



UNIVERSIDAD NACIONAL AUTÓNOMA DE MEXICO

PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS QUÍMICAS

**ACOPLAMIENTO MOLECULAR DE INHIBIDORES PROPUESTOS A LA
PROTEÍNA PRINCIPAL (M^{pro}) del SARS-CoV-2**

TESIS

PARA OPTAR POR EL GRADO DE

MAESTRO EN CIENCIAS

PRESENTA

Q. Didier Nivón Ramírez

Dr. Rodolfo Gómez Balderas
Facultad de Estudios Superiores Cuautitlán

Ciudad de México, marzo 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.



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS QUÍMICAS

ACOPLAMIENTO MOLECULAR DE INHIBIDORES PROPUESTOS A LA
PROTEÍNA PRINCIPAL (M^{pro}) del SARS-CoV-2

TESIS

PARA OPTAR POR EL GRADO DE

MAESTRO EN CIENCIAS

PRESENTA

Q. DIDIER NIVÓN RAMÍREZ

DR. RODOLFO GÓMEZ BALDERAS



Ciudad de México, marzo 2022

Jurado Asignado

Presidente	Dr. Enrique Ramón Ángeles Anguiano
Vocal	Dr. José Luis Medina Franco
Vocal	Dra. Laura Domínguez Dueñas
Vocal	Dr. José Correa Basurto
Secretario	Dr. Armando Hernández García

El presente trabajo se realizó en el Laboratorio de Química Teórica y Computacional de la Facultad de Estudios Superiores Cuautitlán, Campo 1 de la Universidad Nacional Autónoma de México, bajo la tutela del Dr. Rodolfo Gómez Balderas, con el apoyo de la beca para maestría otorgada por el Consejo Nacional de Ciencias y Tecnología CONACYT (No. De registro becario: 926574).

Agradecimientos

-Agradezco especialmente a mis padres por todo el apoyo que han depositado en mí, a mi madre por su compañía, consejos y desvelos a mi lado. A mi padre por sus enseñanzas y comprensión durante todo este tiempo.

-Agradezco a mi novia (ajolita) por su cariño y apoyo a lo largo de este proyecto, así como su paciencia.

-Agradezco a mis amigos de la universidad por su compañía su apoyo a lo largo de todo este tiempo.

-Agradezco a la Facultad de Estudios Superiores Cuautitlán por el uso de sus instalaciones y equipo, así como de su apoyo en estos tiempos de pandemia.

-Agradezco a mi asesor por el tiempo invertido en su apoyo a la culminación de este proyecto, así como de la oportuna asesoría para los trabajos y materias cursadas.

-Agradezco al CONACYT por su apoyo de beca para el posgrado (No. De registro becario: 926574), la cual me permitió terminar en tiempo y forma mi trabajo de investigación.

-Agradezco al proyecto LANCAD-UNAM-DGTIC-058 de la DGTIC-UNAM por los recursos de cómputo para la realización del proyecto.

-Esta investigación fue realizada gracias al Programa UNAM-PAPIIT IN221621. Agradezco a la DGAPA-UNAM la beca recibida para la obtención del grado de maestría.

Índice

1. Objetivos	6
• Objetivo general	
• Objetivos particulares	
2. Introducción	7
3. Antecedentes	9
• Estudio sobre los coronavirus	
• Proteína principal (M ^{pro}) del SARS-CoV-2	
• Bases de datos y estudio de BIOFACQUIM	
• Coeficientes de partición y cálculo de LogP	
• Estudio de la polaridad en proteínas	
4. Metodología	17
• Estudio computacional de la M ^{pro}	
• Estudio de la polaridad en el sitio activo de la M ^{pro}	
• Análisis de la base de datos BIOFACQUIM	
• Preparación de las estructuras	
• Cribado virtual	
• Selección de estructuras y estudios de dinámica molecular	
5. Resultados y análisis	21
6. Conclusiones	46
7. Referencias	48

Resumen

Con el propósito de investigar posibles estructuras que tengan efecto terapéutico en la proteína principal del SARS-CoV-2 llamada M^{pro}, se determinó el sitio de unión para la proteína, empleando metodologías de búsqueda basadas en diferentes algoritmos y calificaciones de sitio. A partir de los resultados obtenidos, se realizó un análisis de la polaridad y del tamaño del sitio empleando métodos de cálculo basado en química cuántica y modelos de solvatación. Utilizando la base de datos BIOFACQUIM, se seleccionaron aquellas moléculas que cumplieran con los criterios de tamaño y polaridad antes mencionados. Realizando sobre estas, estudios de docking, dinámica molecular y cálculo de energía de unión empleando el modelo MM/GBSA. Los resultados obtenidos muestran a la molécula metilarbutin como un posible candidato a estudios posteriores de optimización en el desarrollo de fármacos.

1 Objetivos

Objetivo General

Identificar inhibidores potenciales para la proteína M^{pro} de SARS-CoV-2, a partir de la búsqueda y selección de estructuras de la base de datos BIOFACQUIM, empleando herramientas de análisis basadas en Química computacional y de diseño de fármacos.

Objetivos particulares

- Analizar la naturaleza química de la proteína principal del SARS-CoV-2 (M^{pro}), identificando los posibles sitios de unión a través del uso de herramientas computacionales.
- Estudiar y analizar la base de datos BIOFACQUIM, seleccionando aquellas moléculas que cumplan con los parámetros de polaridad y tamaño de sitio establecidos.
- Realizar estudios de acoplamiento molecular (*docking*) y dinámica molecular a las moléculas seleccionadas previamente, con el propósito de encontrar compuestos que servirán para estudios posteriores de optimización de fármacos.

2 Introducción

Conforme ha pasado el tiempo, el desarrollo de las grandes urbes y las pobres condiciones de vida que se tienen en ellas, las condiciones han favorecido la aparición de virus y bacterias, lo cual ha generado un efecto negativo en la población mundial.¹ Gran parte de estos organismos, han generado la aparición de enfermedades con gran peligrosidad como las bacterias: salmonela, la bacteria *helicobacter pylori*, la bacteria de la gonorrea o los estafilococos. Además de estas bacterias, la presencia de virus en la población se ha convertido en un problema debido a la peligrosidad de algunos como: el ébola, la rabia, la influenza, el dengue entre otros. Dentro de los virus que afectan a la población, existen algunos que deben tomarse en cuenta como los coronavirus. Estas estructuras llamadas así por la forma que presentan, han tenido una presencia importante en la población desde los años 40s,² sin embargo, no fue sino hasta el año 2003 que estos virus afectaron de manera importante a la población; causando problemas en la respiración e incluso ocasionando varios decesos. Es por ello que los coronavirus que afectan al humano van de la mano con un Síndrome Respiratorio Agudo Grave (SARS, por sus siglas en inglés), nombrándose como: SARS-CoV. Recientemente, en el año 2019 apareció una nueva cepa del SARS-CoV-1 denominada como SARS-CoV-2, la cual provoca la enfermedad de COVID-19. Este virus nuevo encontrado en China, presentó mayor rango de contagio y fuerza al atacar órganos como el pulmón, el estómago, el hígado, el riñón y el cerebro,³ además se encontró que su propagación se observa en tejidos como en los pulmones, boca, intestino y colon entre otros.⁴ Debido a esta alta propagación y la cantidad de receptores en el cuerpo donde la enfermedad puede posicionarse, hasta el momento (2021) se tiene reportado hasta 4 millones y medio de personas fallecidas por esta enfermedad. Como respuesta a esta situación, la investigación de vacunas

basadas en la modificación del ARNm pretende llevar el código genético a la célula y que esta, produzca las proteínas que serán detectadas por el sistema inmune como un agente externo a combatir. Aunque la efectividad de estas vacunas es alta, existen diferentes problemas respecto a la distribución y fabricación de estas vacunas alrededor del mundo. Es por todos estos inconvenientes, que las principales farmacéuticas han centrado sus esfuerzos en desarrollar fármacos que puedan inhibir o desactivar proteínas útiles en el crecimiento o replicación del virus en el organismo. En este trabajo se utilizarán herramientas computacionales para realizar un análisis de la proteína principal o también llamada M^{Pro}, encargada de replicar el virus de SARS-CoV-2 en el organismo, determinando los posibles sitios de unión con moléculas, así como la naturaleza química de los aminoácidos en los sitios. Además, empleando una base de datos mexicana con estructuras químicas encontradas en productos naturales, seleccionar a aquellas moléculas que cumplan con los criterios químicos del sitio. Todo esto desde el punto de vista computacional, empleando herramientas como acoplamiento molecular (*docking*), dinámica molecular y cálculos basados en química cuántica.

3 Antecedentes

3.1 Estudio sobre los coronavirus

De manera general, los coronavirus son un tipo de virus que obtiene su nombre gracias a las glicoproteínas tipo corona que tiene en su exterior (Fig. 1).⁵ Una de sus características principales, es que en su interior se conserva una de las cadenas de ARN más grandes conocidas.⁶ Además, en comparación a otros tipos de virus, este comparte algunas otras propiedades como: el modo de infectar y la replicación en el citoplasma con relación a las membranas celulares.^{6,7}

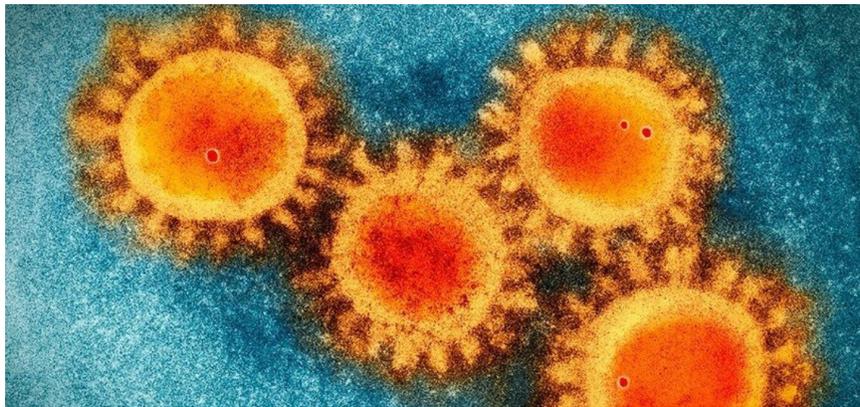


Figura 1. Morfología de los coronavirus vista desde un microscópico.

El interés por la investigación de los coronavirus ha ido en aumento, conforme se presentan nuevas cepas de este. Ya que al existir tan solo 10 genomas registrados en el año 2003, la presencia del nuevo SARS-CoV en el 2008, aumento considerablemente la investigación de este tipo de virus, con la inclusión de 26 genomas más.² Existen alrededor de 7800 artículos científicos relacionados a distintas investigaciones de coronavirus, específicamente a su biodiversidad y aspectos genómicos.⁸ Aunque este tipo de virus tiene mucho tiempo entre nosotros y en distintas variantes, no todas las cepas afectan a los humanos. El primer coronavirus descubierto fue aislado de embriones de pollo en 1937, seguido de otros más en los años 40 incluido el virus de la hepatitis en

ratones.⁹⁻¹¹ Existen al menos 4 grupos de coronavirus (alfa, beta, gamma y delta), delimitados principalmente por los estudios filogenéticos realizados sobre ellos. En este trabajo nos enfocaremos en el SARS-CoV-2, siendo la cepa más reciente y con mayor número de contagios sobre las personas. Este nuevo virus perteneciente al grupo de los β -coronavirus, es de tipo RNA-positivo con una gran similitud al SARS-CoV-1 y el MERS-CoV pertenecientes al mismo grupo.

Una vez que las glicoproteínas espiga (*spike*) se acoplan a las enzimas de angiotensina de tipo 2 o también llamadas ACE2 en el organismo,¹² el RNA del virus se traduce directamente por los ribosomas de la célula huésped codificando 4 proteínas estructurales y 16 no estructurales (NSPs) las cuales llevan a cabo funciones intracelulares (Fig. 2).¹³ Una gran cantidad de estas proteínas han sido estudiadas y se han definido como posibles objetivos para desarrollar potenciales fármacos o antivirales contra el SARS-CoV-2, tenemos el caso del fármaco Remdesivir, el cual inhibe una polimerasa de RNA y ha sido aprobado como tratamiento preventivo para la enfermedad producida por este virus.^{14,15}

Apertura de Genomas y descripción de algunas proteínas

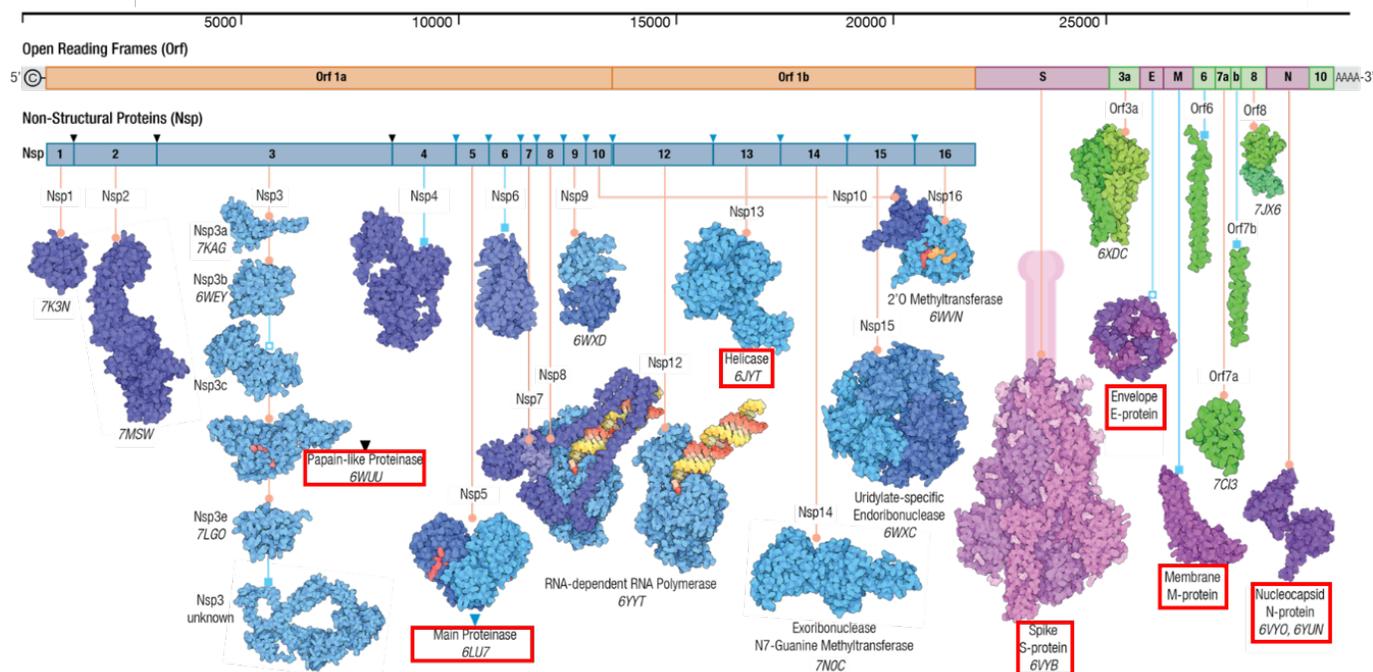


Figura 2. Descripción del genoma del virus SARS-CoV-2 e ilustración de algunas proteínas.

