aragao, D., Williams, M., Flaig, R., Hall, D., McAuley, K., Mazzorana, M., Stuart, D., von Delft, F., & Walsh, M. (2020). SARS-CoV-2 main protease with unliganded active site (2019-nCoV, coronavirus disease 2019, COVID-19). RCS<u>B. https://</u>doi.org/10.2210/pdb6yb7/pdb

Peng, Q., Peng, R., Yuan, B., Zhao, J., Wang, M., Wang, X., Wang, Q., Sun, Y., Fan, Z., Qi, J., Gao, G. F., & Shi, Y. (2020). Structural and Biochemical Characterization of the nsp12-nsp7-nsp8 Core Polymerase Complex from SARS-CoV-2. Cell Reports, 31(11), 107774. https://doi.org/10.1016/j.celrep.2020.107774

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera?A visualization system for exploratory research and analysis. Journal of Computational Chemistry, 25(13), 1605–1612. <u>https://doi.org/10.1002/jcc.20084</u>

Ren, Y., Shu, T., Wu, D., Mu, J., Wang, C., Huang, M., Han, Y., Zhang, X. Y., Zhou, W., Qiu, Y., & Zhou, X. (2020). The ORF3a protein of SARS-CoV-2 induces apoptosis in cells. Cellular & Molecular Immunology, 17(8), 881–883. https://doi.org/10.1038/s41423-020-0485-9

Romano, M., Ruggiero, A., Squeglia, F., Maga, G., & Berisio, R. (2020). A Structural View of SARS-CoV-2 RNA Replication Machinery: RNA Synthesis, Proofreading and Final Capping. Cells, 9(5), 1267. https://doi.org/10.3390/cells9051267

Römer, W., Lam, Y. H., Fischer, D., Watts, A., Fischer, W. B., Göring, P., Wehrspohn, R. B., Gösele, U., & Steinem, C. (2004). Channel Activity of a Viral Transmembrane Peptide in Micro-BLMs: Vpu1- 32from HIV-1. Journal of the American Chemical Society, 126(49), 16267–16274. https://doi.org/10.1021/ja0451970

Ropp, P. J., Spiegel, J. O., Walker, J. L., Green, H., Morales, G. A., Milliken, K. A., Ringe, J. J., & Durrant, J. D. (2019). Gypsum-DL: an open-source program for preparing small-molecule libraries for structure-based virtual screening. Journal of Cheminformatics, 11(1). https://doi.org/10.1186/s13321-019-0358-3

Scott, C., & Griffin, S. (2015). Viroporins: structure, function and potential as antiviral targets. Journal of General Virology, 96(8), 2000–2027. https://doi.org/10.1099/vir.0.000201

Shang, B., Wang, X. Y., Yuan, J. W., Vabret, A., Wu, X. D., Yang, R. F., Tian, L., Ji, Y. Y., Deubel, V., & Sun, B. (2005). Characterization and application of monoclonal antibodies against N protein of SARS-coronavirus. Biochemical and Biophysical Research Communications, 336(1), 110–117. https://doi.org/10.1016/j.bbrc.2005.08.032

Sorokina, M., & Steinbeck, C. (2020). Review on natural products databases: where to find data in 2020. Journal of Cheminformatics, 12(1), 0. https://doi.org/10.1186/s13321-020-00424-9

Stützle, Thomas, Exner, Thomas E. (2007). An ant colony optimization approach to flexible protein–ligand docking. Swarm Intelligence, 1, 115-134. https://doi.org/10.1007/s11721-007-0006-9

Surya, W., Li, Y., & Torres, J. (2018). Structural model of the SARS coronavirus E channel in LMPG micelles. Biochimica et Biophysica Acta (BBA–) - Biomembranes, 1860(6), 1309–1317. https://doi.org/10.1016/j.bbamem.2018.02.017

ten Brink, T., & Exner, T. E. (2009). Influence of Protonation, Tautomeric, and Stereoisomeric States on Protein–Ligand Docking Results. Journal of Chemical Information and Modeling, 49(6), 1535–1546. https://doi.org/10.1021/ci800420z

Testa, B., Carrupt, P., Gaillard, P., Billois, F., & Weber, P. (1996). Lipophilicity in molecular modeling. Pharmaceutical Research, 13(3), 335–343. https://doi.org/10.1023/a:1016024005429

Testa, B., Carrupt, P., Gaillard, P., Billois, F., & Weber, P. (1996). Lipophilicity in molecular modeling. Pharmaceutical Research, 13(3), 335–343. https://doi.org/10.1023/a:1016024005429

Tomar, S., Johnston, M. L., St. John, S. E., Osswald, H. L., Nyalapatla, P. R., Paul, L. N., Ghosh, A. K., Denison, M. R., & Mesecar, A. D. (2015). Ligand-induced Dimerization of Middle East Respiratory Syndrome (MERS) Coronavirus nsp5 Protease (3CLpro). Journal of Biological Chemistry, 290(32), 19403–19422. https://doi.org/10.1074/jbc.m115.651463

Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of computational chemistry, 31(2), 455–461. https://doi.org/10.1002/jcc.21334

Tseng, Y. T., Chang, C. H., Wang, S. M., Huang, K. J., & Wang, C. T. (2013). Identifying SARS-CoV Membrane Protein Amino Acid Residues Linked to Virus-Like Particle Assembly. PLoS ONE, 8(5), e64013. https://doi.org/10.1371/journal.pone.0064013

Ujike, M., & Taguchi, F. (2015). Incorporation of Spike and Membrane Glycoproteins into Coronavirus Virions. Viruses, 7(4), 1700–1725. https://doi.org/10.3390/v7041700

V'kovski, P., Kratzel, A., Steiner, S., Stalder, H., & Thiel, V. (2020). Coronavirus biology and replication: implications for SARS-CoV-2. Nature Reviews Microbiology, 19(3), 155–170. https://doi.org/10.1038/s41579-020-00468-6

Viswanathan, T., Arya, S., Chan, S. H., Qi, S., Dai, N., Misra, A., Park, J. G., Oladunni, F., Kovalskyy, D., Hromas, R. A., Martinez-Sobrido, L., & Gupta, Y. K. (2020). Structural

basis of RNA cap modification by SARS-CoV-2. Nature Communications, 11(1). https://doi.org/10.1038/s41467-020-17496-8

Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., & Veesler, D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell, 181(2), 281–292.e6. https://doi.org/10.1016/j.cell.2020.02.058

Wang, C., Li, W., Drabek, D., Okba, N. M. A., van Haperen, R., Osterhaus, A. D. M. E., van Kuppeveld, F. J. M., Haagmans, B. L., Grosveld, F., & Bosch, B. J. (2020). A human monoclonal antibody blocking SARS-CoV-2 infection. Nature Communications, 11(1). https://doi.org/10.1038/s41467-020-16256-y

Wang, C., Takeuchi, K., Pinto, L. H., & Lamb, R. A. (1993). Ion channel activity of influenza A virus M2 protein: characterization of the amantadine block. Journal of Virology, 67(9), 5585–5594. https://doi.org/10.1128/jvi.67.9.5585-5594.1993

Wang, Q., Zhang, Y., Wu, L., Niu, S., Song, C., Zhang, Z., Lu, G., Qiao, C., Hu, Y., Yuen, K. Y., Wang, Q., Zhou, H., Yan, J., & Qi, J. (2020). Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. Cell, 181(4), 894–904.e9. https://doi.org/10.1016/j.cell.2020.03.045

Weeks, S., de Graef, S., & Munawar, A. (2020). X-ray Crystallographic Structure of Orf9b from SARS-CoV-2. RCS<u>B. https://</u>doi.org/10.2210/pdb6z4u/pdb

Wilson, L., Gage, P., & Ewart, G. (2006). Hexamethylene amiloride blocks E protein ion channels and inhibits coronavirus replication. Virology, 353(2), 294–306. https://doi.org/10.1016/j.virol.2006.05.028

Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., Hu, Y., Tao, Z. W., Tian, J. H., Pei, Y. Y., Yuan, M. L., Zhang, Y. L., Dai, F. H., Liu, Y., Wang, Q. M., Zheng, J. J., Xu, L., Holmes, E. C., & Zhang, Y. Z. (2020). Author Correction: A new coronavirus associated with human respiratory disease in China. Nature, 580(7803), E7. https://doi.org/10.1038/s41586-020-2202-3

Ye, Q., West, A. M. V., Silletti, S., & Corbett, K. D. (2020). Architecture and self-assembly of the SARS-CoV -2 nucleocapsid protein. Protein Science, 29(9), 1890–1901. https://doi.org/10.1002/pro.3909

Zhang, Y., Zhang, J., Chen, Y., Luo, B., Yuan, Y., Huang, F., Yang, T., Yu, F., Liu, J., Liu, B., Song, Z., Chen, J., Pan, T., Zhang, X., Li, Y., Li, R., Huang, W., Xiao, F., & Zhang, H. (2020). The ORF8 Protein of SARS-CoV-2 Mediates Immune Evasion through Potently Downregulating MHC-I. Bio Rxiv. https://doi.org/10.1101/2020.05.24.111823

Zheng, Y. X., Wang, L., Kong, W. S., Chen, H., Wang, X. N., Meng, Q., Zhang, H. N., Guo, S. J., Jiang, H. W., & Tao, S. C. (2021). Nsp2 has the potential to be a drug target revealed by global identification of SARS-CoV-2 Nsp2-interacting proteins. Acta Biochimica et Biophysica Sinica, 53(9), 1134–1141. https://doi.org/10.1093/abbs/gmab088

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., Niu, P., Zhan, F., Ma, X., Wang, D., Xu, W., Wu, G., Gao, G. F., & Tan, W. (2020). A Novel Coronavirus from Patients with Pneumonia in China, 2019. New England Journal of Medicine, 382(8), 727–733. https://doi.org/10.1056/nejmoa2001017

