



**Facultad de Ingeniería
Escuela de Ingeniería Civil en Bioinformática**

Estudio estructural del rol inhibitorio de derivados de naftaleno frente a la enzima PLpro de SARS-CoV.

**Alumno: Luis Castillo Campos
Tutor: Dr. Julio Caballero Ruiz
Co-Tutor: Dr. José Luis Velázquez-Libera
Prof. Informante: Dra. Ariela Vergara-Jaque**

Talca - Chile
2022

CONSTANCIA

La Dirección del Sistema de Bibliotecas a través de su unidad de procesos técnicos certifica que el autor del siguiente trabajo de titulación ha firmado su autorización para la reproducción en forma total o parcial e ilimitada del mismo.



Talca, 2022

Agradecimientos

Pienso que la mayoría de las personas en su camino como estudiante de pregrado, se sienten inmersos en una montaña rusa de emociones. Es un proceso largo, lleno de interrogantes, no solo en el aspecto académico, sino que también en el ámbito personal. Muchas cosas no suceden como las esperamos, mientras que las que sí ocurren, las celebramos como si fuera la mejor victoria de la vida. Como todos, he tenido buenos y malos momentos en mi etapa universitaria, los cuales no habrían sido de la misma forma sin aquellas personas que me acompañaron en todo este viaje. Por ello, me gustaría agradecer a los causantes de todas esas palabras de ánimo, abrazos, conversaciones y risas, que contribuyeron de un modo u otro a que toda esta marcha llegue a su destino.

Primeramente, me gustaría agradecer a mis profesores de Tesis, al Dr. Julio Caballero Ruiz, por darme la oportunidad de formar parte de este proyecto y confiar en mí. Por la disponibilidad para resolver mis dudas y por los feedbacks entregados, los cuales en más de una ocasión abrieron mi mente al 100%. De la misma manera agradezco a mi co-tutor, Dr. José Luis Velázquez-Libera, por el apoyo, por estar presente y por todas las reuniones a lo largo del semestre. Es realmente un gusto ser parte de esto.

Además de ellos, quiero agradecer a la Dra. Ariela Vergara-Jaque, directora de la escuela de Ingeniería Civil Bioinformática. Es muy probable que todas las personas con las que he compartido, conozcan lo agradable que fue para mí formar parte de su grupo en dos de mis prácticas realizadas. Agradezco profundamente la posibilidad que me entregó de aprender desde mi primera práctica, por guiarme, aconsejarme y permitirme ver el lado positivo de las cosas. Por las palabras de apoyo y por alimentar mi entusiasmo por la ciencia y la investigación, de corazón, muchas gracias.

Hago extensivo mis agradecimientos a todos los docentes y funcionarios que contribuyeron tanto en mi formación profesional, como personal.

En este apartado no podían faltar aquellos con los que he compartido a lo largo de estos años, aquellos que han visto de cerca, en primera persona, algunos éxitos y por supuesto resultados no deseados. Quiero agradecer a mis amig@s, sonará cursi, pero sin ellos toda esta pasarela no habría tenido el mismo glamour y color. A Nicolás Garrido, Ricardo Hernández, Francisca Rodríguez, a las emprendedoras: Paulina Hormazabal y Victoria Fuentealba, a mi pana de batallas, Raúl Rojas; Biotecnología, Gestión de innovación y emprendimiento, Recursos humanos, Fundamentos de la administración, Ingeco, Procesos Bioindustriales, Taller de proyectos biotecnológicos y Evolución Molecular, realmente son bastantes, gracias por el aguante y la dedicación, más aún cuando estábamos en desventaja, siempre pensaré que fuimos una buena dupla. A Javier Díaz, alias “Potrillo”, Gonzalo Espinoza y Patricio Briones, quienes me han entregado buenos momentos y gratas conversaciones. Al club de guapos, Matías Rojas, Claudio Illanes y Claudio Letelier. Gracias por las palabras de ánimo y apoyo, son todos realmente un amor. Y por último, no por eso menos importante, quiero agradecer a mi bestie Paloma Espinoza, a Javi Rosales y Luciano Peña, con quienes he compartido en mayores ocasiones, ya sea en mi día a día, coincidiendo en módulos o vía redes sociales. Siento que han creído en mí, mucho más de lo que yo lo he hecho, me han entregado confianza, han lidiado con mi carácter, con mis momentos de rabia y tristeza. Realmente no se que es lo que viene, pero independiente de ello, voy a recordar siempre lo mucho que han contribuido en mi vida personal y académica. Mucho amor y éxito en todo lo que se propongan.

Agradezco también a mi novio Cristófer Vásquez, quien me ha apoyado por más de 5 años. Me incentivaste a priorizar mis estudios y a lograr mis sueños. Viste de cerca el esfuerzo y dedicación que he puesto, sobre todo en ramos que me resultaban complicados. Estuviste conmigo cuando me encerraba intentando comprender algunas materias. Me entregaste la tranquilidad, amor y apoyo que necesitaba. Te amo y agradezco infinitamente.

Finalmente quiero agradecer a mi familia, en especial a mi círculo más cercano: mamá, papá y hermano. Dicen que el amor de una familia no es cuantificable y así es, para mí ha sido infinito. Gracias por todo lo que se han dedicado a mí, por haberme animado a no rendirme nunca. Por esos interminables “tú puedes”, “eres capaz”. Gracias por haberme hecho ver cosas sobre mí y por hacerme mejorar. Estoy seguro que todas las cosas que hemos vivido,

me han ayudado a llegar a este punto. Ahora no soy mis miedos, mis errores o inseguridades. No soy otra cosa que lo que muestro ahora mismo. Y en este momento, tengo la oportunidad de amar profundamente a otros, las cosas que hago y ser una persona que me hace sentir orgulloso. Esto se los debo, vieron cómo empezó este viaje y tengo la fortuna de que podrán verlo acabar. Esta tesis va para ustedes, los amo siempre.

Resumen: La proteasa SARS-CoV PLpro está asociada a diferentes funciones que ocasionan el antagonismo de la respuesta inmune innata antiviral del huésped, como lo es la deubiquitinación y la desISGilación de las proteínas de las células de este. Dada la función que desempeña, la atención en la PLpro como diana terapéutica para el diseño de fármacos contra el SARS-CoV, MERS-CoV y actualmente el SARS-CoV-2, ha ido en aumento y se han reportado compuestos candidatos para detener la actividad de esta enzima. Entre estos, se reportó un conjunto de inhibidores no covalentes basados en naftaleno, que incluye compuestos que muestran una alta capacidad inhibitoria contra la PLpro de SARS-CoV. La comprensión de las características estructurales que facilitan la interacción entre la PLpro de coronavirus y sus ligandos es de gran utilidad para entender el funcionamiento de esta enzima y su capacidad para ser intervenida con inhibidores más efectivos. En este trabajo se revelaron características estructurales de los derivados del naftilo que explican su actividad contra la enzima PLpro de SARS-CoV, a través de un estudio detallado de las interacciones ligando-receptor. El protocolo empleado se inició con el método docking molecular, el cual permitió considerar las orientaciones de los ligandos en el sitio de unión de la proteasa. Posterior a ello, se realizaron diversos análisis quimiométricos: se utilizó el servidor LigRMSD para comparar las orientaciones de los inhibidores estudiados, se realizó un análisis de huellas de interacción para describir las interacciones recurrentes entre los residuos de la proteína y los grupo del ligando en los complejos y finalmente se construyeron modelos 3D-QSAR para explicar la relación entre la estructura de los compuestos y su actividad inhibitoria. Se describieron las interacciones entre diferentes porciones del ligando y varios residuos claves (Asp165, Tyr269, Gln270 y Tyr265), además de la ocupación de estos inhibidores en los subsitios S3 y S4 de la proteasa. Se encontró que la serie de compuestos derivados del naftaleno adoptan orientaciones similares en el sitio de unión de PLpro, con una alta semejanza con las orientaciones de los ligandos co-cristalizados. Sumado a esto, los modelos 3D-QSAR construidos en base a diferentes estrategias de alineamiento, no cumplieron en su totalidad con los requisitos estadísticos para ser considerados modelos con alto poder predictivo. En base a estos resultados, se propone el uso de una nueva metodología para explicar la actividad biológica de la serie de inhibidores.