3.2 Proteína principal (M^{pro}) del SARS-CoV-2

Como se mencionó anteriormente, el genoma viral del SARS-CoV-2 consiste en 6 aperturas de lectura (Orf, por sus siglas en inglés), donde el Orf 1 dividido en a y b, decodifican 16 proteínas no estructurales y 4 proteínas estructurales: spike, envelope, membrana y nucleopside.¹³ Cuando el ARN viral entra en la célula huésped, los ribosomas de la célula lo utilizan para producir una cadena polipeptídica, que contiene todas las proteínas del virus. Pero estas moléculas no generan actividad biológica, hasta que la cadena se separa y las proteínas pueden actuar de manera individual. Es por esto que las proteasas enzimáticas juegan un papel fundamental en la separación de las cadenas polipeptídicas y en el proceso de maduración de las proteínas virales.¹⁶ Dentro del genoma del SARS-CoV-2, existen dos proteasas enzimáticas encargadas de este proceso: la proteasa tipo papaína (PL^{pro}) y la proteasa tipo 3C (M^{pro} o 3CL^{pro}). La proteasa principal o

M^{pro} (Figura 3) es una proteasa tipo cisteína, la cual contiene un conjunto catalítico entre la cisteína 145 y la histidina 41 donde el sustrato se une entre los dominios I y II.¹⁷

Las características químicas y la reactividad del conjunto catalítico HIS-CYS, hace de la M^{pro} un objetivo favorable para estudiar moléculas químicas con propiedades bioactivas que puedan inhibir la función biológica de replicación del virus.^{18,19}

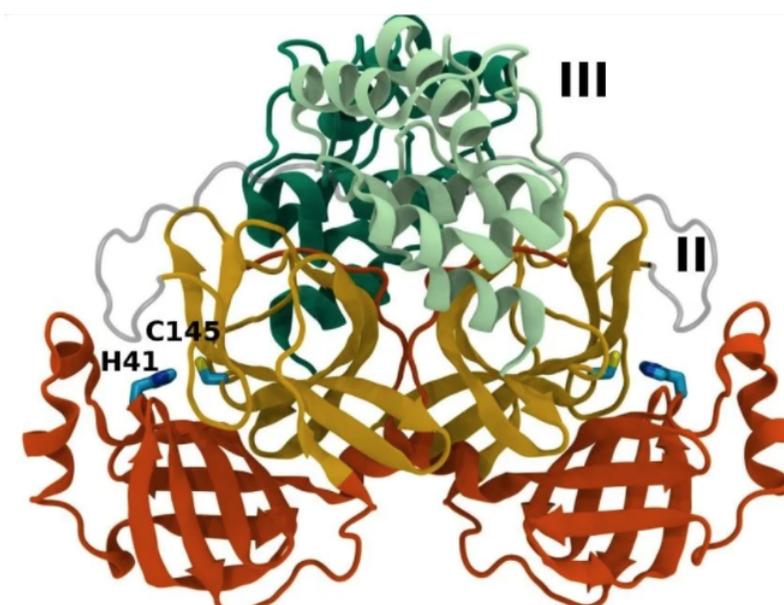


Figura 3. Estructura de la M^{pro} con sus dominios y el conjunto catalítico cys-his.

Derivado de la emergencia sanitaria actual, la investigación computacional relacionada a la búsqueda de inhibidores sobre la M^{pro}, ha crecido a lo largo del tiempo.^{17,20,16,13} Tenemos como ejemplo al inhibidor PF-00835231 desarrollado por Pfizer como inhibidor a la M^{pro} el cual continúa en desarrollo a tal punto, que por ahora se encuentra en ensayos clínicos. También se observa la investigación realizada por la empresa Diamond, en la cual se evaluaron más de 1500 estructuras acopladas, determinando la energía de unión experimental en los complejos.²¹ Existen demasiados artículos científicos de investigación donde relacionan estudios de docking, cribado molecular y dinámica molecular

de diferentes estructuras a la proteína M^{pro}, sin embargo, debido a la emergencia por publicar investigación y propuestas, se han omitido los detalles de la preparación de los ligantes, estados protonados de los aminoácidos contenidos en las proteínas, análisis conformacional de las poses en los cálculos de docking e incluso coherencia química de las interacciones entre los ligantes acoplados y los aminoácidos interactuantes.^{22,23}

3.3 Bases de datos y estudio de BIOFACQUIM

El estudio de la química y el análisis de sus estructuras desde el punto de vista computacional, ha progresado a lo largo de los años.²⁴ Ya sea en la construcción de bases de datos privadas, de tipo gubernamental o perteneciente a universidades. Uno de los principales usos que tienen las bases de datos, son el buscar o proponer estructuras químicas que puedan servir como compuestos líderes en el acoplamiento con alguna estructura objetivo.²⁵

Existe una gran diversidad de bases de datos en función de su contenido, sin embargo, las de mayor popularidad son aquellas en donde las moléculas registradas se encuentran caracterizadas a partir de una planta o producto natural.^{26,27} Dentro de las bases de datos más importantes, con este tipo de características tenemos: la base de datos de productos naturales universales (UNDP),²⁸ la base de datos de china medicinal tradicional (TCM),²⁹ la base de datos brasileña NuBBE_{DB} o la base de datos herbal de Vietnam.^{30,31}

Aunque en México, existe una gran diversidad de plantas y productos naturales que pudieran tener algún tipo de actividad biológica, el registro o caracterización de sus estructuras en bases de datos es limitado. Como respuesta a esta problemática, en la Facultad de Química de la UNAM, se desarrolló la base de datos BIOFACQUIM,³² la cual contiene estructuras caracterizadas y recopiladas de productos naturales. En este trabajo, haremos uso de la base de datos

BIOFACQUIM, como una biblioteca de búsqueda, análisis y selección de estructuras para encontrar compuestos químicos, los cuales puedan presentar actividad biológica sobre la M^{pro}.

3.4 Coeficientes de partición y cálculo de LogP

El coeficiente de partición o coeficiente de reparto es un parámetro fisicoquímico que nos indica la preferencia que tiene una molécula a disolverse en solventes de polaridad baja o alta.³³ Este parámetro se define como el cociente o razón de una sustancia disuelta en distintos solventes que tienen distintas polaridades, la ecuación que lo define es la siguiente:

$$K = \frac{[\text{sustancia}]_1}{[\text{sustancia}]_2} \quad (1)$$

Donde el valor 1 y 2 corresponden a los solventes donde fue disuelta la sustancia. Este valor tiene varias aportaciones al área de la bioquímica, química orgánica o farmacéutica. Específicamente en el área farmacéutica, el cálculo del coeficiente de partición se lleva a cabo empleando como solventes n-octanol y agua, denotando a este valor como P_{ow} . Normalmente este número se da en términos logarítmicos ($\text{Log}P_{ow}$), siendo valores negativos la preferencia por solventes polares o hidrofílicos (agua) y valores positivos la preferencia por solventes con baja polaridad o hidrofóbicos (n-octanol). La importancia de este valor, es que corresponde a un compendio de parámetros fisicoquímicos, los cuales nos permiten predecir las propiedades de administración, distribución, metabolismo y excreción de un fármaco también conocidas como parámetros ADME.³⁴ Dependiendo el valor de $\text{Log}P_{ow}$ obtenido para un fármaco, podemos predecir la posibilidad de que la estructura pueda ser absorbida por la capa lipídica del intestino a través de un proceso llamado transporte transcelular.³⁵

Es por esta razón, que el $\text{Log}P_{\text{ow}}$ debe ser calculado al momento de diseñar fármacos que serán utilizados vía oral. Existen diferentes metodologías experimentales que nos permiten calcular este valor,^{36,37} sin embargo, cuando se analizan bases de datos tan extensas, esta metodología experimental puede resultar muy cara y tardada. Como respuesta a esta problemática, se ha optado por calcular el $\text{Log}P_{\text{ow}}$ utilizando métodos computacionales, esto es, fragmentando la estructura problema y asignando a cada fragmento una contribución al LogP final, dando como resultado un valor de $\text{Log}P_{\text{Promedio}}$. Este método computacional de contribuciones, es útil para determinar el $\text{Log}P_{\text{ow}}$ de muchas estructuras en poco tiempo, sin embargo, el porcentaje de error entre el valor calculado y el obtenido experimentalmente aumenta considerablemente.³⁸ A manera de mejorar los valores de LogP calculados con el método mencionado anteriormente, recientemente, se ha optado por cálculos basados en química cuántica empleando modelos de solvatación.^{36,39,40} La idea fundamental de esta metodología se centra en realizar optimizaciones de geometría, manteniendo la molécula dentro de una cavidad que simula el solvente.⁴¹⁻⁴³ Aunque estos métodos nos permiten determinar energías de solvatación o propiedades fisicoquímicas con mayor exactitud que los métodos empíricos, el costo computacional aumenta sustancialmente conforme la cantidad de electrones aumenta.

3.5 Estudio de la polaridad en Proteínas

Cuando se realizan estudios computacionales sobre una enfermedad, la selección y búsqueda de las proteínas responsables de propagar o acelerar la enfermedad tiene mucha importancia para el diseño de fármacos. Cuando se desea estudiar el sitio de unión, el análisis de la naturaleza química es muy importante, especialmente cuando se desea conocer el grado de encierro o

polaridad de una serie de aminoácidos. Esto nos permite seleccionar moléculas que cumplan con los criterios y puedan acoplarse mejor al sitio. Para determinar estos valores en proteínas, se pueden emplear distintas técnicas experimentales como HPLC⁴⁴ o análisis de rayos X,⁴⁵ sin embargo, la precisión de la medición dependerá principalmente de la pureza de la muestra y la cantidad de proteína aislada, lo cual puede llegar a resultar caro y en algunos casos muy difícil de conseguir. Como respuesta a esta problemática, desde el punto de vista computacional, el análisis de la polaridad para proteínas puede llevarse a cabo utilizando el método de Rekker.⁴⁶ Este método consiste en evaluar la polaridad en función de los fragmentos (f_n) que contiene cada aminoácido, multiplicando por el número de veces que este fragmento aparece (a_n). La suma de sus fragmentos y su frecuencia de aparición (Ecuación 2), nos permite determinar el coeficiente de partición por cada aminoácido.

$$\log(P) = \sum_{n=1}^N a_n f_n \quad (2)$$

Aunque este método es utilizado como posible respuesta a la polaridad de las moléculas, sus valores pueden no acercarse al número real o su porcentaje de error es apreciable. Hoy en día existen softwares que nos permiten realizar evaluaciones de tipo cualitativas sobre el grado de hidrofobicidad en un sitio, así como la obtención de otros descriptores químicos. Sin embargo, también la metodología computacional basada en cálculos cuánticos se puede considerar como una opción favorable para la obtención de parámetros de polaridad bastante acertados, aunque caros computacionalmente.

4 Metodología

4.1 Estudio computacional de la M^{pro}

Con el propósito de determinar la naturaleza química de los aminoácidos que conforman a la proteasa M^{pro} y así detectar los posibles sitios de unión que existen en ella, se utilizó el software SiteMap,⁴⁷ perteneciente a la suite computacional Maestro. Para comparar los resultados obtenidos, se utilizó un software de acceso libre llamado POCASA,⁴⁸ el cual también realiza un análisis de la proteína y determina el sitio con mayor afinidad a unirse un fármaco.

4.2 Estudio de la polaridad en el sitio activo de la M^{pro}

Una vez seleccionado el sitio de unión propuesto a partir de los cálculos de búsqueda anteriores, se realizó un análisis de la polaridad de los aminoácidos que conforman la cavidad a través del cálculo del coeficiente de partición en octanol-agua ($\text{Log}P_{\text{ow}}$) para cada aminoácido. Esto se realizó, haciendo una serie de cálculos basados en química cuántica con el software TURBOMOLE⁴⁹ bajo la aproximación de la teoría de funcionales de la densidad (DFT) y la aproximación de solventes reales COSMO.^{42,50} Los cálculos se llevaron a cabo utilizando el funcional BP86WVN y la base de cálculo Def-TZVP,⁵¹ calculando a partir de los resultados obtenidos, las energías de solvatación con el software COSMOtherm, recopilando estos valores de energía y las apariciones del mismo aminoácido en el sitio (frecuencias de aparición) en una tabla. Con los resultados obtenidos, se calculó un $\text{Log}P_{\text{ow}}$ promedio que será el valor que describe la polaridad general del sitio.

4.3 Análisis de la base de datos BIOFACQUIM

Para encontrar las posibles estructuras químicas que pudieran tener un efecto inhibitorio sobre la M^{pro}, se realizó la descarga de la base de datos BIOFACQUIM y utilizando el software DataWarrior,⁵² se calcularon los siguientes parámetros fisicoquímicos: el LogP, peso molecular, donadores y aceptores de puentes de hidrógeno (HBD y HBA) así como el volumen de las estructuras. Con estos datos, se realizó un análisis de la diversidad de la base, así como la dispersión de las moléculas en el espacio químico en función de las propiedades químicas antes descritas.

4.4 Preparación de las estructuras

Se realizó una separación de estructuras, empleando los criterios de volumen y $\text{Log}P_{\text{ow}}$ promedio para el sitio de unión, calculados en el apartado 4.1 y 4.2 respectivamente. La preparación de las estructuras resultantes, se realizó con el software LigPrep,⁵³ estableciendo los estados protonados con el software Epik.^{54,55}

4.5 Cribado virtual

Se realizó un estudio de acoplamiento molecular (*docking*) con las estructuras preparadas en el apartado 4.4 empleando el módulo Glide^{56,57}. Las conformaciones resultantes por cada ligante fueron analizadas, recopilando las interacciones más importantes de las estructuras con los aminoácidos del sitio. Además de la comparación de la energía de afinidad, respecto a inhibidores reportados en la literatura previamente.^{58,59}

4.6 Selección de estructuras y estudios de dinámica molecular

De las moléculas analizadas en el apartado 4.5, se seleccionaron a aquellas moléculas que tuvieran las mayores energías de afinidad e incluyeran interacciones con aminoácidos del sitio considerados importantes. Posteriormente se realizaron estudios de dinámica molecular, utilizando el software de simulación Desmond.⁶⁰ empleando el *forcefield* OPLS4,^{61,62} manteniendo el volumen y temperatura del sistema con un ensamble NVT, empleando una temperatura y tiempo de simulación de 300 K y 100 ns respectivamente. De los resultados obtenidos, se analizó la estabilidad del ligante en el sitio a partir del cálculo de la desviación cuadrática media (RMSD, por sus siglas en inglés).

Diagrama de Flujo



5 Resultados y análisis

5.1 Estudio computacional de la M^{pro}

El análisis químico y estructural de las proteínas, tienen una gran relevancia cuando se desea conocer los sitios o zonas donde se pueda inhibir su función y así detener el progreso de la enfermedad. La metodología de búsqueda basada en puntos de sitio y sustentada bajo el método de mallas de Goodford.⁶³ Permite evaluar la superficie de la proteína, encontrando aquellos sitios en donde existan cavidades, así como una breve descripción de la naturaleza química de los aminoácidos en el sitio. Esto a través de descriptores químicos como: aceptores de puentes de hidrógeno (rojo), donadores de puentes de hidrógeno (azul) o zonas en donde el grado de encierro entre los aminoácidos permite formar zonas hidrofóbicas (amarillo).^{64,65}

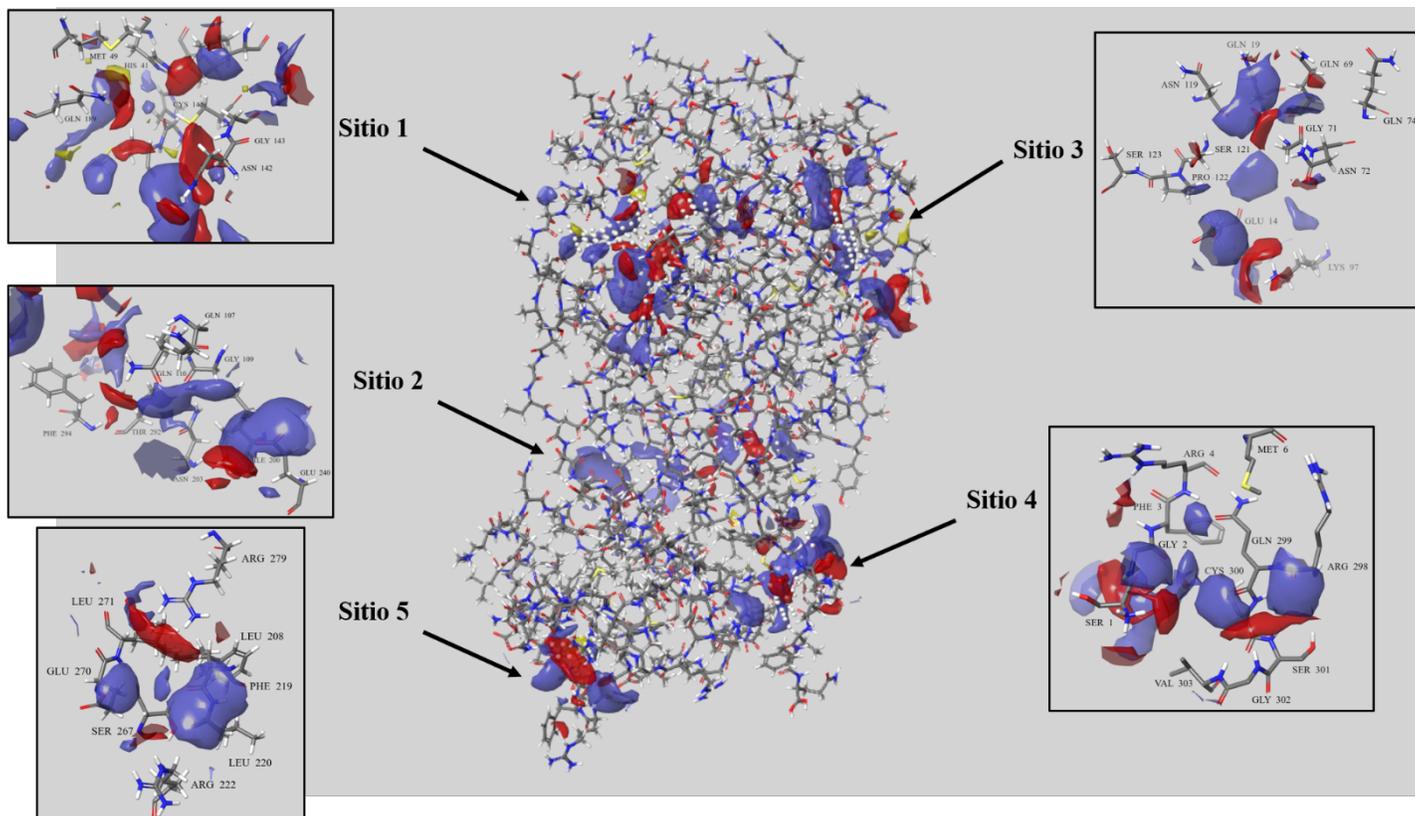


Figura 3. Protómero de la proteína M^{pro} analizada con el *software* SiteMap.