Ziebuhr, J., Gorbalenya, A. E., & Snijder, E. J. (2000). Virus-encoded proteinases and proteolytic processing in the Nidovirales. Journal of General Virology, 81(4), 853–879. https://doi.org/10.1099/0022-1317-81-4-853

Lubin, J. H., Zardecki, C., Dolan, E. M., Lu, C., Shen, Z., Dutta, S., Westbrook, J. D., Hudson, B. P., Goodsell, D. S., Williams, J. K., Voigt, M., Sarma, V., Xie, L., Venkatachalam, T., Arnold, S., Alfaro Alvarado, L. H., Catalfano, K., Khan, A., McCarthy, E., . . . Burley, S. K. (2021, 9 octubre). Evolution of the SARS-CoV -2 proteome in three dimensions (3D) during the first 6 months of the COVID -19 pandemic. Proteins: Structure, Function, and Bioinformatics, 90(5), 1054-1080. https://doi.org/10.1002/prot.26250

Lebeau, G., Vagner, D., Frumence, T., Ah-Pine, F., Guillot, X., Nobécourt, E., Raffray, L., & Gasque, P. (2020, August 18). Deciphering SARS-CoV-2 Virologic and Immunologic Features. International Journal of Molecular Sciences, 21(16), 5932. https://doi.org/10.3390/ijms21165932

Kim, Y., Jedrzejczak, R., Maltseva, N. I., Wilamowski, M., Endres, M., Godzik, A., Michalska, K., & Joachimiak, A. (2020, May 2). Crystal structure of Nsp15 endoribonuclease NendoU from SARS-CoV -2. Protein Science, 29(7), 1596–1605. <u>https://doi.org/10.1002/pro.3873</u>

Hata, H., Phuoc Tran, D., Marzouk Sobeh, M., & Kitao, A. (2021). Binding free energy of protein/ligand complexes calculated using dissociation Parallel Cascade Selection Molecular Dynamics and Markov state model. Biophysics and Physicobiology, 18(0), 305–316. https://doi.org/10.2142/biophysico.bppb-v18.037

Carugo, O., & Pongor, S. (2008, December 31). A normalized root-mean-spuare distance for comparing protein three-dimensional structures. Protein Science, 10(7), 1470–1473. <u>https://doi.org/10.1110/ps.690101</u>

Newman, J.A., Imprachim, N., Yosaatmadja, Y., Gileadi, O. (2022) Crystal structure of SARS-CoV-2 NSP14 in complex with 7MeGpppG. <u>10.2210/pdb7QIF/pdb</u>

Zhou Z, Huang C, Zhou Z, Huang Z, Su L, Kang S, Chen X, Chen Q, He S, Rong X, Xiao F, Chen J, Chen S. (2021) Structural insight reveals SARS-CoV-2 ORF7a as an immunomodulating factor for human CD14+ monocytes. iScience., Mar 19;24(3):102187. doi: 10.1016/j.isci.2021.102187

Wang, X., Sun, Z. & Zhou, X. (2022). Crystal structure of the Membrane protein of a SARS-COV-2-related coronavirus. <u>https://www.rcsb.org/structure/7Y9B</u>

Baell, J. B. (2013, 2 diciembre). PAINS: Relevance to Tool Compound Discovery and
Fragment-BasedDiscovery and
PUBLISHING.https://www.publish.csiro.au/ch/CH13551CSIROPUBLISHING.

Landis MS, Bhattachar S, Yazdanian M, Morrison J. Commentary (2018): Why Pharmaceutical Scientists in Early Drug Discovery Are Critical for Influencing the Design and Selection of Optimal Drug Candidates. AAPS PharmSciTech. Jan;19(1):1-10. doi: 10.1208/s12249-017-0849-3.

Wang, N. N., Dong, J., Deng, Y. H., Zhu, M. F., Wen, M., Yao, Z. J., Lu, A. P., Wang, J. B. & Cao, D. S. (2016). ADME Properties Evaluation in Drug Discovery: Prediction of Caco-2 Cell Permeability Using a Combination of NSGA-II and Boosting. Journal of Chemical Information and Modeling, 56(4), 763-773. https://doi.org/10.1021/acs.jcim.5b00642

Lu, J., Zhang, P., Zou, X. W., Zhao, X. Q., Cheng, K. G., Zhao, Y. L., Bi, Y., Zheng, M. Y. & Luo, X. M. (2017). In Silico Prediction of Chemical Toxicity Profile Using Local Lazy Learning. Combinatorial Chemistry & High Throughput Screening, 20(4). https://doi.org/10.2174/1386207320666170217151826

Seidel, T., Wieder, O., Garon, A. & Langer, T. (2020). Applications of the Pharmacophore Concept in Natural Product inspired Drug Design. Molecular Informatics, 39(11), 2000059. <u>https://doi.org/10.1002/minf.202000059</u>