Abstract: The SARS-CoV protease PLpro is associated with different functions that cause antagonism of the host innate antiviral immune response, such as deubiquitination and deISGylation of host cell proteins. Given the role it plays, attention on PLpro as a therapeutic target for the design of drugs against SARS-CoV, MERS-CoV and currently SARS-CoV-2, has been increasing and candidate compounds have been reported to reduce the activity of this enzyme. Studying the structural features that facilitate the interaction between coronavirus PLpro and its ligands is of great utility in understanding the functioning of this enzyme and its ability to be targeted with more effective inhibitors. In this work, structural features of naphthyl derivatives against SARS-CoV PLpro enzyme were revealed through a detailed study of ligand-receptor interactions. The protocol started with the molecular docking method, which allowed to consider the ligand orientations at the protease binding site. Subsequently, several chemometric analyses were performed: the LigRMSD server was used to compare the orientations of the studied inhibitors, an interaction fingerprint analysis was performed to describe the recurrent interactions between the protein residues and the ligand groups in the complexes; finally, 3D-QSAR models were constructed to explain the relationship between the structure of the compounds and their inhibitory activity. The interactions between different portions of the ligand and key residues (Asp165, Tyr269, Gln270 and Tyr265) were described, in addition to the occupancy of these inhibitors at the S3 and S4 sub-sites of the protease. It was found that the serie of naphthalene-derived compounds adopted similar orientations at the PLpro binding site, with high similarity to the orientations of co-crystallized ligands. In addition, the 3D-QSAR models constructed based on different alignment strategies did not fully meet the statistical requirements to be considered models with high predictive power. Based on these results, a new methodology is proposed to explain the biological activity of the inhibitor series.

ÍNDICE DE CONTENIDOS

1. Introducción	10
1.1 Síndrome respiratorio agudo grave causado por coronavirus	10
1.2 Organización molecular de la poliproteína viral	11
1.3 Aspectos estructurales y actividad de la proteasa PLpro	13
1.4 Inhibidores de SARS-CoV PLpro derivados del naftaleno	15
1.5 Alcance de esta tesis	20
2. Planteamiento de hipótesis	22
3. Objetivos	22
4. Materiales y métodos.	23
4.1 Preparación de compuestos derivados del naftaleno y estructura de la proteína	24
4.2 Docking molecular	32
4.3 Análisis post-docking	34
4.3.1 LigRMSD	34
4.3.2 Interaction fingerprinting	35
4.4 3D-QSAR	35
5. Resultados y Discusión	38
5.1 Poses de acoplamiento molecular	38
5.2 Comparación entre las poses	44
5.3 Interacciones con residuos en el sitio de unión de SARS-CoV PLpro	48
5.4 Modelos de correlación 3D para describir actividades diferenciales	52
6. Conclusiones y propuestas futuras.	59
6.1 Conclusiones	59
6.2 Propuestas futuras	60
7. Referencias	61

ÍNDICE DE FIGURAS

Figura 1.1. Escisiones proteolíticas por PLpro y 3CLpro	12
Figura 1.2. Organización de dominios estructurales de SARS-CoV PLpro	13
Figura 1.3. Ejemplos de derivados de naftaleno, inhibidores de la PLpro de SARS-CoV	16
Figura 1.4. Características estructurales de los compuestos derivados de naftaleno	17
Figura 1.5. Inhibidores derivados del compuesto 15g	18
Figura 4.1. Metodología	24
Figura 4.2. Residuos participantes en la interfaz proteína-ligando	33
Figura 5.1. Estructuras acopladas dentro del sitio de unión de SARS-CoV PLpro	39
Figura 5.2. Serie de compuestos derivados del naftaleno en el sitio de unión de PLpro	41
Figura 5.3. Mejores poses obtenidas para la estructura monomérica y dimérica de PLpro	44
Figura 5.4. Alineamiento estructural de compuestos co-cristalizados	45
Figura 5.5. Similitud de las poses de docking respecto a las referencias 3E9S y 4OW0	47
Figura 5.6. Interacciones presentes en los compuestos acoplados	49
Figura 5.7. Alineamiento de compuestos derivados del naftaleno	53
Figura 5.8. Errores en la predicción de la actividad de los compuestos estudiados	55

ÍNDICE DE TABLAS

Tabla 4.1. Primer conjunto de compuestos derivados del naftaleno	25
Tabla 4.2. Segundo conjunto de compuestos derivados del naftaleno	26
Tabla 4.3. Tercer conjunto de compuestos derivados del naftaleno	27
Tabla 4.4. Cuarto conjunto de compuestos derivados del naftaleno	28
Tabla 4.5. Valores de RMSD de los residuos en la interfaz proteína-ligando	30
Tabla 5.1. Resultados del servidor LigRMSD para compuestos co-cristalizados	45
Tabla 5.2. Información estadística de los modelos 3D-QSAR	54
Tabla 5.3. Variación del parámetro estadístico Q2, en ausencia de diferentes compuestos	56

1. INTRODUCCIÓN

1.1. Síndrome respiratorio agudo grave causado por Coronavirus

Las pandemias son, en su mayoría, brotes de infecciones que se generalizan como resultado de la propagación de una infección entre humanos. El término "pandemia" no ha sido definido a gran escala por muchos textos médicos, pero presenta algunas características claves, como la amplia extensión geográfica, el movimiento de la enfermedad, la novedad, la gravedad, las altas tasas de ataque y explosividad, la mínima inmunidad de la población, la infecciosidad y el contagio, las cuales se han asociado con impactos negativos en la salud, la economía, la sociedad y la seguridad de las comunidades nacionales y mundiales (Qui et al., 2017). Se han registrado muchos brotes de enfermedades y pandemias importantes en la historia, como la gripe española (Taubenberger & Morens, 2006), la gripe de Hong Kong (Cockburn et al., 1969), el H7N9 (Watanabe et al., 2014), el ébola (Del Rio & Guarner, 2015) y el SARS (Cherry & Krogstad, 2004), siendo este último el centro de atención a nivel global desde fines del año 2019.

El síndrome respiratorio agudo severo (SARS) es la primera enfermedad infecciosa emergente del siglo XXI que ha sido altamente transmisible y mortal, siendo informada por primera vez en la provincia de Guangdong, China, en noviembre de 2002 (Loeb, 2006), afectando a tres continentes e infectando a más de 8.000 personas con una tasa de mortalidad global del 9.2%, (CDC, 2003).

Estudios desarrollados en torno a la primera variante de la enfermedad determinaron que el agente etiológico del SARS corresponde al coronavirus (CoV) nombrado SARS-CoV, el cual se aisló a partir de pacientes con síndrome respiratorio agudo. El SARS-CoV y las variantes posteriores son considerados virus altamente patógenos que presentan amenazas a gran escala debido a dos factores: (1) no se cuenta con avances terapéuticos en el momento en que afectan a la población (vacunas) y (2) el virus proviene de una familia caracterizada por sus frecuentes mutaciones (lo cual afecta el desarrollo de fármacos específicos). De esta forma, la única opción de detener la propagación del virus es mediante medidas de cuarentena rigurosas (Demmler & Ligon, 2003).

Los coronavirus (CoV) se han convertido en los principales patógenos de los brotes emergentes de enfermedades respiratorias. Son miembros de la subfamilia *Coronavirinae* (familia *Coronaviridae*), la cual consta de cuatro géneros: Alphacoronavirus, Betacoronavirus, Gammacoronavirus y Deltacoronavirus, siendo los dos primeros capaces de causar infecciones agudas y prolongadas del tracto respiratorio en mamíferos, particularmente en humanos (Cui & Shi, 2019). Estos virus poseen estructuras esféricas, aproximadamente de 125 nm de diámetro, rodeadas de cúspides que se proyectan desde la superficie, dando la apariencia de una corona solar, lo que determina el nombre de “coronavirus”. En su interior se encuentra el RNA monocatenario, de sentido positivo, que puede aislarse en diferentes especies animales. Este logra atravesar las barreras de las especies y causar en los seres humanos enfermedades que van desde un resfriado simple, hasta patologías más graves (Perlman & Netland, 2009).