Durante la selección del sitio donde pueda unirse un ligante, es importante considerar los siguientes parámetros: el tamaño del sitio determinado ya sea mediante herramientas computacionales o reportes experimentales, los puntos farmacóforos descritos computacionalmente o reportados de determinaciones cristalográficas, el grado de encierro de la cavidad o la exposición al solvente que tiene el sitio.^{66,67} Otro parámetro a considerar en la selección de un sitio viable para unir moléculas, son los reportes experimentales que indiquen la existencia de aminoácidos susceptibles o prioritarios para que la proteína o sistema cumpla con su función biológica. Los resultados obtenidos de SiteMap mostrados en la Fig.3 nos muestran 5 sitios de unión posibles, cada uno con una calificación obtenida a partir de la cantidad de descriptores químicos descritos anteriormente. Si observamos el sitio No. 5 y 4, nos damos cuenta de que nos encontramos en los extremos del protómero de la M^{Pro}, esta zona se encuentra altamente expuesta al solvente, por lo que su cavidad además de tener carácter hidrofílico, es de tamaño pequeño. Aunque el software SiteMap propone estos sitios, no se consideran como lugares viables para la unión por la pequeña profundidad de las cavidades lo cual se ve reflejado en la calificación otorgada por el software a cada uno de los sitios (calf. 0.60), la misma situación la observamos para el sitio No. 3 (calf. 0.61), donde observamos que su ubicación en la zona superior, no es favorable para un acoplamiento, esto debido al tamaño pequeño de la cavidad formada (101.53 Å³) y la presencia mínima de zonas hidrofóbicas. Con respecto al sitio No. 2 (Fig.4), ubicado a la mitad del protómero y con un tamaño de 251.42 Å³, tiene una calificación más alta que los sitios reportados anteriormente (0.77). Sin embargo, aunque el tamaño de cavidad es mayor, los descriptores químicos reportados en el sitio indican una nula zona hidrofóbica. A diferencia del sitio No. 2, el sitio No. 1 con una profundidad de cavidad de 293.26 Å³ y una calificación de 0.98 contiene en su interior algunos aminoácidos importantes previamente reportados como

potenciales puntos farmacóforos: CYS145, HIS41, MET145, GLY143, GLU166 y GLN189.^{19,58,68-70} Esta información complementaria, nos permite asignar al sitio No. 1 como el lugar viable para realizar estudios de *docking* en el interior de la cavidad para nuestras moléculas seleccionadas. La importancia de los estudios computacionales en la búsqueda de sitios en proteínas, es la posibilidad de predecir zonas en la proteína con mayor precisión en sistemas donde no se tiene una investigación experimental previa. Esto permite avanzar a nivel teórico el desarrollo de fármacos que puedan acoplarse eficientemente a las proteínas de la enfermedad estudiada.

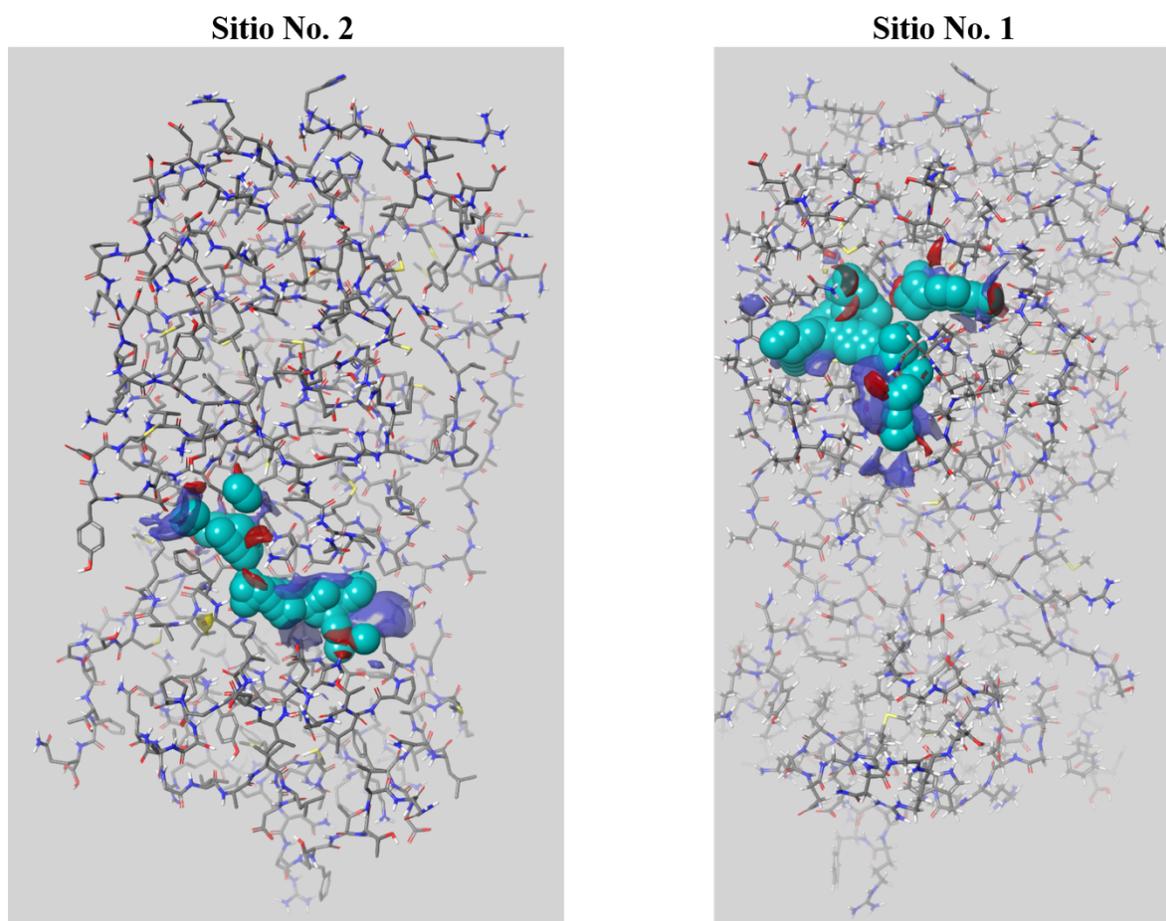


Figura 4. Distribución del Sitio 1 y 2 en el protómero de la M^{pro}, calculados con el *software* SiteMap.

En la Tabla 1 se muestran las calificaciones asignadas a los sitios reportados por el *software* SiteMap. La teoría detrás se basa en generar una malla de puntos de sitio dentro de las cavidades de la proteína, asignando una calificación al sitio con la siguiente ecuación:

$$\text{SiteScore} = 0.0733^{2(n)} + 0.6688e - 0.20p \quad (3)$$

Donde el valor n es el número de puntos de sitio (máximo 100), el número e corresponde al grado de encierro y p es la puntuación relacionada a la hidrofobicidad con un límite de 1 para sitios cargados o altamente polares. Se ha demostrado que a partir de calificaciones como 0.8 en adelante se puede diferenciar entre sitios de unión prometedores y los que no lo son.^{65,71}

Tabla 1. Calificaciones obtenidas para cada sitio obtenido para la Mpro

No. de Sitio	Calificación	Tamaño de sitio Å ³
Sitio 1	0.98	293.26
Sitio 2	0.77	251.42
Sitio 3	0.61	101.53
Sitio 4	0.60	80.95
Sitio 5	0.60	68.26

Con el propósito de comparar la búsqueda computacional de SiteMap, se realizó el análisis de la M^{pro} empleando otro *software* de detección llamado POCASA.⁴⁸ Esta herramienta computacional, emplea puntos de malla para rellenar cada posible cavidad de la proteína objetivo. Posteriormente asigna una calificación a las cavidades, realizando el cálculo de la profundidad de la cavidad (VD , por sus siglas en inglés) empleando la siguiente ecuación

$$VD_i = \sum_{j=1}^N d(g_j), \quad g_j \in P_i \quad (4)$$

Donde $d(g_j)$ es la profundidad del punto de sitio g_j y N es el número de puntos en el sitio P_i .⁴⁸ El alcance de esta metodología computacional, permite determinar sitios donde pueda existir una posible unión con fármacos, llegando incluso a predecir la disposición geométrica que el ligante va a adquirir en el sitio.⁴⁸ En la Fig. 5 se observa la presencia de 3 sitios de unión posibles, los cuales coinciden con el sitio 1, 3 y 4 obtenidos por el *software* SiteMap. Los resultados obtenidos con el cálculo hecho por POCASA, asignan como mejor sitio de unión a la zona donde se ubican los aminoácidos: MET145, GLY143, GLU166 y GLN189, mostrando una coincidencia con los aminoácidos reportados previamente por SiteMap para el sitio No. 1.

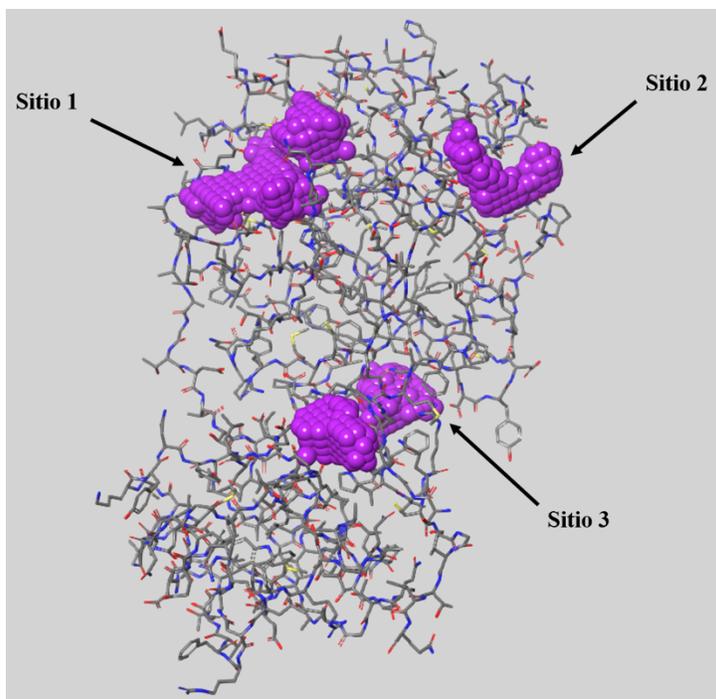


Figura 5. Protómero de la proteína M^{pro} analizada con el software POCASA

Como resultado del mapeo por parte de SiteMap y POCASA, se recopilaron los volúmenes calculados para el sitio que tuvo la mayor calificación, obteniendo un valor de 293.26 Å³ con el *software* SiteMap y un valor de 216 Å³ con el *software* POCASA. A manera de visualizar a detalle la composición de

aminoácidos en el sitio, en la Figura 6 se muestran los aminoácidos más importantes en el sitio de elección.

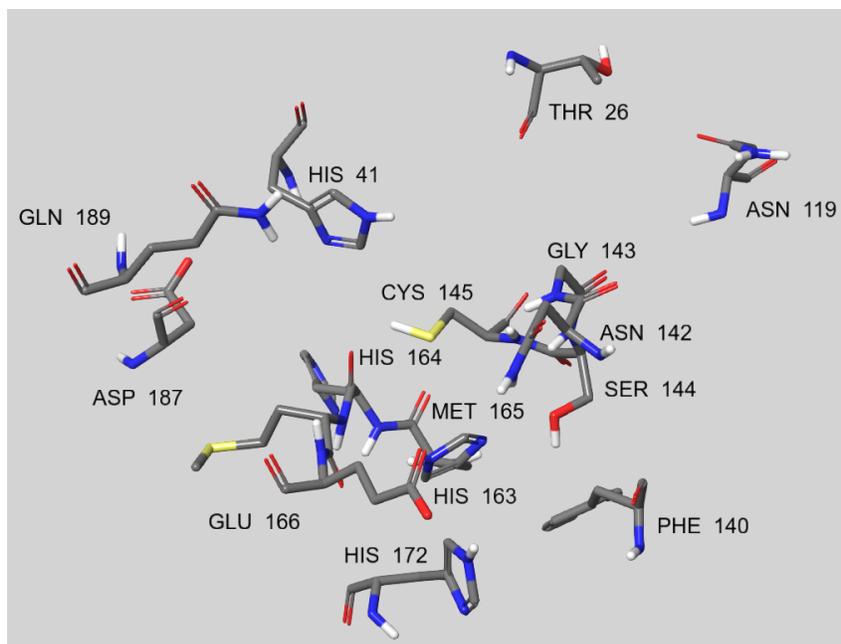


Figura 6. Aminoácidos principales en el sitio reportado por SiteMap y POCASA.

Como se puede observar en la Figura 6, gran parte de los aminoácidos presentes en el sitio tienen átomos que confieren un carácter polar en el área, lo cual influye en la selección de moléculas que puedan cumplir con las propiedades del sitio. Para determinar la polaridad o grado de hidrofobicidad que tiene el sitio, se necesitan herramientas computacionales que determinen la naturaleza de los aminoácidos, promediando una polaridad en el sitio.

5.2 Estudio de la polaridad en el sitio activo de la M^{pro}

Como se observa en la Figura 6, existe una gran variedad de aminoácidos que componen a la cavidad de unión para la M^{pro} . La naturaleza química de los aminoácidos presentes en el sitio, presentan una polaridad específica que favorece su solubilidad en presencia de solventes polares o no polares. La afinidad de esta solubilidad mencionada, se conoce a través del coeficiente de reparto denotado por LogP_{ow} . Para calcular este valor, se utilizan metodologías

computacionales basadas en modelos de solvatación, en donde la molécula objetivo se rodea de una cavidad que simula al solvente.⁷²⁻⁷⁶ Este tipo de modelos “implícitos”, tienen gran importancia al momento de determinar algunas propiedades de mezclado, pues nos permite evaluar diferentes solventes a través de modificar la constante dieléctrica del sistema. Empleando el modelo de solvatación basado en química cuántica para solventes reales conocido como COSMO-RS^{77,78,76} y el software TURBOMOLE, se realizaron cálculos de optimización de geometría a los aminoácidos del sitio (Fig. 6). Obteniendo en cada ciclo de optimización, la distribución de carga generada alrededor de la molécula conocida como perfil sigma⁷⁶. Utilizando la energías de solvatación en octanol agua resultantes (ΔG_o y ΔG_w) se calcularon los valores de $\text{Log}P_{ow}$ utilizando las ecuaciones de Kundi y Ho,⁴¹ como se muestran:

$$\Delta\Delta\text{transf} = \Delta G_o - \Delta G_w$$

$$\text{Log}P_{ow} = \frac{-\Delta\Delta\text{transf}}{RT \ln(10)} \quad (5)$$

Como se observa en la Figura 7, los perfiles de carga calculados para algunos aminoácidos se muestran como una malla de colores, donde el color rojo representa la presencia de electrones o sustituyentes donadores de puentes de hidrógeno, por otro lado, en color naranja se observan zonas donde la presencia de electrones es menor. Finalmente, en color verde se muestran las zonas donde la presencia de electrones es nula, lo cual podría representar zonas hidrofóbicas.