1.2 Organización molecular de la poliproteína viral

Investigaciones realizadas hasta la fecha han revelado procedimientos importantes a considerar para detener la replicación viral de los coronavirus. Uno de los procesos esenciales en el ciclo de vida del coronavirus es la escisión de una poliproteína viral multidominio en 16 componentes maduros denominados proteínas no estructurales (Nsps), que se ensamblan en complejos para ejecutar la síntesis de ARN viral (Baker, 2008). Incluidas dentro de este complejo residen dos cisteínas proteasas, necesarias para formar el ensamble replicasa de los virus, integrado por una proteasa similar a la papaína (PLpro) y una proteasa similar a 3C (3CLpro) (Yang et al., 2006). PLpro es responsable de la escisión de las tres primeras posiciones de su poliproteína, mientras que 3CLpro escinde las 11 ubicaciones restantes, liberando un total de 16 proteínas no estructurales (Nsp1 a Nsp16) (Figura 1.1) (Lee et al., 2015).

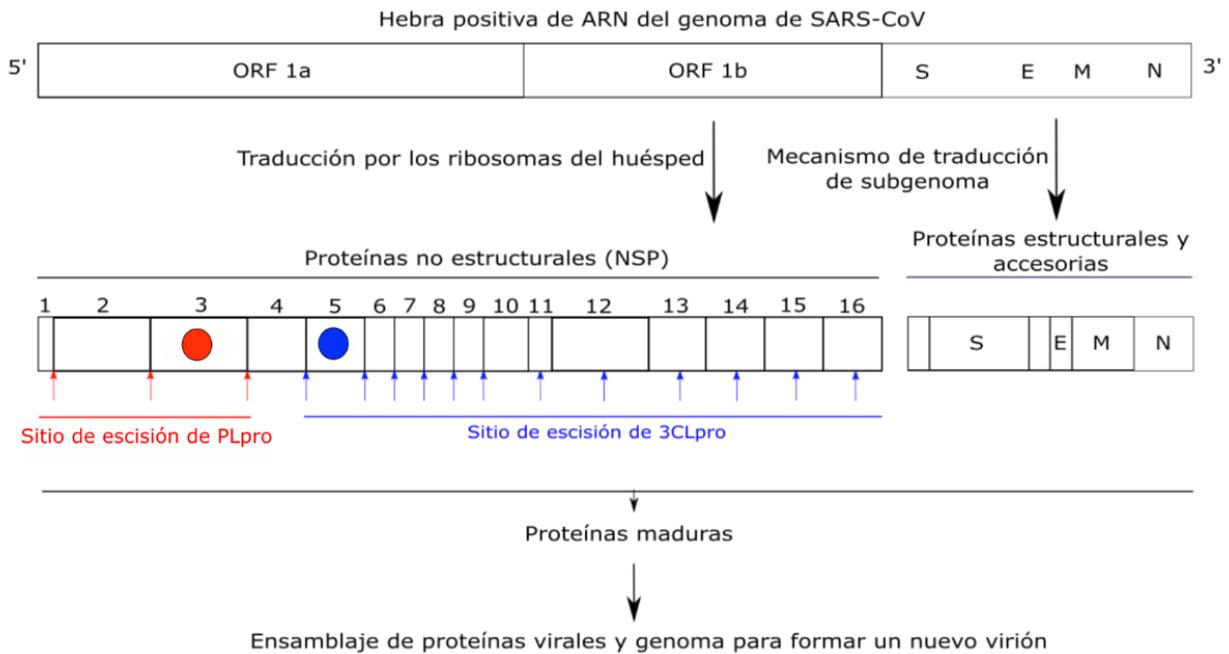


Figura 1.1: Escisiones proteolíticas por PLpro y 3CLpro. El ARN de SARS-CoV ingresa a la célula huésped, utilizando los ribosomas de este para iniciar el proceso de traducción, el cual resulta en 16 proteínas no estructurales tras la escisión autoproteolítica por las enzimas PLpro y 3CLpro. Las flechas coloreadas indican los sitios de escisión de cada una de las proteasas. Los círculos rojo y azul, corresponden a las posiciones de PLpro y 3CLpro, respectivamente. Las proteínas estructurales (S = proteína Spike, E = proteína de envoltura, M = proteína de membrana y N = proteína de nucleocápside) se traducen mediante un mecanismo de traducción de subgenoma (Mody et al., 2021). Todas las proteínas maduras se ensamblan junto con el genoma ARN monocatenario de sentido positivo para formar un nuevo virión.

Desde el año 2003, la estructura y actividad de la proteasa 3CLpro se encuentra reportada y el conocimiento sobre sus aspectos estructurales y mecánicos han proporcionado múltiples vías como puntos de partida para el diseño basado en la estructura de compuestos antivirales dirigidos contra el sitio activo de esta proteasa (Ghosh et al., 2006). Otro de los objetivos igual de viable para la intervención terapéutica del SARS, basándose en los eventos bioquímicos críticos para la replicación viral, corresponde a PLpro. Esta proteína, además de desempeñar un papel esencial durante la replicación del virus, es una enzima clave en la patogénesis de los coronavirus, asociándose al procesamiento de la poliproteína replicasa viral (Harcourt et al., 2004), deubiquitinación (Barretto et al., 2005) y deISGilación (Lindner

et al., 2007) de las proteínas de la célula huésped. Estas actividades enzimáticas desencadenan el antagonismo de la respuesta inmune innata antiviral del huésped.

1.3 Aspectos estructurales y actividad de la proteasa PLpro

Considerando la acción de la proteasa PLpro, los análisis estructurales y de actividad enzimática han sido fundamentales para el diseño y síntesis de inhibidores contra esta proteína. PLpro posee estructuras cristalográficas reportadas y su arquitectura consta de cuatro dominios, donde tres de ellos dan lugar a una conformación de mano derecha con diferentes dominios de palma, compuesta de una hoja β de seis hebras (β 8-13); el pulgar formado por cuatro hélices prominentes (α 4-7) y los dedos, constituidos por una hoja β de cuatro hebras, retorcidas y antiparalelas (β 4-7). El cuarto de estos dominios corresponde a un dominio terminal independiente similar a los dominios de ubiquitina (Ratia et al, 2006) (Figura 1.2). Actualmente, existe evidencia experimental de que SARS-CoV PLpro es capaz de reconocer e hidrolizar sustratos que contengan la secuencia consenso LXGG, siendo éste un motivo de escisión reconocido y que coincide con el perfil de especificidad (LRGG) de las enzimas desubiquitantes (DUB) de HAUSP (Hu et al, 2002). Además de su actividad DUB, PLpro también posee actividad desISGilante que se requiere para la replicación viral y la evasión de la respuesta inmune del huésped (Clemente et al., 2020).