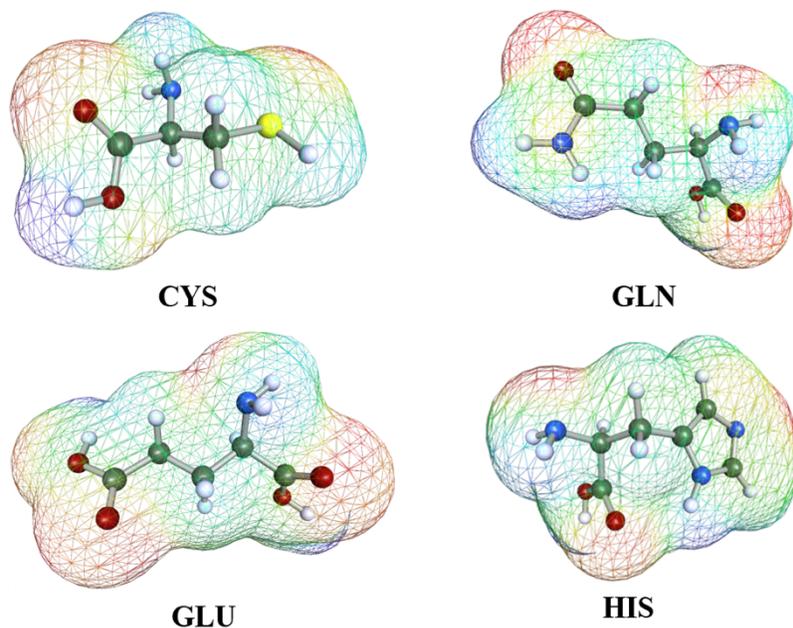


Figura 7. Aminoácidos calculados bajo la aproximación COSMO, mostrando en mallas de color su perfil de densidad.

En la tabla 2 se muestran las energías de solvatación calculadas para los solventes octanol agua, así como el valor de LogP_{ow} por cada aminoácido y la frecuencia de aparición de cada aminoácido en el sitio. Como se puede observar, gran parte de los valores de LogP_{ow} resultantes son en su mayoría negativos, lo cual nos indicaría que gran parte del sitio tiene preferencia por solventes polares o moléculas con sustituyentes polares.

Tabla 2. Valores de LogP_{ow} calculados para los aminoácidos presentes en el sitio de la M^{pro} , así como la frecuencia de aparición.

Aminoácido	ΔG_{w} kJ/mol	ΔG_{o} kJ/mol	$-\Delta\Delta\text{transf}$ kJ/mol	LogP_{ow}	Frecuencia de aparición
GLN	-73.90	-57.34	16.56	-2.90	1
ASP	-65.37	-52.29	13.08	-1.80	1
HIS	-66.37	-62.28	4.09	-1.25	4
THR	-59.98	-52.12	7.86	-1.38	1
GLU	-70.09	-59.56	10.53	-1.84	1
CYS	-43.55	-41.25	2.3	-0.41	1
GLY	-46.81	-37.03	9.78	-1.74	1
ASN	-65.37	-52.29	13.08	-2.30	2
MET	-41.17	-44.60	-3.43	-1.84	1
SER	-56.38	-46.81	9.57	-1.37	1
PHE	-55.38	-55.17	0.21	-0.02	1

Conociendo la cantidad de veces que los aminoácidos aparecen (frecuencia de aparición) y los valores de LogP_{ow} calculados bajo esta teoría, se calculó un LogP_{ow} promedio denotado como $\text{LogP}_{\text{promedio}}$ que considere todas las medidas de los aminoácidos así como su frecuencia de aparición. Encontrando un valor de $\text{LogP}_{\text{promedio}} = -1.4$, observamos la preferencia del sitio por moléculas polares o con alta afinidad al solvente (agua). Es importante recordar que este valor pretende aproximar la naturaleza química del sitio, sin considerar como influye en la polaridad el grado de encierro y la conformación estructural de los aminoácidos. Como se observa en la Fig. 8, en color verde se reporta el mapeo de las zonas hidrofílicas en el sitio calculadas por SiteMap. Como se puede apreciar, en gran parte del sitio persisten aminoácidos con sustituyentes donadores de puentes de hidrógeno, así como una visible exposición del sitio al solvente. Tomando en cuenta estas observaciones, podemos considerar que aquellas moléculas que se deseen acoplar al sitio, deberán encontrarse en un valor de LogP_{ow} comprendido en el intervalo de -1.4 a 0.

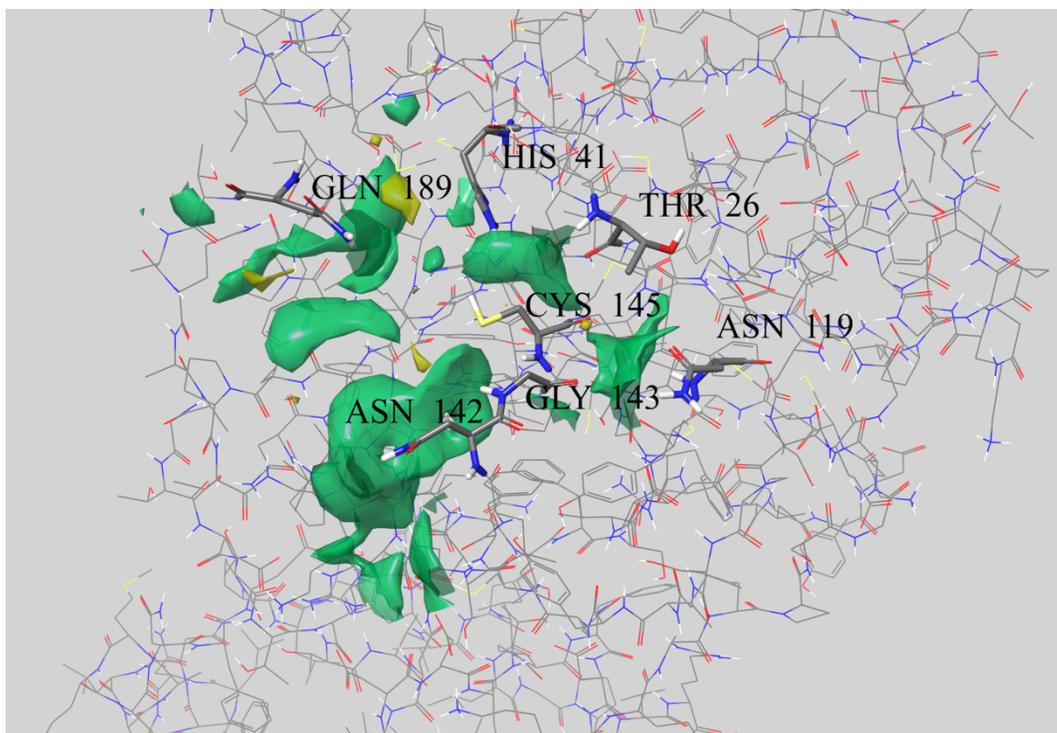


Figura 8. Descripción de la naturaleza química del sitio catalítico de la M^{pro}. En color verde se observan las zonas hidrofílicas del sitio.

5.3 Análisis de la Base de Datos BIOFACQUIM

Las bases de datos químicas, representan un compendio de estructuras moleculares que contienen distintas propiedades fisicoquímicas. Estos valores nos sirven para delimitar a aquellas estructuras que cumplan con los criterios de estudio. La base de datos BIOFACQUIM fue recopilada de una búsqueda exhaustiva de artículos científicos mexicanos en donde se describiera a detalle el proceso de aislamiento, purificación y proceso de caracterización de cada una de las moléculas. Como se describe en el artículo de BIOFACQUIM,³² la base de datos está compuesta de 423 compuestos químicos, donde 316 fueron aislados de 49 diferentes tipos de plantas, 98 fueron aislados de 19 tipos de hongos y 9 de propóleos mexicanos. Con respecto a las propiedades fisicoquímicas contenidas en las moléculas, se menciona que el peso molecular tiene una media entre 340.5 a 412 g/mol (Fig. 9a) con un parecido cercano a la base de datos brasileña NuBBE_{DB}.³⁰ Como se observa en la Fig. 9b, los valores

de LogP_{ow} calculados en DataWarrior para las estructuras de BIOFACQUIM⁷⁹ muestra que el intervalo de valores se ubica entre -0.3 a 5, indicando un carácter hidrofóbico por parte de la mayoría de moléculas en concordancia a lo reportado.³²

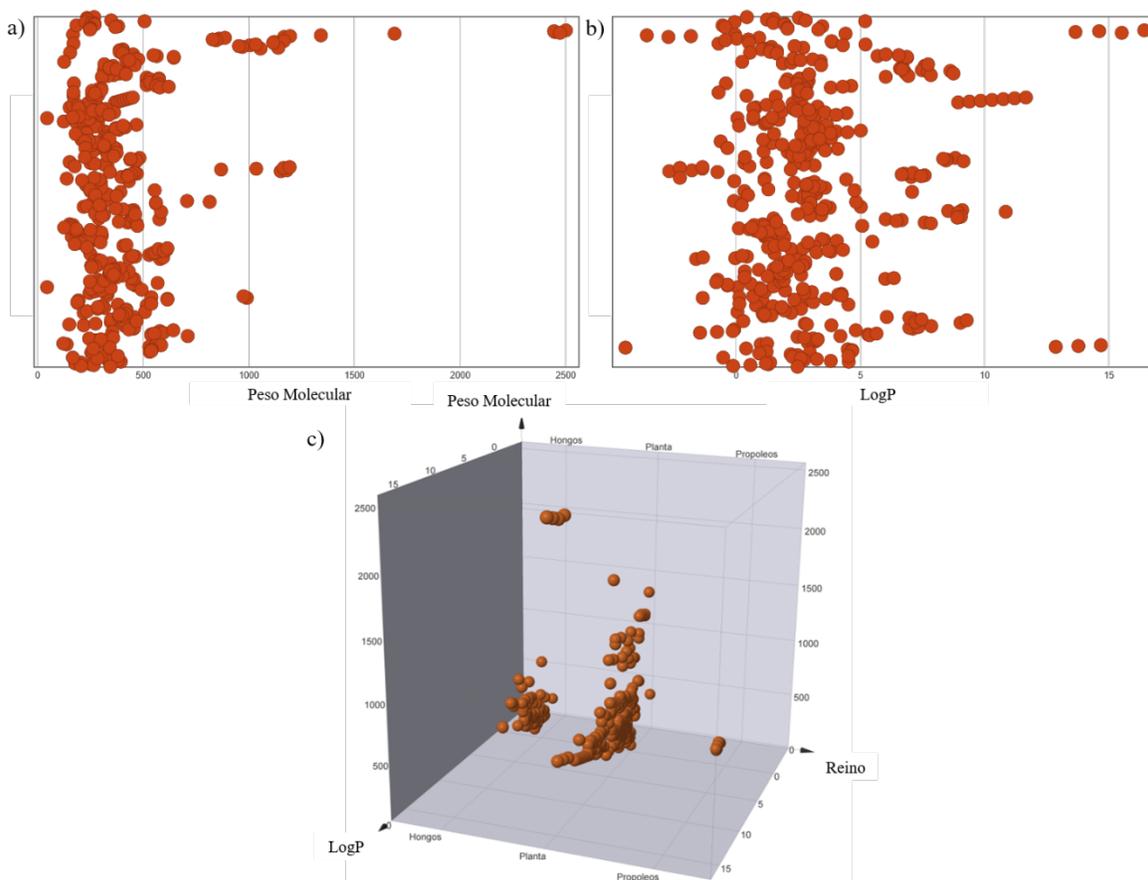


Figura 9. (a) Distribución de las moléculas en función del peso molecular. (b) Distribución de las moléculas en función del valor de LogP_{ow} . (c) Gráfico que nos indica la diversidad de las moléculas en función del organismo.

5.4 Preparación de las estructuras

Una vez conocida de manera general la composición de la base de datos BIOFACQUIM, procederemos a aplicar los criterios de selección a las estructuras con el propósito de delimitar la base de datos y trabajar con las estructuras restantes. Recordemos que el primer criterio de selección corresponde al tamaño de sitio calculado por SiteMap y POCASA, indicando que solo aquellas moléculas que tengan un volumen igual o menor a 254.63 \AA^3 pueden acoplarse en el sitio. El segundo criterio de selección relacionado a polaridad del sitio calculada a través de la aproximación COSMO, nos indica que aquellas moléculas que se encuentren en el intervalo de LogP_{ow} entre -1.4 a 0, podrán acoplarse correctamente en el sitio. Como resultado de aplicar estos criterios de selección, se encontraron 8 moléculas cumplieron con los criterios de selección (Fig. 10). Como se puede observar, las moléculas filtradas contienen en su estructura una cantidad considerable de átomos de oxígeno y nitrógeno, responsables de aportar el carácter hidrofílico a la molécula, así como favorecer una mayor cantidad de interacciones tipo puentes de hidrógeno con los aminoácidos del sitio seleccionado.

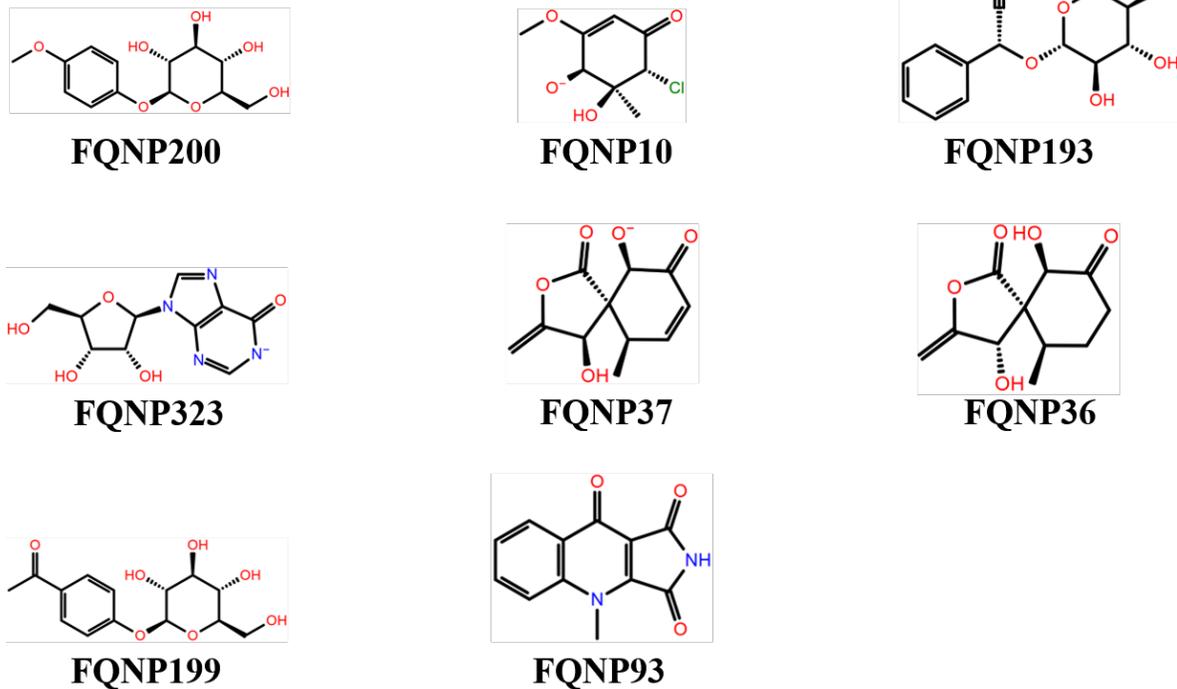


Figura 10. Estructuras resultantes de los criterios propuestos.