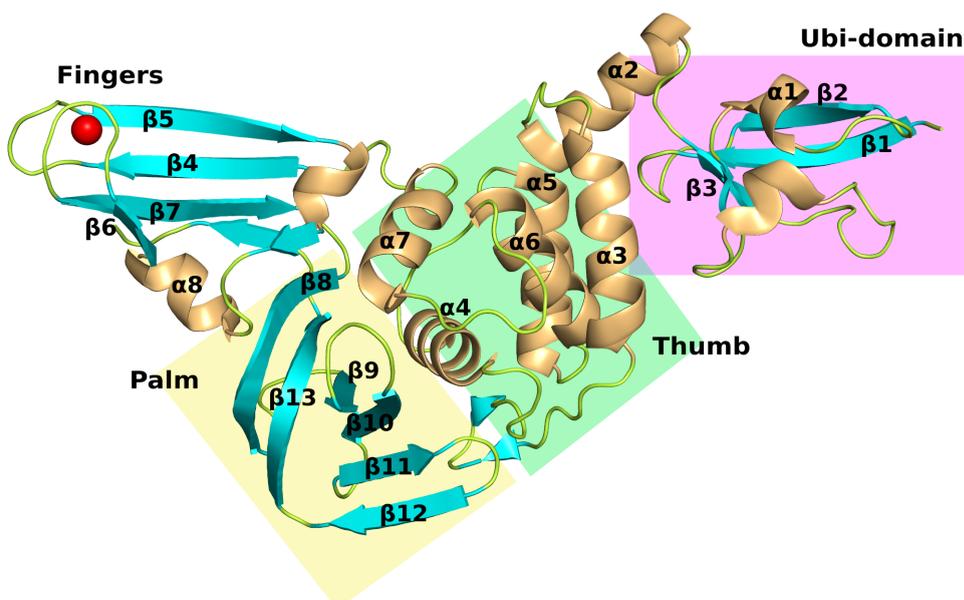


Figura 1.2: Organización de dominios estructurales de SARS-CoV PLpro. La ubicación del dominio ubiquitina (Ubl) se representa mediante el cuadro rosado, mientras que el las regiones que representan el pulgar y la palma se muestran por medio de los recuadros verde y amarillo, respectivamente. Las hélices α aparecen en color naranja y las láminas β en color azul, enumeradas y representadas como cartoons. Finalmente, el átomo de zinc presente en la estructura, se grafica de color rojo en la zona superior izquierda de la figura.

En cuanto al sitio activo, PLpro presenta una tríada catalítica compuesta por los aminoácidos Cys112-His273-Asp287, situándose el primer residuo (Cys112) en el extremo N-terminal del dominio del pulgar, resultando el átomo de azufre de la cadena lateral en interacción con el átomo de nitrógeno de la histidina catalítica (His273). Por otra parte, uno de los átomos de oxígeno de la cadena lateral del ácido aspártico (Asp287) se encuentra situado a una distancia de 2.7 Å con respecto al nitrógeno de la histidina catalítica en el dominio de la palma, formando un enlace de hidrógeno. Este arreglo estructural resulta en una triada formada por estos residuos catalíticos en la intersección de los dominios del pulgar y palma (Barretto et al., 2005).

Los residuos que conforman la tríada catalítica no son los únicos importantes en la estructura de PLpro, pues su acción conjunta con el “bucle de bloqueo” (BL2) es clave al momento de analizar el mecanismo de unión de potenciales inhibidores. El BL2, comprende a los residuos 267 - 272 de la estructura y presenta una longitud similar a los bucles del mismo tipo presentes en HAUSP, teniendo ambos dos glicinas en cada uno de los extremos (G267 - G272). Estas glicinas son 2 de los 11 residuos que se conservan entre PLpro y HAUSP, siendo significativos para impartir flexibilidad a la región del bucle. Finalmente, un aminoácido catalogado como esencial para la función enzimática de SARS-CoV PLpro, corresponde a W107, el cual sería partícipe en la estabilización de intermediarios tetraédricos cargados negativamente formados durante la catálisis (Ratia et al., 2006).

Si bien es esencial la presencia del residuo de triptófano W107 para SARS-CoV PLpro, no lo es para MERS-CoV PLpro. El análisis de la conservación de aminoácidos en los subsitios enzimáticos del MERS-CoV PLpro revela la sustitución de los residuos E168, Y265 y W107 en SARS-CoV PLpro por los residuos R170, F271 y L108 en MERS-CoV PLpro (Baez-Santos et al.,

2014). Las diferencias que se presentan en los aminoácidos de los sitios enzimáticos de las PLpro en distintos coronavirus explicarían la selectividad que tienen variados compuestos diseñados para ser inhibidores de SARS-CoV PLpro frente a la proteasa de otros coronavirus.

1.4 Inhibidores de SARS-CoV PLpro derivados del naftaleno

La gran cantidad de funciones realizadas por PLpro en la patogénesis viral, apuntan a esta proteasa como un objetivo atractivo para diseñar fármacos antivirales. Hasta la fecha se han realizado investigaciones para identificar inhibidores potenciales de PLpro por medio de estudios que incluyen el uso de técnicas computacionales (Baez-Santos et al., 2014; Ghosh et al., 2009, 2010). El conocimiento de las relaciones existentes entre los genomas SARS-CoV-2, MERS-CoV y SARS-CoV al codificar una única PLpro, enfoca a esta enzima como una diana tentativa para la creación de inhibidores. La utilización de métodos computacionales para investigar sobre un gran número de compuestos sintéticos, tal como lo es el cribado de alto rendimiento (HTS), ha permitido a la comunidad científica realizar correlaciones estructura-actividad de diversos candidatos para inhibir PLpro. Dada la homología que presenta la PLpro con las enzimas deubiquitinantes humanas y a que existe sobre el 50% de DUB en humanos que también son cisteínas proteasas (Daviet & Colland, 2008), es de gran importancia que los compuestos que se desarrollen como inhibidores de PLpro sean altamente selectivos.

Existen investigaciones asociadas al diseño de potenciales moléculas capaces de inhibir el mecanismo de acción de PLpro. Ratia y col. (2008) reportaron la evaluación de 50.080 compuestos diversos, de los cuales solo el 0,04% (equivalente a 17 de estos), lograron tener una capacidad inhibitoria >35% sobre PLpro, tras la realización de un cribado primario. Ensayos posteriores permitieron descubrir el compuesto GRL0617, un derivado de naftaleno que resultó ser un inhibidor potente frente PLpro. A partir de este compuesto, los autores desarrollaron una serie de nuevos inhibidores de PLpro (ejemplo en la [Figura 1.3](#)) (Baez-Santos et al., 2014; Ghosh et al., 2009, 2010).

Ratia y col. introdujeron estos compuestos con la finalidad de que actúen como inhibidores competitivos no covalentes/reversibles de PLpro, al unirse a los subsitios S3-S4 de la enzima,

generando cambios conformacionales en el bucle articulado por dos residuos de glicina (G267 y G272). Tales compuestos inducen la reorientación de las cadenas laterales de Y269, generando el cierre del bucle de modo que sostenga el inhibidor al cuerpo de la proteína, logrando así inhibir el sitio activo de PLpro.

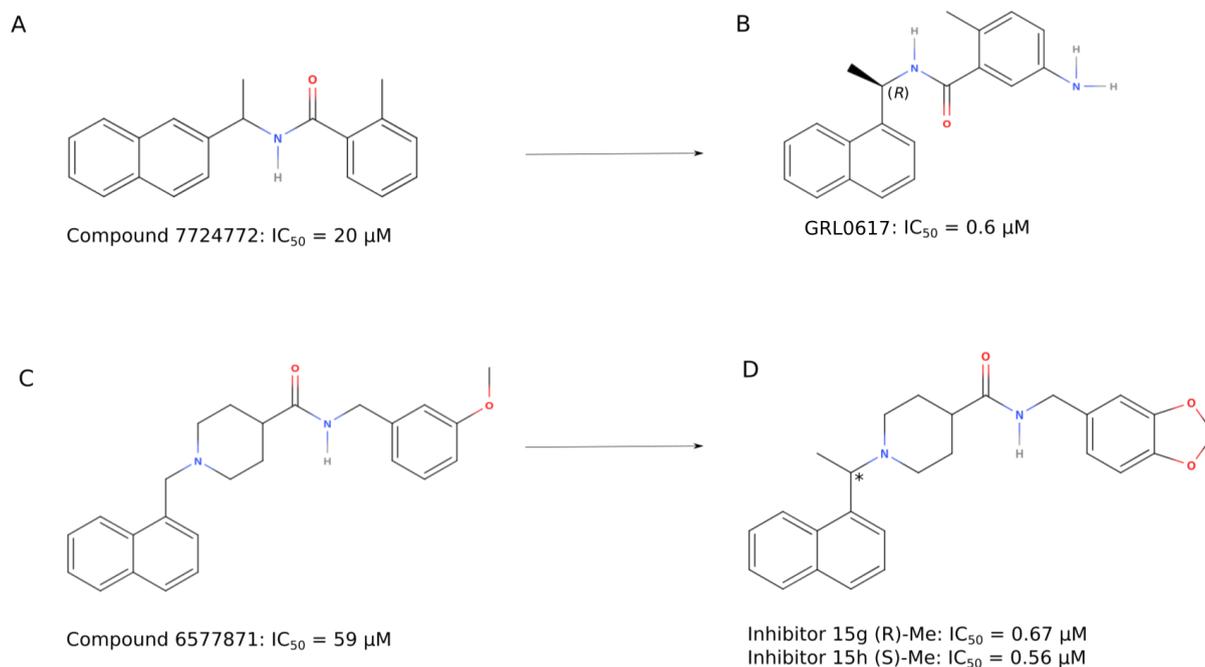
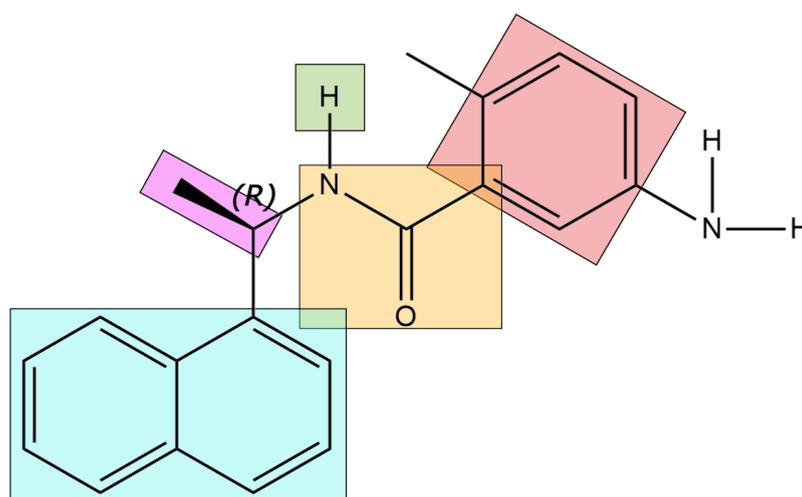


Figura 1.3: Ejemplos de derivados de naftaleno, inhibidores de la PLpro de SARS-CoV. (A) Compuesto 7724772 obtenido por HTS, a partir del cual se desarrolló el inhibidor GRL0617 (B). (C) Compuesto 6577871 obtenido por HTS, a partir del cual se desarrollaron los inhibidores 15g y 15h. El centro quiral de los isómeros derivados del compuesto 6577871 se indica con un asterisco.

La información sobre los inhibidores eficaces de PLpro de esta serie, permitió a los investigadores desarrollar mejores candidatos. Los análisis de estructura-actividad aplicados a los datos de HTS permitieron la exploración de otros sustituyentes en determinadas posiciones de la estructura 7724772 ($IC_{50} = 20,1 \pm 1,1 \mu\text{M}$) (Figura 1.3A). A partir de estos estudios se obtuvo el compuesto GRL0617, que aumentó considerablemente la potencia inhibitoria frente a PLpro ($IC_{50} = 0,6 \pm 0,1 \mu\text{M}$) (Baez-Santos et al., 2014). Sumado a la síntesis de GRL0617, también surgió 6577871 mediante el uso de HTS, exhibiendo un valor de IC_{50} elevado ($59 \mu\text{M}$). La optimización de este compuesto derivó en el potente inhibidor 15g (IC_{50}

= 0,32 μM), que en conjunto con su enantiómero 15h ($\text{IC}_{50} = 0.56\mu\text{M}$) fueron capaces de inhibir la replicación viral de SARS-CoV.

Los compuestos GRL0617, 15g y 15h comparten una serie de características químicas y estructurales. Presentan el grupo naftilo adyacente a un centro estereogénico que contiene un grupo metilo y un donante de enlace de hidrógeno centrado en nitrógeno. Además, comparten conectores de carboxamida de orientación opuesta a los grupos bencenoides (Figura 1.4). Relaciones de estructura-actividad han demostrado que la presencia de un grupo metilo en la posición mencionada anteriormente es un factor clave para la potencia de esta serie de inhibidores (Ghosh et al., 2006).



GRL0617

Figura 1.4: Características estructurales que comparten cada uno de los compuestos derivados del naftaleno, utilizando como ejemplo el inhibidor GRL0617. Grupo naftilo representado por recuadro celeste. Grupo metilo en centro estereogénico y donante de enlace de hidrógeno, destacados en cuadros rosado y verde, respectivamente. Finalmente, grupo carboxamida en coloración naranja adyacente al grupo bencenoide (rojo claro).

En el año 2014, Baez-Santos y col. reportaron nuevos compuestos de esta serie obtenidos a partir del inhibidor 15g, presentando variaciones estructurales importantes. Uno de estos corresponde al inhibidor 3e ($\text{IC}_{50} = 0,39 \pm 0,01 \mu\text{M}$), el cual en lugar de un grupo dioxolano enlazado al anillo aromático, presenta un sustituyente acetamido en la posición meta del

benceno. Además, obtuvieron los derivados 3j y 3k que poseen un sustituyente fluoruro en las posiciones meta y para del benceno, respectivamente; presentando el último de estos una mejora apreciable en la actividad ($IC_{50} = 0,15 \pm 0,01 \mu M$) en comparación con cualquiera de las otras moléculas reportadas hasta el momento (Baez-Santos et al., 2014) (Figura 1.5).

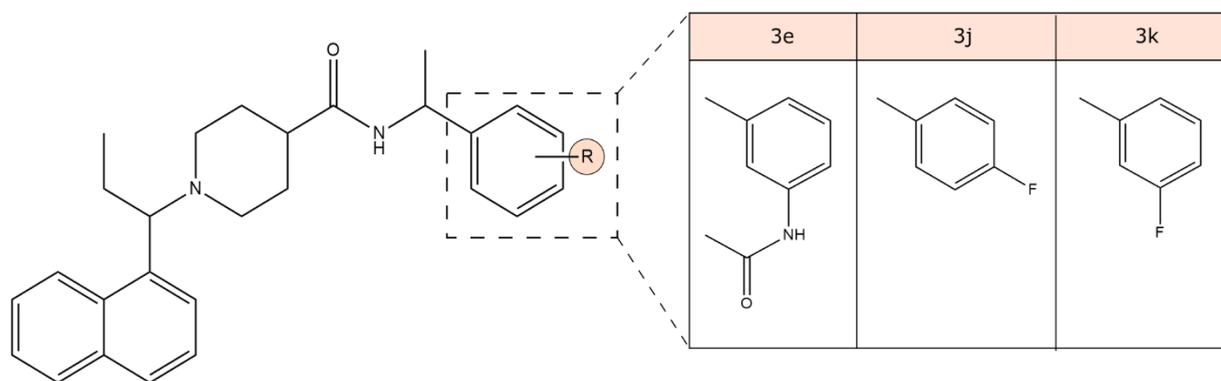


Figura 1.5: Inhibidores derivados del compuesto 15g. La estructura general del compuesto es similar a la de todas las moléculas derivadas del naftilo, presentando variaciones en el sustituyente R para cada uno de los casos. Como ejemplo se representan los compuestos 3e (R = 3-NHCOMe), 3j (R = 4-F) y 3k (R = F) (parte derecha de la imagen).

Los derivados de naftaleno de esta segunda generación no exhibieron citotoxicidad y mostraron alta selectividad frente a la diana terapéutica. Además, presentaron una estabilidad metabólica significativamente mejorada en comparación con las moléculas sintéticas de las que provienen, como 15g, por lo que podrían ser utilizados como un punto de partida importante para avanzar en el desarrollo de fármacos capaces de combatir el coronavirus.