Algunos parámetros fisicoquímicos de las estructuras seleccionadas (Fig. 8), se muestran en la Tabla 3

Tabla 3. Propiedades fisicoquímicas de las estructuras seleccionadas

Estructura	Peso Molecular (g/mol)	LogP _{ow}	Volumen Å ³
FQNP200	287.28	-0.745	264.4
FQNP10	206.62	-0.124	180.3
FQNP193	295.29	-1.242	287.9
FQNP323	268.22	-2.266	234.3
FQNP37	224.21	-0.625	199.9
FQNP36	226.22	-0.350	204.7
FQNP199	298.29	-0.804	282.3
FQNP93	228.21	-0.065	198.5

5.5 Cribado virtual

Para estudiar en términos aproximados el posicionamiento a nivel molecular de una estructura dentro de una macroestructura como una proteína, la herramienta computacional de acoplamiento o *docking* ha sido ampliamente utilizada en la investigación de nuevas moléculas y sus interacciones con proteínas objetivo.⁸⁰ La metodología de docking consiste en predecir la posición de la estructura en el sitio o también llamada pose, así como el cálculo relacionado con la energía de unión entre el ligante y la proteína.⁸¹ El fundamento del docking se encuentra en la teoría de “llave y herradura” de Fischer,⁸² considerando que las moléculas se encuentran rígidas y no se considera a detalle la naturaleza química de las especies, es decir un *docking* rígido.⁸⁰ En este trabajo, empleando el software Glide,^{56,57,83} se realizaron estudios de *docking* entre las moléculas seleccionadas en el apartado 5.4 y la proteína M^{pro}, considerando la aproximación de Koshland.^{84,85} Esta aproximación también conocida como *docking* flexible, permite un mayor acomodo entre el ligante y el receptor a través de mover ambas especies descritas por un campo de fuerza previamente seleccionado.^{86–}⁸⁹ Este cálculo presenta una mayor precisión a la posible disposición del ligante en el sitio, sin embargo computacionalmente es más tardado dependiendo de la cantidad de moléculas que se deseen acoplar, por lo que el filtrado de moléculas realizado anteriormente es muy importante. Como se muestra en la Fig. 11, el acoplamiento de la Prunasina (FQNP193)⁹⁰ con el protómero M^{pro} nos muestra interacciones principales con los aminoácidos: GLU166, ASN142, SER144 y GLY143, donde la mayor parte de estas son puentes de hidrógeno, además es importante observar la presencia de interacciones tipo $\pi-\pi$ entre el sustituyente fenilo y el aminoácido HIS41 perteneciente a los aminoácidos importantes en el sitio de unión. Este tipo de interacciones tienen gran importancia al momento de estudiar posibles inhibiciones a una proteasa o

proteína con función específica, ya que la idea principal es evitar que su función biológica se lleve a cabo a través de bloquear espacialmente el complejo catalítico que lleva a cabo la función.⁹¹

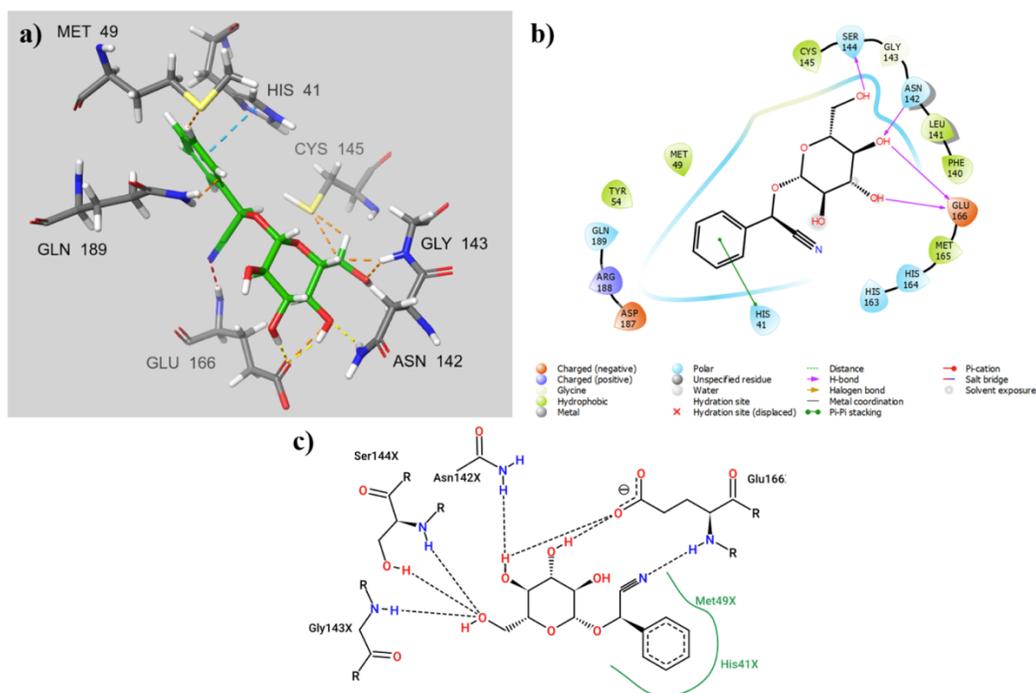


Figura 11. (a) Mapa de interacciones 3D de la molécula Prunasina a la M^{PRO}. (b) Diagrama de interacción 2D obtenida con Maestro. (c) Diagrama de interacción obtenido con ProteinsPlus.

Como resultado de los estudios de docking aplicados sobre las moléculas seleccionadas, en la Tabla 4 se recopilan la cantidad de veces que un aminoácido aparece en el sitio (frecuencia de aparición), así como el tipo de interacción que se está llevando a cabo. Como se observa, la mayoría de las interacciones son puentes de hidrógeno (HB, por sus siglas en inglés) entre los ligantes y los aminoácidos: GLU166, GLY143, GLN189 y SER144. Es importante mostrar que la HIS41 perteneciente a los aminoácidos importantes del sitio de la M^{PRO} presenta dos tipos de interacciones con las moléculas acopladas, la primera como puente de hidrógeno y la segunda como una interacción de tipo $\pi - \pi$.

Tabla 4. Conteo de frecuencias para las interacciones por aminoácido de las estructuras acopladas a la proteína M^{pro}.

Aminoácido	Frecuencia de aparición	Tipo de interacción
ASP187	2	HB
ASN142	2	HB
HIS41	4,1	$\pi - \pi$, HB
GLU166	14	HB
GLY143	7	HB
SER144	4	HB
CYS145		
GLN189	5	HB
ARG188	2	HB
THR190	4	HB

HB, significa puentes de hidrógeno

Como se observa en la Tabla 5, los valores más altos de energías de afinidad obtenidos en los cálculos de docking, corresponden a aquellas moléculas que interaccionan directamente con los aminoácidos reactivos en el sitio, ya sea como un puente de hidrógeno con CYS145 o una interacción con el anillo de HIS41. En la tabla 5 se muestran las energías de afinidad de los cálculos de *docking* para las 5 estructuras que presentaron la mayor energía de afinidad en su acoplamiento a la M^{pro}, así como su comparación con inhibidores previamente propuestos en la literatura.^{17,92} Las energías de afinidad obtenidas en los estudios de *docking* nos muestra que la prunasina presenta el valor máximo debido a su interacción con HIS41, sin embargo, aunque la prunasina comparte la misma energía de afinidad que el inhibidor N3 y supera energéticamente al inhibidor TG3, se observa que el inhibidor UED tiene un mayor valor energético. Esto puede deberse a que todos estos inhibidores reportados previamente presentan un mayor tamaño respecto a la prunasina, lo cual se ve reflejado en mayores interacciones en el sitio y por consiguiente en un posible aumento en la energía de afinidad.

Tabla 5. Conteo de frecuencias para las interacciones por aminoácido de las estructuras acopladas a la proteína M^{pro}

Compuesto	Energía de afinidad (Kcal/mol)
prunasina (FQNP193) ⁹⁰	-8.46
piceína (FQNP199) ⁹⁰	-6.93
metilarbutin (FQNP200) ⁹⁰	-6.68
massarigenina (FQNP37) ⁹³	-4.92
quinolonimida (FQNP93) ⁹⁴	-4.06
N3 (PDB ID: 6LU7) ¹⁷	-8.49
TG3 (PDB ID: 7JT7) ⁹²	-7.93
UED (PDB ID: 7JSU) ⁹²	-9.64

5.6 Selección de estructuras y estudios de dinámica molecular

La aproximación de los cálculos de docking realizados sobre la M^{pro}, nos permiten observar una posible configuración de acomodo de los ligantes en el sitio, sin embargo, aunque la posición final es aproximada, la contribución del solvente en el acomodo del ligante en el sitio es fundamental para conocer la posible estabilidad entre el complejo formado. Con el propósito de analizar el acomodo y la posible estabilidad del complejo, la metodología computacional de la dinámica molecular nos permite estudiar el acomodo de las moléculas que conforman al sistema (complejo ligante-proteína y solvente) a través de evaluar el movimiento de estas en el sistema. El detalle de esta metodología consiste en describir en términos de energía potencial, las interacciones de las especies y las coordenadas internas de las moléculas (enlaces, ángulos, ángulos diedros) en un documento conocido como campo de fuerza o *forcefield*.⁹⁵⁻⁹⁷. Utilizando el software Desmond,⁶⁰ se realizó la asignación de los parámetros de interacción y coordenadas internas a la M^{pro} así como a las estructuras que tuvieron las energías de afinidad más altas e interacciones con aminoácidos como HIS41 o CYS145 (prunasina, piceína y metilarbutin) empleando el *forcefield* OPLS4.^{61,62} El campo de fuerza de potenciales optimizados para simulaciones de líquidos o OPLS por sus siglas en inglés, tiene gran relevancia en los estudios

enfocados al desarrollo de fármacos y moléculas de interés biológico. Además, en su versión más reciente (OPLS4), se incluyen parámetros de interacción, distancias de enlace y ángulos, calculados por métodos experimentales y cálculos cuánticos actualizados. Permitiendo evaluar en una mejor forma detalles en la conformación de complejos ligante-proteína y prediciendo algunos valores como energías de enlace con mayor precisión.^{98,99} Las dinámicas moleculares de los complejos seleccionados se llevaron a cabo empleando un ensamble de tipo NVT a una temperatura de 300 K, por un tiempo de 100 ns recopilando el estado de sistema cada 2 ns. De las trayectorias recopiladas, se evaluó la estabilidad de los ligantes y el grado de relajación de la proteína a partir de calcular la desviación cuadrática media (RMSD, por sus siglas en inglés) del ligante y proteína. Como se observa en la Fig. 12a, los resultados de la dinámica molecular del complejo prunasina-M^{Pro} en el tiempo inicial (0.0 ns), la estructura se encuentra acoplada en el sitio y mantiene las principales interacciones reportadas en los cálculos de docking, sin embargo, conforme pasa el tiempo en la dinámica (100 ns) se observa un desplazamiento del ligante hacia afuera del sitio, llegando incluso a salir completamente y no tener ningún tipo de interacción (Fig. 12b). Es importante considerar que, en la unión de un ligante a un sitio receptor, el acomodo del solvente en el sitio juega un papel importante en la disposición final del ligante. El cálculo del RMSD mostrado en la Fig. 13, representa numéricamente y en términos de distancia, la relajación de la proteína respecto a su forma cristalizada, así como la estabilidad del ligante en el sitio acoplado en los cálculos de *docking*.

A diferencia del complejo de prunasina, el acoplamiento a 0.0 ns del complejo piceína-M^{pro} (Fig. 14a) no presenta alguna interacción con los aminoácidos HIS41 o CYS145, sin embargo, se observan importantes interacciones de puentes de hidrógeno con: GLU166, THR190 y ASP187. Conforme avanza la dinámica molecular del complejo (100 ns), observamos una configuración final en el que solo sobresalen las interacciones con los aminoácidos GLU166 y GLN189, así como interacciones con las moléculas del solvente.

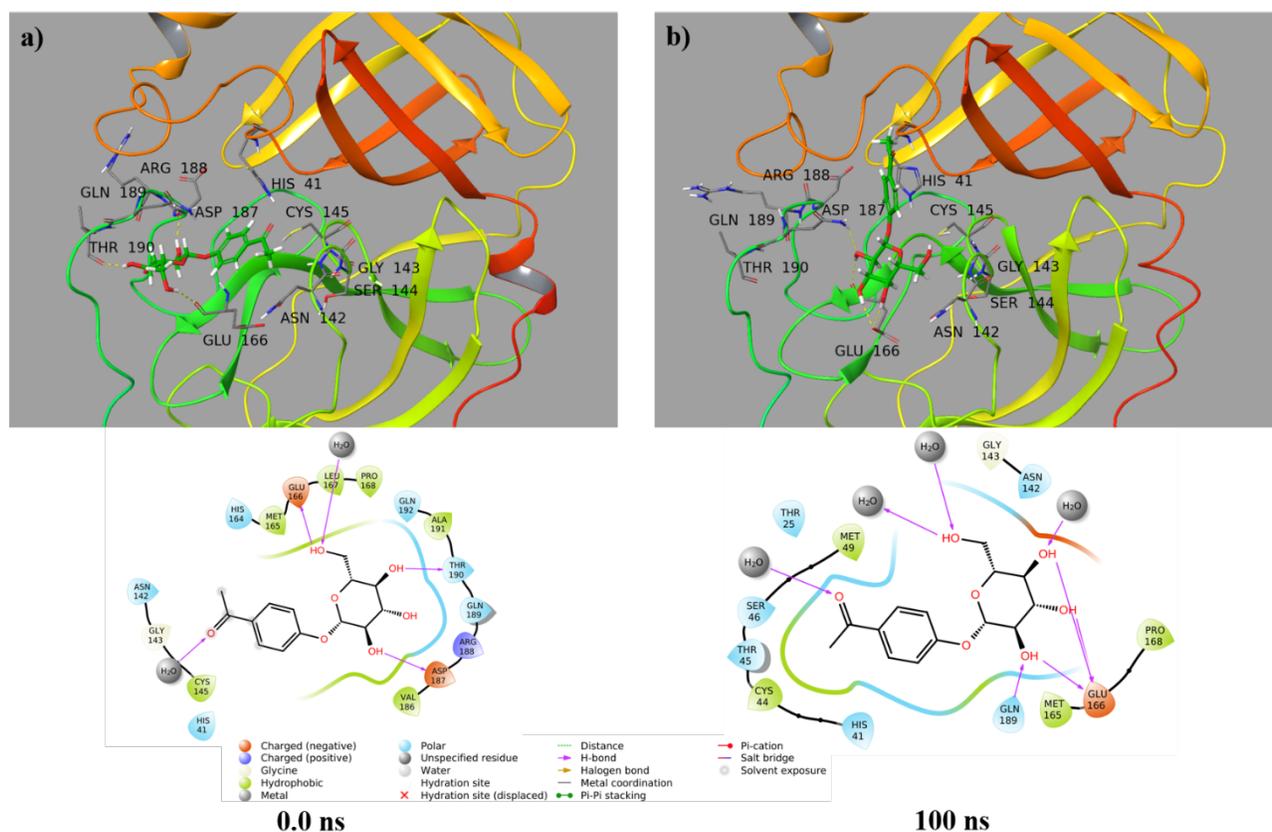


Figura 14. (a) Configuración y diagrama 2D de piceína a un tiempo de 0.0 ns. (b) Configuración y diagrama 2D de piceína a un tiempo de 100 ns.

A diferencia del complejo anterior, se observa que la piceína tiene una mayor presencia en el sitio de la M^{pro} (Fig. 14). Esta observación puede confirmarse en el cálculo de RMSD para el ligante en el sitio (Fig. 15), en donde observamos que a partir del paso No. 60, la fluctuación en el valor es mínima indicando una estabilidad en el sitio.

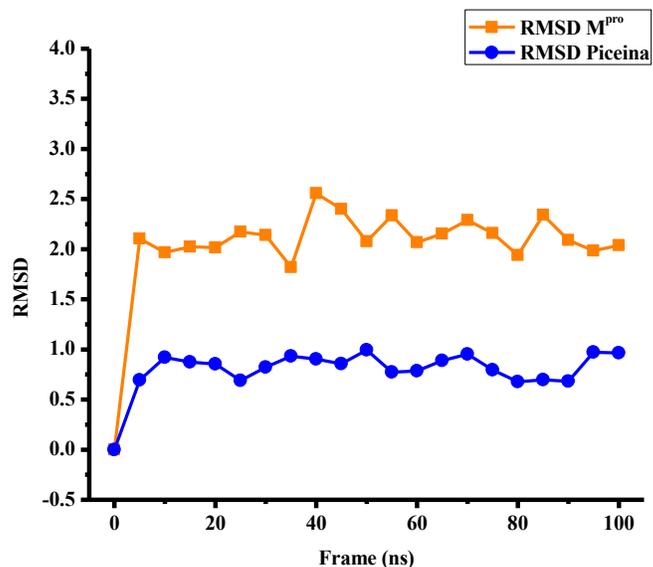


Figura 15. Cálculo de RMSD para el ligante piceína en el complejo con la M^{pro} (línea azul). Valores de RMSD de la relajación de la proteína M^{pro} en disolución acuosa (línea naranja).