Si bien la estructura general de la SARS-PLpro es similar a la de la MERS-PLpro, la comparación estructural minuciosa entre las dos enzimas PLpro utilizando todas las estructuras disponibles, reveló diferencias importantes para el desarrollo de inhibidores (Lee et al., 2015). Basándose en ello y en los antecedentes asociados a los residuos con los cuales interactúan las moléculas derivadas de naftilo, estos serían efectivos para SARS-CoV-PLpro y SARS-CoV-2-PLpro, no así para MERS-CoV-PLpro. Esto se debe a que MERS-CoV-PLpro

presenta una variación en los residuos responsables de la unión del inhibidor en el bucle de bloqueo BL2, reemplazando los aminoácidos Y269 y Q270 (en SARS-CoV-PLpro) por T274 y A275 en MERS-CoV PLpro, lo que dificulta la formación de interacciones de unión críticas para tipos similares de inhibidores (Freitas et al., 2020).

Recientemente, a través de una caracterización bioquímica de las actividades desubiquitinantes y desISGilantes en SARS-CoV-2, se reveló que PLpro, para este nuevo coronavirus, parece funcionar más parecido a la MERS-CoV PLpro que a su contraparte en SARS-CoV, pese a que difiere con la primera en 54 residuos. A pesar de esta diferencia, los aminoácidos que recubren el sitio activo, así como también los sitios S3 y S4 cercanos son idénticos entre SARS-CoV y SARS-CoV-2, incluyéndose de esta forma los residuos en el bucle BL2, imprescindibles para unir inhibidores basados en naftaleno.

Conociendo estos antecedentes y aun cuando el interés en las terapias contra los coronavirus había disminuido debido a la efectiva detención del brote de SARS-CoV en el año 2003; hoy en día existe una necesidad urgente de terapias contra SARS-CoV-2, basadas en el desarrollo de fármacos antivirales capaces de inhibir e impedir la replicación de los coronavirus en humanos. Como se ha presentado con anterioridad, centrar la atención en la diana proteica PLpro usando la información reunida en los inhibidores derivados de naftaleno contra la variante presente en SARS-CoV, es relevante para investigar a futuro la acción de esta clase de inhibidores contra el coronavirus SARS-CoV-2. Cabe destacar que la serie de compuestos derivados de naftaleno seleccionados para el presente estudio, es la más numerosa que se ha reportado hasta la fecha contra la PLpro de coronavirus. Esta está conformada por una serie de compuestos con diversas características químicas que determinan sus capacidades inhibitorias. Por lo tanto, es de gran relevancia para la presente investigación revelar cómo las variaciones de sustituyentes químicos en estos compuestos determinan las variaciones en las actividades frente a PLpro. Esto contribuiría a racionalizar el proceso de optimización de moléculas con actividad inhibitoria contra PLpro y a entender cómo establecer interacciones favorables en el sitio de unión de la proteína para esta clase de compuestos.

1.5 Alcance de esta tesis

Los coronavirus zoonóticos son importantes patógenos virales cuya especie más reciente, el SARS-CoV-2, ha sido causante de una emergencia a nivel mundial debido a su rápida propagación. La infección por este tipo de virus es responsable del Síndrome respiratorio agudo severo (SARS), enfermedad que ha provocado un gran número de muertes en diferentes zonas del mundo. Como se mencionó anteriormente, a través del presente estudio se centrará la atención en la proteasa PLpro del virus SARS-CoV, la cual es participe en la replicación viral y tiene una función en la evasión de la respuesta inmune innata del huésped. Por esta razón, el diseño de inhibidores de SARS-CoV PLpro es un objetivo crítico para luchar contra las infecciones por coronavirus.

SARS-CoV PLpro comparte una similitud de secuencia del 82.80% con la cepa homóloga de SARS-CoV-2 y un 100% de identidad en los residuos que forman el sitio de unión de moléculas pequeñas (Amin et al., 2021). La importante similitud entre SARS-CoV PLpro y SARS-CoV-2 PLpro inspira a la comunidad a investigar más a fondo estas proteasas y así también, los inhibidores reportados hasta la fecha, con la esperanza de descubrir compuestos capaces de actuar sobre las proteínas de SARS-CoV-2. Bajo este marco investigativo, diferentes autores han utilizado como punto de partida a los inhibidores derivados del naftaleno y la información previa reportada, para abordar el escenario de salud actual.

Pitsillou y col. investigaron las características de unión de los compuestos GRL-0617 y 3k contra la proteasa similar a la papaína (PLpro) de SARS-CoV-2 y evaluaron su efecto sobre la actividad desubiquitinante de la PLpro. Sus hallazgos, indicaron además, otros candidatos que podrían actuar como inhibidores de la PLpro, que se consideran compuestos líderes potenciales (Pitsillou et al., 2020). A mediados del año 2020, Bosken sometió el compuesto 3k de la serie congénica a simulación molecular frente a SARS-CoV-2 PLpro. Este estudio tuvo la finalidad de explorar la interfaz ligando-receptor y ofrecer estrategias para mejorar el andamio de los compuestos derivados del naftaleno mediante la adición de sustituyentes y

la participación de residuos en un espacio infrautilizado del sitio de unión (Bosken et al., 2020).

Gosh por otra parte, a inicios del año 2021, utilizó un enfoque computacional 2D, basado en fragmentos, para correlacionar estructuras moleculares y su actividad biológica, utilizando 29 inhibidores de la serie reportada por Ratia. De este estudio se desprende la importancia del fragmento de benzamida para la formación de enlaces con los residuos D165 y Q270. Así mismo, declaran al sustituyente naftaleno como responsable de las actividades inhibitorias de SARS-CoV PLpro (Gosh et al., 2021).

Como se puede observar, el mundo científico ha utilizado enfoques multidisciplinares que involucran metodologías de química computacional para contribuir a este tema de investigación. El efecto inhibitorio de los derivados del naftaleno ya ha sido estudiado a nivel experimental y parte de estos compuestos han sido sometidos a análisis de relación estructura-actividad (SAR y QSAR). Sin embargo, es necesario caracterizar los complejos formados por PLpro y estos inhibidores reversibles, con la finalidad de conocer en detalle el modo de unión de los compuestos y comprender de qué manera contribuyen los diferentes residuos de la proteasa en la formación de interacciones que modulan la inhibición de PLpro.

De esta forma, considerar las diferentes estructuras de la PLpro en conjunto con la serie de compuestos derivados del naftaleno, permitirá describir características farmacofóricas y químicas esenciales para la formación del complejo proteína-ligando. Para esto, se obtendrán las poses activas de la serie congénica más amplia de inhibidores de la PLpro de SARS-CoV mediante docking y se realizará un protocolo que incluye una comparación entre las orientaciones de los ligandos utilizando el servidor LigRMSD (Velázquez-Libera et al., 2020), análisis de huellas dactilares de interacción (interaction fingerprint) y 3D-QSAR. En los últimos tiempos, el interés por los derivados de naftilo como inhibidores de PLpro ha aumentado enormemente, por lo que esta información será relevante para estudios posteriores orientados al diseño racional de potentes inhibidores de PLpro.

2. HIPÓTESIS

Considerando las diferencias entre las estructuras de los compuestos derivados de naftaleno, se plantea que es posible dilucidar a través de modelado computacional, el efecto de los sustituyentes en las diferentes actividades de los compuestos actuando frente a la proteasa PLpro de SARS-CoV.

3. OBJETIVOS

- Preparar las estructuras de los compuestos derivados del naftaleno y las estructuras proteicas de SARS-CoV PLpro, considerando los estados de protonación a pH fisiológico.
- Realizar estudios de docking molecular entre los compuestos derivados de naftaleno y la proteína PLpro de SARS-CoV.
- Realizar el análisis de los resultados de docking en cuanto a la similitud entre las poses y las interacciones recurrentes entre los ligandos y los residuos del sitio activo de PLpro.
- Desarrollar modelos 3D-QSAR para explicar cómo contribuyen los diferentes grupos químicos a las diferencias de actividad de los compuestos.

4. MATERIALES Y MÉTODOS

Para lograr cada uno de los objetivos propuestos, se siguió el esquema de trabajo expuesto en la [figura 4.1](#). En este estudio, se identificaron series congéneres para construir modelos de correlación que describan las actividades diferenciales de los compuestos derivados del naftaleno frente a la proteasa SARS-CoV PLpro, utilizando herramientas de acoplamiento molecular y QSAR. La mayoría de los programas que fueron empleados para el estudio se ejecutaron en la interfaz Maestro de Schrödinger (Maestro, Schrodinger, LLC, Nueva York, NY, EE. UU.). Las moléculas sintéticas se dibujaron por medio del editor de Maestro, mientras que el proceso de preparación de los ligandos se realizó a través de *LigPrep* (LigPrep, Schrödinger, LLC, New York, NY, 2021). Para la preparación de la proteína se utilizó el componente *Protein Preparation Wizard* (Protein Preparation Wizard, Schrodinger LLC, New York, NY, USA, 2016) que ofrece la misma suite. El docking molecular de los ligandos se efectuó con *Glide* ((Glide, Schrödinger, LLC, New York, NY, 2021)), con detalles explicados en la [sección 4.2](#). Los análisis sobre la calidad de las poses obtenidas para cada compuesto frente a la proteína se desarrollaron utilizando el servidor LigRMSD (Velázquez-Libera et al., 2020), mientras que la identificación de interacciones recurrentes entre los residuos presentes en el sitio de unión de PLpro y sus ligandos, se realizó por medio del panel de interacción de huellas dactilares (IFP). Finalmente, *Phase* (Phase, Schrodinger LLC, Nueva York, NY, EE. UU., 2016) fue usado para el tratamiento de los modelos 3D-QSAR.

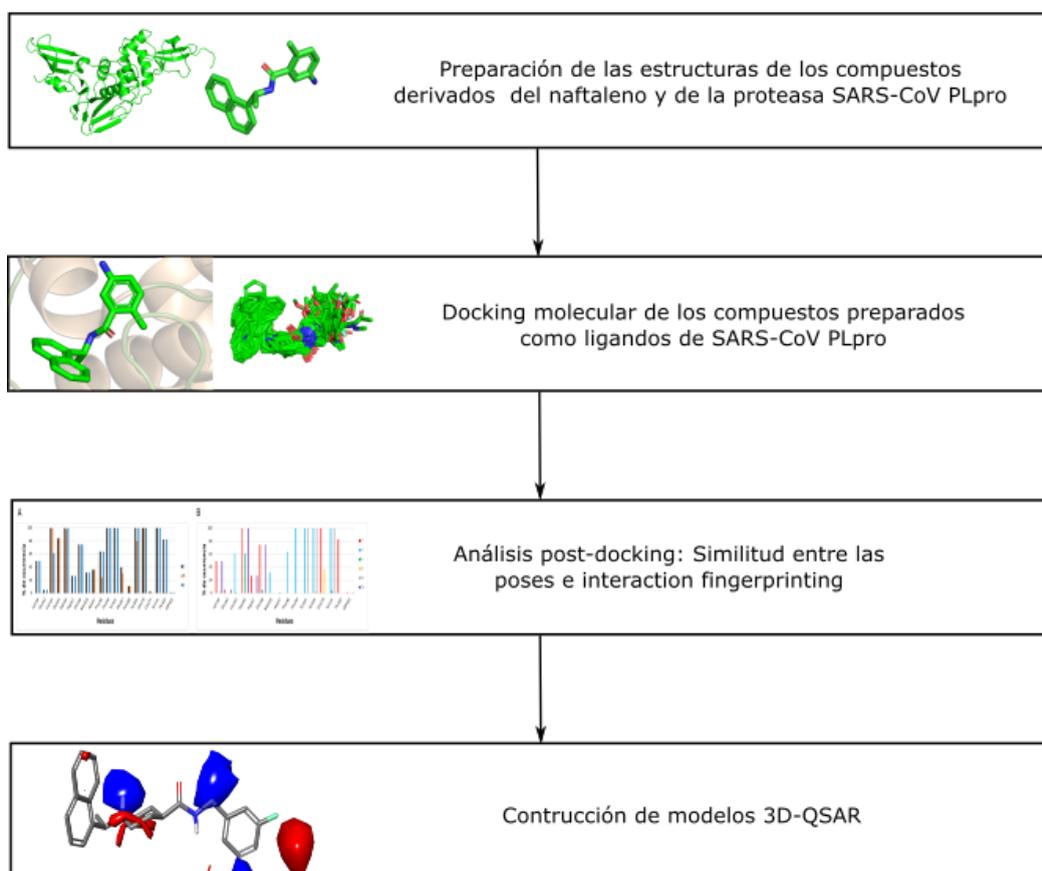


Figura 4.1: Metodología. Flujo de trabajo para la caracterización estructural de los compuestos derivados del naftaleno interactuando con SARS-CoV PLpro.

4.1 Preparación de compuestos derivados del naftaleno y estructura de la proteína.

Para realizar el estudio de interacción proteína-ligando, se recopilaron 75 estructuras de los compuestos derivados del naftaleno procedentes de las series reportadas por Ratia et al. (Ratia et al., 2008; Gosh et al., 2009, 2010; Baez-Santos et al., 2014). Las estructuras de los compuestos que se estudiaron se encuentran en las [tablas 4.1-4.4](#), las cuales fueron dibujadas utilizando el editor molecular de *Maestro* (Maestro, Schrodinger, LLC, Nueva York, NY, EE. UU.). El pre-procesamiento de los compuestos se realizó utilizando la herramienta

LigPrep (LigPrep, Schrödinger, LLC, New York, NY, 2021). De esta forma, se convirtieron todos los compuestos dibujados en dos dimensiones a su forma tridimensional (3D), siendo estos tratados posteriormente mediante una serie de pasos enfocados en la preparación del ligando. El estado de protonación de cada uno de los inhibidores fue estimado utilizando *Epik* (Epik, Schrödinger, LLC, New York, NY, 2021) bajo un valor de pH fisiológico de 7. Para todas las estructuras generadas, se consideraron aquellas que se encontraban adecuadamente protonadas. Esto se definió teniendo en cuenta los diferentes complejos proteína-ligando presentes en PDB. Como último paso de preparación, se inspeccionaron visualmente las representaciones moleculares.

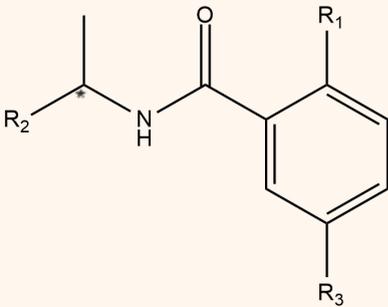
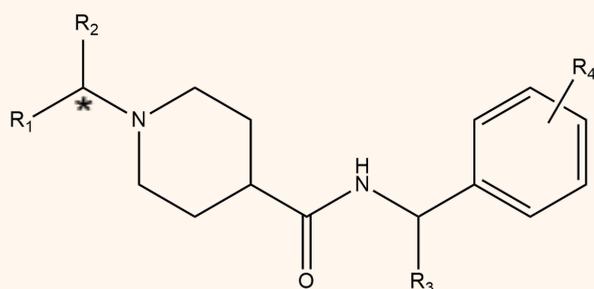
				
Compuesto	R ₁	R ₂	R ₃	IC ₅₀ (μM)
7724772 (S)	Me	2-naphthyl	H	>200
7724772 (R)	Me	2-naphthyl	H	8.7 ± 0.7
A3 (R)	Cl	2-naphthyl	H	14.5 ± 0.9
A4 (R)	Et	2-naphthyl	H	>200
A5 (R)	Me	1-naphthyl	H	2.3 ± 0.1
A6 (R)	Me	1-naphthyl	NHAc	2.6 ± 0.1
A7 (R)	Me	1-naphthyl	NO ₂	7.3 ± 0.9
GRL0617	Me	1-naphthyl	NH ₂	0.6 ± 0.1

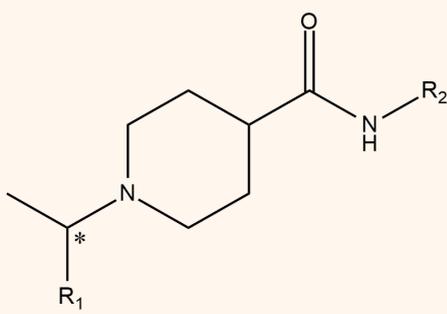
Tabla 4.1: Primer conjunto de compuestos derivados del naftaleno seleccionados para el presente estudio.