Finalmente, en la Fig. 16, el complejo metilarbutin-M^{pro} presenta durante su conformación inicial (0.0 ns) la interacción de tipo puentes de hidrógeno con los aminoácidos: GLN189, GLU166, ASN142 y SER144. Además, se observa una interacción de tipo $\pi - \pi$ con el aminoácido HIS41. Conforme el tiempo en la dinámica continúa (100 ns), la configuración del ligante se relaja y las interacciones con el sitio se mantienen en su mayoría incluyendo la interacción con el aminoácido HIS41. El acomodo final de esta molécula y la presencia de interacciones con aminoácidos previamente reportados como importantes en el sitio (HIS41 o GLU166) sugieren una mayor estabilidad del ligante en el sitio.

Esto puede confirmarse en la magnitud del valor de RMSD calculado para el ligante en el sitio (Fig. 17), mostrando una separación de casi 0.5 Å

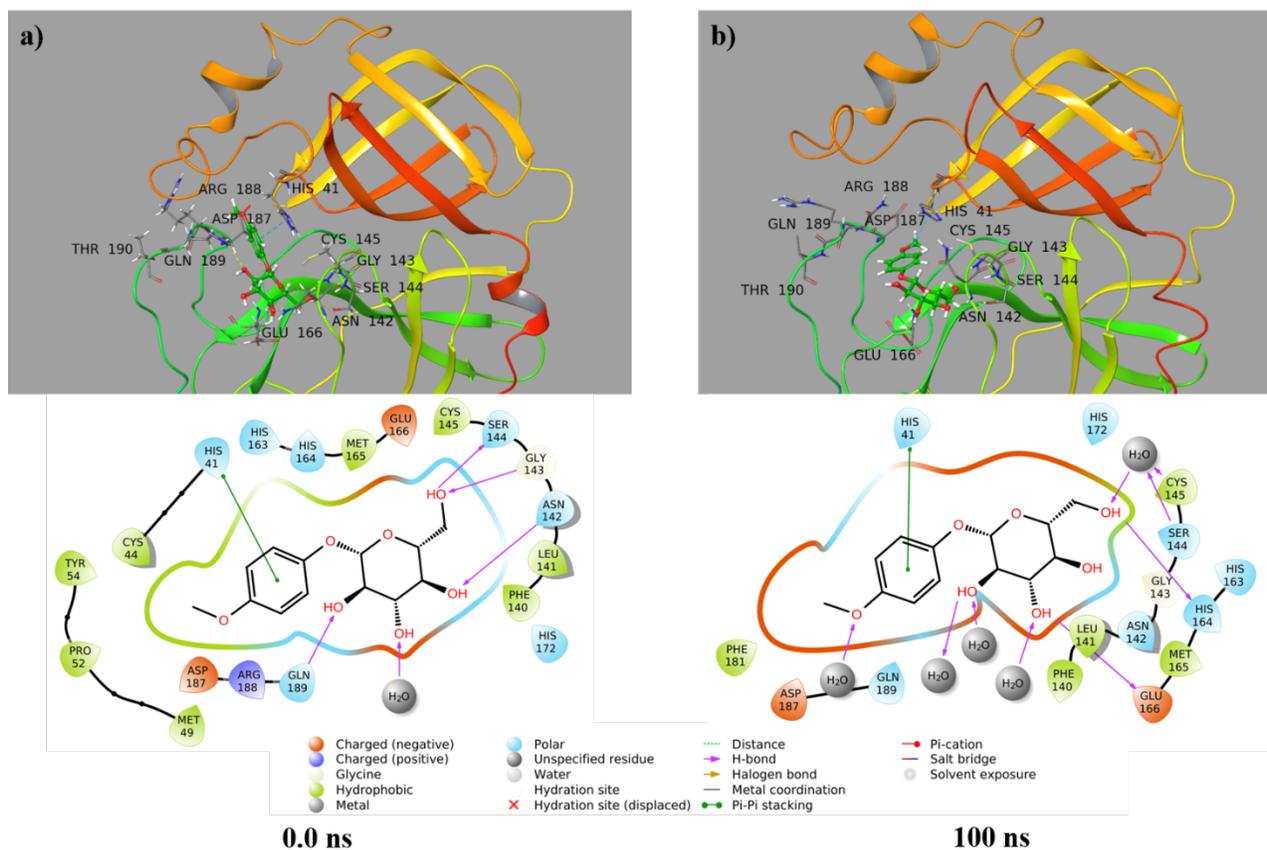


Figura 16. (a) Configuración y diagrama 2D de metilarbutin a un tiempo de 0.0 ns. (b) Configuración y diagrama 2D de metilarbutin a un tiempo de 100 ns.

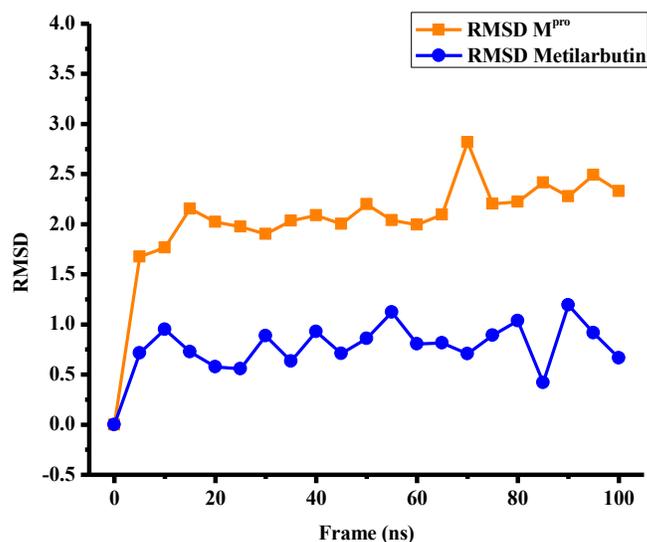


Figura 17. Cálculo de RMSD para el ligante metilarbutin en el complejo con la M^{pro} (línea azul). Valores de RMSD de la relajación de la proteína M^{pro} en disolución acuosa (línea naranja).

La estabilidad de los ligantes en el sitio, es una propiedad útil para evaluar la disposición de los ligantes en el sitio y como esto influye en la selección de las mejores estructuras químicas. Dentro de la investigación de moléculas con posibles propiedades terapéuticas, el estudio para obtener la energía de unión correspondiente al ligante L y el receptor R denotada como $\Delta G_{\text{unión}}$ se ha calculado a través de técnicas experimentales como la calorimetría isotérmica entre otras. Sin embargo, debido al costo que se requiere para evaluar la energía de unión por cada molécula que se desea acoplar a un receptor, los valores obtenidos de los cálculos de *docking* han tenido relevancia. Estos valores de energía se encuentran limitados por la metodología de cálculo que tiene el software para determinar el valor, por lo que se requiere una metodología computacional más precisa respecto al cálculo de la energía de unión. El modelo basado en mecánica molecular y la aproximación del modelo generalizado de Born (MM/GBSA)¹⁰⁰ ha demostrado tener una mayor precisión respecto a la energía de afinidad obtenida en los cálculos de *docking*.^{101,102} De manera

general, la metodología MM/GBSA nos dice que la energía de unión ΔG_{union} se encuentra definida por las energías libres de los reactantes menos los productos como se muestra:

$$\Delta G_{\text{union}} = \langle G_{\text{RL}} \rangle - \langle G_{\text{R}} \rangle - \langle G_{\text{L}} \rangle \quad (6)$$

Este método desarrollado por Kollman,¹⁰¹ ha tenido una gran popularidad en los últimos años debido a la utilidad en el diseño de proteínas,¹⁰² estudio de interacciones entre proteínas,¹⁰³ estabilidad de conformeros y cálculo de energías de afinidad.¹⁰³⁻¹⁰⁷ Empleando el modelo MM/GBSA incorporado en el software Maestro, se calcularon los valores de ΔG_{union} para los complejos calculados por dinámica molecular. En la tabla 5 podemos observar que el complejo de Prunasina presenta la menor energía de unión, confirmando el resultado obtenido por dinámica molecular en donde el ligante fue expulsado del sitio de unión propuesto en los cálculos de *docking*. Por otro lado, La diferencia energética entre el complejo con Piceína y el Metilarbutin puede deberse en su mayoría a la cantidad de interacciones con el solvente que tiene el Metilarbutin respecto a la piceína. Además, dentro de las interacciones con mayor importancia podemos observar que el metilarbutin conserva una importante interacción de tipo $\pi - \pi$ con el aminoácido HIS41, mostrando un mayor anclaje en el sitio y una mayor energía de unión.

Tabla 5. Valores de energías de unión ΔG_{union} calculadas por el método MM/GBSA

Complejos	ΔG_{union} Kcal/mol
Prunasina-M ^{pro}	-13.961
Piceína-M ^{pro}	-24.280
Metilarbutin-M ^{pro}	-46.777

Estos valores energéticos apoyados por los estudios de dinámica molecular, sugieren que el metilarbutin podría ser una molécula con posible actividad

terapéutica o inhibitoria en la proteína M^{pro}. Sin embargo, se requieren estudios más profundos (ensayos clínicos, estudios de toxicidad y ensayos de inhibición) para determinar si esta molécula puede ser candidato a fármaco.

6 Conclusiones

Como respuesta al estudio del virus SARS-CoV-2 que produce la enfermedad del Covid-19, se realizó el estudio computacional de una base de datos compuesta de productos naturales mexicanos (423 compuestos) llamada BIOFACQUIM. A partir del cálculo de descriptores químicos, se determinó que la mayor parte de las moléculas se encuentran en el intervalo de 340 a 412 g/mol, así como en un rango de LogP_{ow} entre -0.3 a 5. Estos valores muestran que las moléculas pertenecientes a la base de datos tienen preferencia por sitios hidrofóbicos o de baja polaridad. Con el propósito de analizar la naturaleza química de la proteína encargada de la replicación en el virus llamada M^{pro} los resultados obtenidos por el *software* SiteMap y calculados por métodos basados en química cuántica (COSMO-RS), muestran la preferencia del sitio por especies polares (-1.4 a 0), así como la preferencia de moléculas con un tamaño igual o menor a 254.63 \AA^3 . Con estos parámetros de selección, se encontraron 8 moléculas, las cuales se acoplaron al sitio de la M^{pro} , utilizando el software Glide en su modalidad de *docking* flexible. Encontrando una mayor energía de afinidad (-8.46 Kcal/mol) para el ligante prunasina, la cual se comparó con algunos inhibidores previamente reportados. Las moléculas que presentaron energías de afinidad superiores o iguales a los inhibidores reportados, así como interacciones con aminoácidos previamente reportados como puntos farmacóforos (prunasina, piceína y metilarbutin) fueron analizadas a través de estudios de dinámica molecular. Los cálculos de RMSD obtenidos de las corridas de dinámica molecular fueron analizados, encontrando una mayor estabilidad para los complejos con piceína y el metilarbutin. Con estos resultados, y los cálculos de energías de unión obtenidas por el método MM/GBSA mostraron que el compuesto metilarbutin tiene una mayor energía de unión, así como una cantidad importante de interacciones con aminoácidos

como: GLU166, HIS141 y GLN189. Esta información nos permite proponer a la metilarbutin para estudios de mayor relevancia, como la reactividad a través de funciones de Fukui, estudios que relacionan química cuántica y mecánica cuántica (QM/MM, por sus siglas en inglés) o incluso modificación de la estructura para mejorar con ello la cantidad de interacciones en el sitio.

7 Referencias

- (1) Neiderud, C.-J. How Urbanization Affects the Epidemiology of Emerging Infectious Diseases. *Infection Ecology & Epidemiology* **2015**, 5 (1), 27060. <https://doi.org/10.3402/iee.v5.27060>.
- (2) Woo, P. C. Y.; Lau, S. K. P.; Huang, Y.; Yuen, K.-Y. Coronavirus Diversity, Phylogeny and Interspecies Jumping. *Exp Biol Med (Maywood)* **2009**, 234 (10), 1117–1127. <https://doi.org/10.3181/0903-MR-94>.
- (3) Hamming, I.; Timens, W.; Bulthuis, M.; Lely, A.; Navis, G.; van Goor, H. Tissue Distribution of ACE2 Protein, the Functional Receptor for SARS Coronavirus. A First Step in Understanding SARS Pathogenesis. *J. Pathol.* **2004**, 203 (2), 631–637. <https://doi.org/10.1002/path.1570>.
- (4) Xu, H.; Zhong, L.; Deng, J.; Peng, J.; Dan, H.; Zeng, X.; Li, T.; Chen, Q. High Expression of ACE2 Receptor of 2019-NCov on the Epithelial Cells of Oral Mucosa. *Int J Oral Sci* **2020**, 12 (1), 8. <https://doi.org/10.1038/s41368-020-0074-x>.
- (5) Kasmi, Y.; Khataby, K.; Souiri, A.; Ennaji, M. M. Coronaviridae: 100,000 Years of Emergence and Reemergence. In *Emerging and Reemerging Viral Pathogens*; Elsevier, 2020; pp 127–149. <https://doi.org/10.1016/B978-0-12-819400-3.00007-7>.
- (6) Enjuanes, L. *Coronavirus Replication and Reverse Genetics*; Current topics in microbiology and immunology; Springer: Berlin New York, 2005.
- (7) Denison, M. R.; Spaan, W. J. M.; van der Meer, Y.; Gibson, C. A.; Sims, A. C.; Prentice, E.; Lu, X. T. The Putative Helicase of the Coronavirus Mouse Hepatitis Virus Is Processed from the Replicase Gene Polyprotein and Localizes in Complexes That Are Active in Viral RNA Synthesis. *J Virol* **1999**, 73 (8), 6862–6871. <https://doi.org/10.1128/JVI.73.8.6862-6871.1999>.
- (8) Cheng, V. C. C.; Lau, S. K. P.; Woo, P. C. Y.; Yuen, K. Y. Severe Acute Respiratory Syndrome Coronavirus as an Agent of Emerging and Reemerging Infection. *Clin Microbiol Rev* **2007**, 20 (4), 660–694. <https://doi.org/10.1128/CMR.00023-07>.
- (9) Cunningham, C. H.; Stuart, H. O. Cultivation of the Virus of Infectious Bronchitis of Chickens in Embryonated Chicken Eggs. *Am J Vet Res* **1947**, 8 (27), 209–212.
- (10) Cheever, F. S.; Daniels, J. B.; Pappenheimer, A. M.; Bailey, O. T. A MURINE VIRUS (JHM) CAUSING DISSEMINATED ENCEPHALOMYELITIS WITH EXTENSIVE DESTRUCTION OF MYELIN. *Journal of Experimental Medicine* **1949**, 90 (3), 181–194. <https://doi.org/10.1084/jem.90.3.181>.
- (11) Doyle, L. P.; Hutchings, L. M. A Transmissible Gastroenteritis in Pigs. *J Am Vet Med Assoc* **1946**, 108, 257–259.
- (12) Xu, X.; Chen, P.; Wang, J.; Feng, J.; Zhou, H.; Li, X.; Zhong, W.; Hao, P. Evolution of the Novel Coronavirus from the Ongoing Wuhan Outbreak and Modeling of Its Spike Protein for Risk of Human Transmission. *Sci. China Life Sci.* **2020**, 63 (3), 457–460. <https://doi.org/10.1007/s11427-020-1637-5>.
- (13) Gordon, D. E.; Jang, G. M.; Bouhaddou, M.; Xu, J.; Obernier, K.; White, K. M.; O’Meara, M. J.; Rezelj, V. V.; Guo, J. Z.; Swaney, D. L.; Tummino, T. A.; Hüttenhain, R.; Kaake, R. M.; Richards, A. L.; Tutuncuoglu, B.; Foussard, H.; Batra, J.; Haas, K.; Modak, M.; Kim, M.; Haas, P.; Polacco, B. J.; Braberg, H.; Fabius, J. M.; Eckhardt, M.; Soucheray, M.; Bennett, M. J.; Cakir, M.; McGregor, M. J.; Li, Q.; Meyer, B.; Roesch, F.; Vallet, T.; Mac Kain, A.; Miorin, L.; Moreno, E.; Naing, Z. Z. C.; Zhou, Y.; Peng, S.; Shi, Y.; Zhang, Z.; Shen, W.; Kirby, I. T.; Melnyk, J. E.; Chorba, J. S.; Lou, K.; Dai, S. A.; Barrio-

Hernandez, I.; Memon, D.; Hernandez-Armenta, C.; Lyu, J.; Mathy, C. J. P.; Perica, T.; Pilla, K. B.; Ganesan, S. J.; Saltzberg, D. J.; Rakesh, R.; Liu, X.; Rosenthal, S. B.; Calviello, L.; Venkataramanan, S.; Liboy-Lugo, J.; Lin, Y.; Huang, X.-P.; Liu, Y.; Wankowicz, S. A.; Bohn, M.; Safari, M.; Ugur, F. S.; Koh, C.; Savar, N. S.; Tran, Q. D.; Shengjuler, D.; Fletcher, S. J.; O'Neal, M. C.; Cai, Y.; Chang, J. C. J.; Broadhurst, D. J.; Klippsten, S.; Sharp, P. P.; Wenzell, N. A.; Kuzuoglu-Ozturk, D.; Wang, H.-Y.; Trenker, R.; Young, J. M.; Cavero, D. A.; Hiatt, J.; Roth, T. L.; Rathore, U.; Subramanian, A.; Noack, J.; Hubert, M.; Stroud, R. M.; Frankel, A. D.; Rosenberg, O. S.; Verba, K. A.; Agard, D. A.; Ott, M.; Emerman, M.; Jura, N.; von Zastrow, M.; Verdin, E.; Ashworth, A.; Schwartz, O.; d'Enfert, C.; Mukherjee, S.; Jacobson, M.; Malik, H. S.; Fujimori, D. G.; Ideker, T.; Craik, C. S.; Floor, S. N.; Fraser, J. S.; Gross, J. D.; Sali, A.; Roth, B. L.; Ruggero, D.; Taunton, J.; Kortemme, T.; Beltrao, P.; Vignuzzi, M.; García-Sastre, A.; Shokat, K. M.; Shoichet, B. K.; Krogan, N. J. A SARS-CoV-2 Protein Interaction Map Reveals Targets for Drug Repurposing. *Nature* **2020**, *583* (7816), 459–468.

<https://doi.org/10.1038/s41586-020-2286-9>.

(14) Eastman, R. T.; Roth, J. S.; Brimacombe, K. R.; Simeonov, A.; Shen, M.; Patnaik, S.; Hall, M. D. Remdesivir: A Review of Its Discovery and Development Leading to Emergency Use Authorization for Treatment of COVID-19. *ACS Cent. Sci.* **2020**, *6* (5), 672–683. <https://doi.org/10.1021/acscentsci.0c00489>.

(15) WHO Solidarity trial consortium; Pan, H.; Peto, R.; Karim, Q. A.; Alejandria, M.; Henao-Restrepo, A. M.; García, C. H.; Kieny, M.-P.; Malekzadeh, R.; Murthy, S.; Preziosi, M.-P.; Reddy, S.; Periago, M. R.; Sathiyamoorthy, V.; Röttingen, J.-A.; Swaminathan, S.; as the members of the Writing Committee, assume responsibility for the content and integrity of this article. *Repurposed Antiviral Drugs for COVID-19 –Interim WHO SOLIDARITY Trial Results*; preprint; Infectious Diseases (except HIV/AIDS), 2020. <https://doi.org/10.1101/2020.10.15.20209817>.

(16) Mengist, H. M.; Dilnessa, T.; Jin, T. Structural Basis of Potential Inhibitors Targeting SARS-CoV-2 Main Protease. *Front. Chem.* **2021**, *9*, 622898. <https://doi.org/10.3389/fchem.2021.622898>.

(17) Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; Duan, Y.; Yu, J.; Wang, L.; Yang, K.; Liu, F.; Jiang, R.; Yang, X.; You, T.; Liu, X.; Yang, X.; Bai, F.; Liu, H.; Liu, X.; Guddat, L. W.; Xu, W.; Xiao, G.; Qin, C.; Shi, Z.; Jiang, H.; Rao, Z.; Yang, H. Structure of Mpro from SARS-CoV-2 and Discovery of Its Inhibitors. *Nature* **2020**, *582* (7811), 289–293. <https://doi.org/10.1038/s41586-020-2223-y>.

(18) Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; Duan, Y.; Yu, J.; Wang, L.; Yang, K.; Liu, F.; Jiang, R.; Yang, X.; You, T.; Liu, X.; Yang, X.; Bai, F.; Liu, H.; Liu, X.; Guddat, L. W.; Xu, W.; Xiao, G.; Qin, C.; Shi, Z.; Jiang, H.; Rao, Z.; Yang, H. Structure of Mpro from SARS-CoV-2 and Discovery of Its Inhibitors. *Nature* **2020**, *582* (7811), 289–293. <https://doi.org/10.1038/s41586-020-2223-y>.

(19) Paul, A. S.; Islam, R.; Parves, M. R.; Mamun, A. A.; Shahriar, I.; Hossain, M. I.; Hossain, M. N.; Ali, M. A.; Halim, M. A. Cysteine Focused Covalent Inhibitors against the Main Protease of SARS-CoV-2. *Journal of Biomolecular Structure and Dynamics* **2020**, 1–20. <https://doi.org/10.1080/07391102.2020.1831610>.

(20) Jin, Z.; Zhao, Y.; Sun, Y.; Zhang, B.; Wang, H.; Wu, Y.; Zhu, Y.; Zhu, C.; Hu, T.; Du, X.; Duan, Y.; Yu, J.; Yang, X.; Yang, X.; Yang, K.; Liu, X.; Guddat, L. W.; Xiao, G.; Zhang, L.; Yang, H.; Rao, Z. Structural Basis for the Inhibition of SARS-CoV-2 Main Protease by Antineoplastic Drug Carmofur. *Nat Struct Mol Biol* **2020**, *27* (6), 529–532. <https://doi.org/10.1038/s41594-020-0440-6>.

- (21) Main Protease Structure and XChem Fragment Screen – Diamond Light Source, <https://www.diamond.ac.uk/Covid-19/for-Scientists/Main-Protease-Structure-and-XChem.html>, Accessed 26 January 2021.
- (22) Muratov, E. N.; Amaro, R.; Andrade, C. H.; Brown, N.; Ekins, S.; Fourches, D.; Isayev, O.; Kozakov, D.; Medina-Franco, J. L.; Merz, K. M.; Oprea, T. I.; Poroikov, V.; Schneider, G.; Todd, M. H.; Varnek, A.; Winkler, D. A.; Zakharov, A. V.; Cherkasov, A.; Tropsha, A. A Critical Overview of Computational Approaches Employed for COVID-19 Drug Discovery. *Chem. Soc. Rev.* **2021**, *50* (16), 9121–9151. <https://doi.org/10.1039/D0CS01065K>.
- (23) Fourches, D.; Muratov, E.; Tropsha, A. Curation of Chemogenomics Data. *Nat Chem Biol* **2015**, *11* (8), 535–535. <https://doi.org/10.1038/nchembio.1881>.
- (24) Hersey, A.; Chambers, J.; Bellis, L.; Patrícia Bento, A.; Gaulton, A.; Overington, J. P. Chemical Databases: Curation or Integration by User-Defined Equivalence? *Drug Discovery Today: Technologies* **2015**, *14*, 17–24. <https://doi.org/10.1016/j.ddtec.2015.01.005>.
- (25) Kumar, A.; Arya, P. K. Database Resources for Drug Discovery. In *Computer-Aided Drug Design*; Singh, D. B., Ed.; Springer Singapore: Singapore, 2020; pp 89–114. https://doi.org/10.1007/978-981-15-6815-2_5.
- (26) Newman, D. J. From Natural Products to Drugs. *Physical Sciences Reviews* **2019**, *4* (4). <https://doi.org/10.1515/psr-2018-0111>.
- (27) Newman, D. J.; Cragg, G. M. Natural Products as Sources of New Drugs from 1981 to 2014. *J. Nat. Prod.* **2016**, *79* (3), 629–661. <https://doi.org/10.1021/acs.jnatprod.5b01055>.
- (28) Gu, J.; Gui, Y.; Chen, L.; Yuan, G.; Lu, H.-Z.; Xu, X. Use of Natural Products as Chemical Library for Drug Discovery and Network Pharmacology. *PLoS ONE* **2013**, *8* (4), e62839. <https://doi.org/10.1371/journal.pone.0062839>.
- (29) Chen, C. Y.-C. TCM Database@Taiwan: The World's Largest Traditional Chinese Medicine Database for Drug Screening In Silico. *PLoS ONE* **2011**, *6* (1), e15939. <https://doi.org/10.1371/journal.pone.0015939>.
- (30) Pilon, A. C.; Valli, M.; Dametto, A. C.; Pinto, M. E. F.; Freire, R. T.; Castro-Gamboa, I.; Andricopulo, A. D.; Bolzani, V. S. NuBBEDB: An Updated Database to Uncover Chemical and Biological Information from Brazilian Biodiversity. *Sci Rep* **2017**, *7* (1), 7215. <https://doi.org/10.1038/s41598-017-07451-x>.
- (31) Nguyen-Vo, T.-H.; Le, T.; Pham, D.; Nguyen, T.; Le, P.; Nguyen, A.; Nguyen, T.; Nguyen, T.-N.; Nguyen, V.; Do, H.; Trinh, K.; Duong, H. T.; Le, L. VIETHERB: A Database for Vietnamese Herbal Species. *J. Chem. Inf. Model.* **2019**, *59* (1), 1–9. <https://doi.org/10.1021/acs.jcim.8b00399>.
- (32) Pilon-Jiménez, B.; Saldívar-González, F.; Díaz-Eufracio, B.; Medina-Franco, J. BIOFACQUIM: A Mexican Compound Database of Natural Products. *Biomolecules* **2019**, *9* (1), 31. <https://doi.org/10.3390/biom9010031>.
- (33) Harris, D. C. *Análisis Químico Cuantitativo*; Revert: Barcelona, 2016.
- (34) Butina, D.; Segall, M. D.; Frankcombe, K. Predicting ADME Properties in Silico: Methods and Models. *Drug Discovery Today* **2002**, *7* (11), S83–S88. [https://doi.org/10.1016/S1359-6446\(02\)02288-2](https://doi.org/10.1016/S1359-6446(02)02288-2).
- (35) Kubinyi, H. Nonlinear Dependence of Biological Activity on Hydrophobic Character: The Bilinear Model. *Farmaco Sci* **1979**, *34* (3), 248–276.
- (36) Tshepelevitsh, S.; Kadam, S. A.; Darnell, A.; Bobacka, J.; Rützel, A.; Haljasorg, T.; Leito, I. LogP Determination for Highly Lipophilic Hydrogen-Bonding Anion Receptor Molecules. *Analytica Chimica Acta* **2020**, *1132*, 123–133.

<https://doi.org/10.1016/j.aca.2020.07.024>.

- (37) Schönsee, C. D.; Bucheli, T. D. Experimental Determination of Octanol–Water Partition Coefficients of Selected Natural Toxins. *J. Chem. Eng. Data* **2020**, *65* (4), 1946–1953. <https://doi.org/10.1021/acs.jced.9b01129>.
- (38) Ioakimidis, L.; Thoukydidis, L.; Mirza, A.; Naeem, S.; Reynisson, J. Benchmarking the Reliability of QikProp. Correlation between Experimental and Predicted Values. *QSAR Comb. Sci.* **2008**, *27* (4), 445–456. <https://doi.org/10.1002/qsar.200730051>.
- (39) Ho, J.; Ertem, M. Z. Calculating Free Energy Changes in Continuum Solvation Models. *J. Phys. Chem. B* **2016**, *120* (7), 1319–1329. <https://doi.org/10.1021/acs.jpcc.6b00164>.
- (40) Zhang, J.; Tuguldur, B.; van der Spoel, D. Correction to Force Field Benchmark of Organic Liquids. 2. Gibbs Energy of Solvation. *J. Chem. Inf. Model.* **2016**, *56* (4), 819–820. <https://doi.org/10.1021/acs.jcim.6b00081>.
- (41) Kundi, V.; Ho, J. Predicting Octanol–Water Partition Coefficients: Are Quantum Mechanical Implicit Solvent Models Better than Empirical Fragment-Based Methods? *J. Phys. Chem. B* **2019**, *123* (31), 6810–6822. <https://doi.org/10.1021/acs.jpcc.9b04061>.
- (42) Klamt, A.; Eckert, F. COSMO-RS: A Novel and Efficient Method for the a Priori Prediction of Thermophysical Data of Liquids. *Fluid Phase Equilibria* **2000**, *172* (1), 43–72. [https://doi.org/10.1016/S0378-3812\(00\)00357-5](https://doi.org/10.1016/S0378-3812(00)00357-5).
- (43) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113* (18), 6378–6396. <https://doi.org/10.1021/jp810292n>.
- (44) Nozaki, Y.; Tanford, C. The Solubility of Amino Acids and Two Glycine Peptides in Aqueous Ethanol and Dioxane Solutions. Establishment of a Hydrophobicity Scale. *J. Biol. Chem.* **1971**, *246* (7), 2211–2217.
- (45) Heijne, G. On the Hydrophobic Nature of Signal Sequences. *Eur. J. Biochem.* **1981**, *116* (2), 419–422. <https://doi.org/10.1111/j.1432-1033.1981.tb05351.x>.
- (46) Mannhold, R.; Rekker, R. F. The Hydrophobic Fragmental Constant Approach Forcalculating Log P in Octanol/Water and Aliphatic Hydrocarbon/Water Systems. *Perspectives in Drug Discovery and Design* **2000**, *18* (1), 1–18. <https://doi.org/10.1023/A:1008782809845>.
- (47) Halgren, T. New Method for Fast and Accurate Binding-Site Identification and Analysis. *Chem Biol Drug Design* **2007**, *69* (2), 146–148. <https://doi.org/10.1111/j.1747-0285.2007.00483.x>.
- (48) Yu, J.; Zhou, Y.; Tanaka, I.; Yao, M. Roll: A New Algorithm for the Detection of Protein Pockets and Cavities with a Rolling Probe Sphere. *Bioinformatics* **2010**, *26* (1), 46–52. <https://doi.org/10.1093/bioinformatics/btp599>.
- (49) Balasubramani, S. G.; Chen, G. P.; Coriani, S.; Diedenhofen, M.; Frank, M. S.; Franzke, Y. J.; Furche, F.; Grotjahn, R.; Harding, M. E.; Hättig, C.; Hellweg, A.; Helmich-Paris, B.; Holzer, C.; Huniar, U.; Kaupp, M.; Marefat Khah, A.; Karbalaei Khani, S.; Müller, T.; Mack, F.; Nguyen, B. D.; Parker, S. M.; Perlt, E.; Rappoport, D.; Reiter, K.; Roy, S.; Rückert, M.; Schmitz, G.; Sierka, M.; Tapavicza, E.; Tew, D. P.; van Wüllen, C.; Voora, V. K.; Weigend, F.; Wodyński, A.; Yu, J. M. TURBOMOLE: Modular Program Suite for *Ab Initio* Quantum-Chemical and Condensed-Matter Simulations. *J. Chem. Phys.* **2020**, *152* (18), 184107. <https://doi.org/10.1063/5.0004635>.
- (50) Klamt, A. Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *J. Phys. Chem.* **1995**, *99* (7), 2224–

2235. <https://doi.org/10.1021/j100007a062>.
- (51) Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R. Auxiliary Basis Sets for Main Row Atoms and Transition Metals and Their Use to Approximate Coulomb Potentials. *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)* **1997**, *97* (1–4), 119–124. <https://doi.org/10.1007/s002140050244>.
- (52) Sander, T.; Freyss, J.; von Korff, M.; Rufener, C. DataWarrior: An Open-Source Program For Chemistry Aware Data Visualization And Analysis. *J. Chem. Inf. Model.* **2015**, *55* (2), 460–473. <https://doi.org/10.1021/ci500588j>.
- (53) *Schrödinger Release 2021-3: LigPrep, Schrödinger, LLC, New York, NY, 2021.*
- (54) Shelley, J. C.; Cholleti, A.; Frye, L. L.; Greenwood, J. R.; Timlin, M. R.; Uchimaya, M. Epik: A Software Program for PK a Prediction and Protonation State Generation for Drug-like Molecules. *J Comput Aided Mol Des* **2007**, *21* (12), 681–691. <https://doi.org/10.1007/s10822-007-9133-z>.
- (55) Greenwood, J. R.; Calkins, D.; Sullivan, A. P.; Shelley, J. C. Towards the Comprehensive, Rapid, and Accurate Prediction of the Favorable Tautomeric States of Drug-like Molecules in Aqueous Solution. *J Comput Aided Mol Des* **2010**, *24* (6–7), 591–604. <https://doi.org/10.1007/s10822-010-9349-1>.
- (56) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47* (7), 1739–1749. <https://doi.org/10.1021/jm0306430>.
- (57) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47* (7), 1750–1759. <https://doi.org/10.1021/jm030644s>.
- (58) Yoshino, R.; Yasuo, N.; Sekijima, M. Identification of Key Interactions between SARS-CoV-2 Main Protease and Inhibitor Drug Candidates. *Sci Rep* **2020**, *10* (1), 12493. <https://doi.org/10.1038/s41598-020-69337-9>.
- (59) Martínez-Ortega, U.; Figueroa-Figueroa, D. I.; Hernández-Luis, F.; Aguayo-Ortiz, R. In Silico Characterization of Masitinib Interaction with SARS-CoV-2 Main Protease. *ChemMedChem* **2021**, *16* (15), 2339–2344. <https://doi.org/10.1002/cmdc.202100375>.
- (60) *Powerful beyond Imagination: SC 06, [2006 IEEE/ACM SC Conference] ; November 11 - 17, 2006, Tampa Convention Center, Tampa, Florida ; Proceedings; Institute of Electrical and Electronics Engineers, Association for Computing Machinery, IEEE Computer Society, IEEE Computer Society, Association for Computing Machinery, Eds.; IEEE: Piscataway, NJ, 2006.*
- (61) Shivakumar, D.; Williams, J.; Wu, Y.; Damm, W.; Shelley, J.; Sherman, W. Prediction of Absolute Solvation Free Energies Using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field. *J. Chem. Theory Comput.* **2010**, *6* (5), 1509–1519. <https://doi.org/10.1021/ct900587b>.
- (62) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12* (1), 281–296. <https://doi.org/10.1021/acs.jctc.5b00864>.
- (63) Goodford, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**,

- 28 (7), 849–857. <https://doi.org/10.1021/jm00145a002>.
- (64) Halgren, T. New Method for Fast and Accurate Binding-Site Identification and Analysis. *Chem Biol Drug Design* **2007**, *69* (2), 146–148. <https://doi.org/10.1111/j.1747-0285.2007.00483.x>.
- (65) Halgren, T. A. Identifying and Characterizing Binding Sites and Assessing Druggability. *J. Chem. Inf. Model.* **2009**, *49* (2), 377–389. <https://doi.org/10.1021/ci800324m>.
- (66) Mikol, V.; Papageorgiou, C.; Borer, X. The Role of Water Molecules in the Structure-Based Design of (5-Hydroxynorvaline)-2-Cyclosporin: Synthesis, Biological Activity, and Crystallographic Analysis with Cyclophilin A. *J. Med. Chem.* **1995**, *38* (17), 3361–3367. <https://doi.org/10.1021/jm00017a020>.
- (67) Biela, A.; Khayat, M.; Tan, H.; Kong, J.; Heine, A.; Hangauer, D.; Klebe, G. Impact of Ligand and Protein Desolvation on Ligand Binding to the S1 Pocket of Thrombin. *Journal of Molecular Biology* **2012**, *418* (5), 350–366. <https://doi.org/10.1016/j.jmb.2012.01.054>.
- (68) Li, J.; Zhou, X.; Zhang, Y.; Zhong, F.; Lin, C.; McCormick, P. J.; Jiang, F.; Luo, J.; Zhou, H.; Wang, Q.; Fu, Y.; Duan, J.; Zhang, J. Crystal Structure of SARS-CoV-2 Main Protease in Complex with the Natural Product Inhibitor Shikonin Illuminates a Unique Binding Mode. *Science Bulletin* **2021**, *66* (7), 661–663. <https://doi.org/10.1016/j.scib.2020.10.018>.
- (69) Forrestall, K. L.; Burley, D. E.; Cash, M. K.; Pottie, I. R.; Darvesh, S. 2-Pyridone Natural Products as Inhibitors of SARS-CoV-2 Main Protease. *Chemico-Biological Interactions* **2021**, 335, 109348. <https://doi.org/10.1016/j.cbi.2020.109348>.
- (70) Tam, N. M.; Nam, P. C.; Quang, D. T.; Tung, N. T.; Vu, V. V.; Ngo, S. T. Binding of Inhibitors to the Monomeric and Dimeric SARS-CoV-2 Mpro. *RSC Adv.* **2021**, *11* (5), 2926–2934. <https://doi.org/10.1039/D0RA09858B>.
- (71) Wang, R.; Fang, X.; Lu, Y.; Yang, C.-Y.; Wang, S. The PDBbind Database: Methodologies and Updates. *J. Med. Chem.* **2005**, *48* (12), 4111–4119. <https://doi.org/10.1021/jm048957q>.
- (72) Barone, V.; Cossi, M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *J. Phys. Chem. A* **1998**, *102* (11), 1995–2001. <https://doi.org/10.1021/jp9716997>.
- (73) Miertuš, S.; Scrocco, E.; Tomasi, J. Electrostatic Interaction of a Solute with a Continuum. A Direct Utilization of AB Initio Molecular Potentials for the Prediction of Solvent Effects. *Chemical Physics* **1981**, *55* (1), 117–129. [https://doi.org/10.1016/0301-0104\(81\)85090-2](https://doi.org/10.1016/0301-0104(81)85090-2).
- (74) Klamt, A.; Moya, C.; Palomar, J. A Comprehensive Comparison of the IEFPCM and SS(V)PE Continuum Solvation Methods with the COSMO Approach. *J. Chem. Theory Comput.* **2015**, *11* (9), 4220–4225. <https://doi.org/10.1021/acs.jctc.5b00601>.
- (75) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113* (18), 6378–6396. <https://doi.org/10.1021/jp810292n>.
- (76) Klamt, A. Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *J. Phys. Chem.* **1995**, *99* (7), 2224–2235. <https://doi.org/10.1021/j100007a062>.
- (77) Klamt, A.; Eckert, F. COSMO-RS: A Novel and Efficient Method for the a Priori Prediction of Thermophysical Data of Liquids. *Fluid Phase Equilibria* **2000**, *172* (1), 43–

72. [https://doi.org/10.1016/S0378-3812\(00\)00357-5](https://doi.org/10.1016/S0378-3812(00)00357-5).
- (78) Eckert, F.; Klamt, A. Fast Solvent Screening via Quantum Chemistry: COSMO-RS Approach. *AIChE J.* **2002**, *48* (2), 369–385. <https://doi.org/10.1002/aic.690480220>.
- (79) Mannhold, R.; Poda, G. I.; Ostermann, C.; Tetko, I. V. Calculation of Molecular Lipophilicity: State-of-the-Art and Comparison of LogP Methods on More than 96,000 Compounds. *Journal of Pharmaceutical Sciences* **2009**, *98* (3), 861–893. <https://doi.org/10.1002/jps.21494>.
- (80) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A Geometric Approach to Macromolecule-Ligand Interactions. *Journal of Molecular Biology* **1982**, *161* (2), 269–288. [https://doi.org/10.1016/0022-2836\(82\)90153-X](https://doi.org/10.1016/0022-2836(82)90153-X).
- (81) Meng, X.-Y.; Zhang, H.-X.; Mezei, M.; Cui, M. Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *CAD* **2011**, *7* (2), 146–157. <https://doi.org/10.2174/157340911795677602>.
- (82) Fischer, E. Einfluss Der Configuration Auf Die Wirkung Der Enzyme. *Ber. Dtsch. Chem. Ges.* **1894**, *27* (3), 2985–2993. <https://doi.org/10.1002/cber.18940270364>.
- (83) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein–Ligand Complexes. *J. Med. Chem.* **2006**, *49* (21), 6177–6196. <https://doi.org/10.1021/jm051256o>.
- (84) Koshland, D. E. Correlation of Structure and Function in Enzyme Action: Theoretical and Experimental Tools Are Leading to Correlations between Enzyme Structure and Function. *Science* **1963**, *142* (3599), 1533–1541. <https://doi.org/10.1126/science.142.3599.1533>.
- (85) Hammes, G. G. Multiple Conformational Changes in Enzyme Catalysis. *Biochemistry* **2002**, *41* (26), 8221–8228. <https://doi.org/10.1021/bi0260839>.
- (86) Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects. *J. Med. Chem.* **2006**, *49* (2), 534–553. <https://doi.org/10.1021/jm050540c>.
- (87) Jiang, F.; Kim, S.-H. “Soft Docking”: Matching of Molecular Surface Cubes. *Journal of Molecular Biology* **1991**, *219* (1), 79–102. [https://doi.org/10.1016/0022-2836\(91\)90859-5](https://doi.org/10.1016/0022-2836(91)90859-5).
- (88) Claußen, H.; Buning, C.; Rarey, M.; Lengauer, T. FlexE: Efficient Molecular Docking Considering Protein Structure Variations1 Edited by J. Thornton. *Journal of Molecular Biology* **2001**, *308* (2), 377–395. <https://doi.org/10.1006/jmbi.2001.4551>.
- (89) Alonso, H.; Bliznyuk, A. A.; Gready, J. E. Combining Docking and Molecular Dynamic Simulations in Drug Design. *Med. Res. Rev.* **2006**, *26* (5), 531–568. <https://doi.org/10.1002/med.20067>.
- (90) Flores-Bocanegra, L.; Pérez-Vásquez, A.; Torres-Piedra, M.; Bye, R.; Linares, E.; Mata, R. α -Glucosidase Inhibitors from *Vauquelinia Corymbosa*. *Molecules* **2015**, *20* (8), 15330–15342. <https://doi.org/10.3390/molecules200815330>.
- (91) Zhuang, W.-R.; Wang, Y.; Cui, P.-F.; Xing, L.; Lee, J.; Kim, D.; Jiang, H.-L.; Oh, Y.-K. Applications of π - π Stacking Interactions in the Design of Drug-Delivery Systems. *Journal of Controlled Release* **2019**, *294*, 311–326. <https://doi.org/10.1016/j.jconrel.2018.12.014>.
- (92) Iketani, S.; Forouhar, F.; Liu, H.; Hong, S. J.; Lin, F.-Y.; Nair, M. S.; Zask, A.; Huang, Y.; Xing, L.; Stockwell, B. R.; Chavez, A.; Ho, D. D. Lead Compounds for the Development of SARS-CoV-2 3CL Protease Inhibitors. *Nat Commun* **2021**, *12* (1), 2016. <https://doi.org/10.1038/s41467-021-22362-2>.

- (93) Verastegui-Omaña, B.; Rebollar-Ramos, D.; Pérez-Vásquez, A.; Martínez, A. L.; Madariaga-Mazón, A.; Flores-Bocanegra, L.; Mata, R. α -Glucosidase Inhibitors from *Malbranchea Flavorosea*. *J. Nat. Prod.* **2017**, *80* (1), 190–195. <https://doi.org/10.1021/acs.jnatprod.6b00977>.
- (94) Del Valle, P.; Martínez, A.-L.; Figueroa, M.; Raja, H.; Mata, R. Alkaloids from the Fungus *Penicillium Spathulatum* as α -Glucosidase Inhibitors. *Planta Med* **2016**, *82* (14), 1286–1294. <https://doi.org/10.1055/s-0042-111393>.
- (95) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed.; Computational science series; Academic Press: San Diego, 2002.
- (96) Shaw, D. E.; Adams, P. J.; Azaria, A.; Bank, J. A.; Batson, B.; Bell, A.; Bergdorf, M.; Bhatt, J.; Butts, J. A.; Correia, T.; Dirks, R. M.; Dror, R. O.; Eastwood, M. P.; Edwards, B.; Even, A.; Feldmann, P.; Fenn, M.; Fenton, C. H.; Forte, A.; Gagliardo, J.; Gill, G.; Gorlatova, M.; Greskamp, B.; Grossman, J. P.; Gullingsrud, J.; Harper, A.; Hasenplaugh, W.; Heily, M.; Heshmat, B. C.; Hunt, J.; Ierardi, D. J.; Iserovich, L.; Jackson, B. L.; Johnson, N. P.; Kirk, M. M.; Klepeis, J. L.; Kuskin, J. S.; Mackenzie, K. M.; Mader, R. J.; McGowen, R.; McLaughlin, A.; Moraes, M. A.; Nasr, M. H.; Nociolo, L. J.; O'Donnell, L.; Parker, A.; Peticolas, J. L.; Pocina, G.; Predescu, C.; Quan, T.; Salmon, J. K.; Schwink, C.; Shim, K. S.; Siddique, N.; Spengler, J.; Szalay, T.; Tabladillo, R.; Tartler, R.; Taube, A. G.; Theobald, M.; Towles, B.; Vick, W.; Wang, S. C.; Wazlowski, M.; Weingarten, M. J.; Williams, J. M.; Yuh, K. A. Anton 3: Twenty Microseconds of Molecular Dynamics Simulation before Lunch. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*; ACM: St. Louis Missouri, 2021; pp 1–11. <https://doi.org/10.1145/3458817.3487397>.
- (97) Pérez-Benito, L.; Keränen, H.; van Vlijmen, H.; Tresadern, G. Predicting Binding Free Energies of PDE2 Inhibitors. The Difficulties of Protein Conformation. *Sci Rep* **2018**, *8* (1), 4883. <https://doi.org/10.1038/s41598-018-23039-5>.
- (98) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12* (1), 281–296. <https://doi.org/10.1021/acs.jctc.5b00864>.
- (99) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118* (45), 11225–11236. <https://doi.org/10.1021/ja9621760>.
- (100) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities. *Expert Opinion on Drug Discovery* **2015**, *10* (5), 449–461. <https://doi.org/10.1517/17460441.2015.1032936>.
- (101) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000**, *33* (12), 889–897. <https://doi.org/10.1021/ar000033j>.
- (102) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate–DNA Helices. *J. Am. Chem. Soc.* **1998**, *120* (37), 9401–9409. <https://doi.org/10.1021/ja981844+>.
- (103) Gohlke, H.; Case, D. A. Converging Free Energy Estimates: MM-PB(GB)SA Studies on the Protein-Protein Complex Ras-Raf. *J. Comput. Chem.* **2004**, *25* (2), 238–250.

<https://doi.org/10.1002/jcc.10379>.

(104) Réblová, K.; Střelcová, Z.; Kulhánek, P.; Beššeová, I.; Mathews, D. H.; Nostrand, K. V.; Yildirim, I.; Turner, D. H.; Sponer, J. An RNA Molecular Switch: Intrinsic Flexibility of 23S rRNA Helices 40 and 68 5'-UAA/5'-GAN Internal Loops Studied by Molecular Dynamics Methods. *J Chem Theory Comput* **2010**, *2010* (6), 910–929.

(105) Homeyer, N.; Gohlke, H. Free Energy Calculations by the Molecular Mechanics Poisson–Boltzmann Surface Area Method. *Mol. Inf.* **2012**, *31* (2), 114–122.

<https://doi.org/10.1002/minf.201100135>.

(106) Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the Performance of the Molecular Mechanics/Poisson Boltzmann Surface Area and Molecular Mechanics/Generalized Born Surface Area Methods. II. The Accuracy of Ranking Poses Generated from Docking. *J. Comput. Chem.* **2011**, *32* (5), 866–877. <https://doi.org/10.1002/jcc.21666>.

(107) Sun, H.; Li, Y.; Shen, M.; Tian, S.; Xu, L.; Pan, P.; Guan, Y.; Hou, T. Assessing the Performance of MM/PBSA and MM/GBSA Methods. 5. Improved Docking Performance Using High Solute Dielectric Constant MM/GBSA and MM/PBSA Rescoring. *Phys. Chem. Chem. Phys.* **2014**, *16* (40), 22035–22045. <https://doi.org/10.1039/C4CP03179B>.