Compuesto	R ₁	R ₂	R ₃	R ₄	IC ₅₀ (μM)
65778771	1-naphthyl	H	H	4-OMe	59.2 ± 7.8
B7a	1-naphthyl	H	H	2-OMe	116 ± 30
B7b	1-naphthyl	H	H	3-OMe	30 ± 3
B15a	1-naphthyl	Me(R)	H	2-OMe	1.21 ± 0.04
B15b	1-naphthyl	Me(R)	H	3-OMe	0.34 ± 0.01
B15c	1-naphthyl	Me(R)	H	4-OMe	0.34 ± 0.01
B15d	2-naphthyl	Me(R)	H	3-OMe	13.2 ± 0.6
B15e	2-naphthyl	Me(R)	H	2-OMe	34.8 ± 4.0
B15f	2-naphthyl	Me(S)	H	3-OMe	5.8 ± 0.1
B15g	1-naphthyl	Me(R)	H	3,4-(-O-CH ₂ -O-)	0.32 ± 0.01
B15h	1-naphthyl	Me(S)	H	3,4-(-O-CH ₂ -O-)	0.56 ± 0.03
B15i	1-naphthyl	H	H	3,4-(-O-CH ₂ -O-)	~45
B15j	2-naphthyl	H	H	3,4-(-O-CH ₂ -O-)	~100
B15k	1-naphthyl	Gem-dimethyl	H	3,4-(-O-CH ₂ -O-)	>200
C1a	1-naphthyl	Et(R, S)	H	3,4-(-O-CH ₂ -O-)	17.2 ± 0.03
C1b	1-naphthyl	CH ₂ OH(R, S)	H	3,4-(-O-CH ₂ -O-)	32.0 ± 4.5
C1c	1-naphthyl	CH ₂ OMe(R, S)	H	3,4-(-O-CH ₂ -O-)	>100
C1d	1-naphthyl	CH ₂ Ph(R, S)	H	3,4-(-O-CH ₂ -O-)	>100
C2a	1-naphthyl	Me(R)	H	H	2.2 ± 0.1
C2b	1-naphthyl	Me(R)	Me(R)	H	13.5 ± 1.2
C2c	1-naphthyl	Me(R)	Me(S)-	H	12.7 ± 0.3

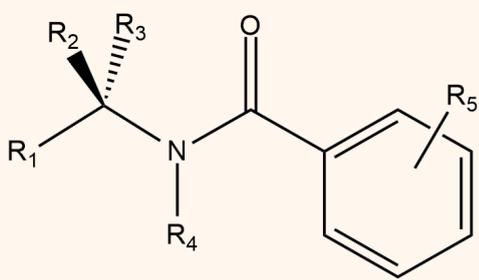
C2d	1-naphthyl	Me(R)	CH ₂ OMe(R)	H	18.0 ± 1.9
C2e	1-naphthyl	Me(R)	CH ₂ OMe(S)	H	1.9 ± 0.1
C3a	1-naphthyl	Me(R)	H	4-Et	0.47 ± 0.01
C3b	1-naphthyl	Me(R)	H	4-CO-NH-Me	0.60 ± 0.02
C3c	1-naphthyl	Me(R)	H	3-CO-NH-Me	0.63 ± 0.01
C3d	1-naphthyl	Me(R)	H	4-NH-CO-Me	5.7 ± 0.5
C3e	1-naphthyl	Me(R)	H	3-NH-CO-Me	0.39 ± 0.01
C3f	1-naphthyl	Me(R)	H	3-CH ₂ -NH-CO-Me	20.4 ± 1.2
C3g	1-naphthyl	Me(R)	H	3-Cl	27.2 ± 4.1
C3h	1-naphthyl	Me(R)	H	4-Cl	0.58 ± 0.02
C3i	1-naphthyl	Me(R)	H	3,4-diF	29.2 ± 2.1
C3j	1-naphthyl	Me(R)	H	4-F	0.49 ± 0.01
C3k	1-naphthyl	Me(R)	H	3-F	0.15 ± 0.01

Tabla 4.2: Segundo conjunto de compuestos derivados del naftaleno seleccionados para el presente estudio.

			
Compuesto	R ₁	R ₂	IC ₅₀ (μM)
C4a	8-quinolinyl(R, S)	3-F-Ph-CH ₂	7.0 ± 0.7
C4b	5-quinolinyl(R, S)	3-F-Ph-CH ₂	4.5 ± 0.2
C4c	5-isoquinolinyl(R, S)	3-F-Ph-CH ₂	6.8 ± 0.3
C4d	1-isoquinolinyl(R, S)	3-F-Ph-CH ₂	30.8 ± 2.6
C5a	1-naphthyl(R)	3-pyridinyl-CH ₂	26.3 ± 2.3

C5b	1-naphthyl(R)	4-pyridinyl-CH ₂	18.3 ± 0.9
C5c	1-naphthyl(R)	2-methoxy-4-pyridinyl-C H ₂	0.35 ± 0.02
C6a	1-naphthyl(R)	4-Cl-Ph-CH ₂ CH ₂	1.6 ± 0.3
C6b	1-naphthyl(R)	3-F-Ph-CH ₂ CH ₂	1.9 ± 0.1

Tabla 4.3: Tercer conjunto de compuestos derivados del naftaleno seleccionados para el presente estudio.

						
Compuesto	R ₁	R ₂	R ₃	R ₄	R ₅	IC ₅₀ (μM)
D5a	2-naphthyl	Me	H	H	3-Me	14.8 ± 5.0
D5b	2-naphthyl	Me	H	H	4-Me	29.1 ± 3.8
D5c	2-naphthyl	Me	H	H	2-OMe	90 ± 26
D5d	2-naphthyl	Me	H	H	3-OMe	13.5 ± 6.8
D5e	2-naphthyl	Me	H	H	4-OMe	149 ± 43
D5f	2-naphthyl	Me	H	H	2,6-diMe	12.1 ± 0.7
D9	2-naphthyl	Me	H	H	4-NH ₂	46.1 ± 13.0
D21	1-naphthyl	Me	H	Me	2-Me	22.6 ± 6.9
D23	1-naphthyl	Me	H	H	4-NH ₂	24.8 ± 1.0
D29	1-naphthyl	Me	Me	H	2-Me and 5-NH ₂	11.1 ± 1.3
D33	1-naphthyl	Me	H	H	2-Me and 5-CN	5.2 ± 0.5
D40	1-naphthyl	Me	H	H	2-CH ₂ OMe and 5-NH ₂	2.7 ± 0.1

D32	1-naphthyl	Me	H	H	2-Me and 5-I	1.4 ± 0.3
D47	1-naphthyl	Me	H	H	2-Me and 5-CH ₂ NHBoc	4.8 ± 0.4
D49	1-naphthyl	Me	H	H	2-Me and 5-CH ₂ NHMe	1.3 ± 0.1
D02	1-naphthyl	Me	H	H	2-Me and 5-CH ₂ NH ₂	0.46 ± 0.03

Tabla 4.4: Cuarto conjunto de compuestos derivados del naftaleno seleccionados para el presente estudio.

Además de la preparación de los ligandos, el procesamiento de la proteína es un paso importante para asegurar que las estructuras tridimensionales estén correctamente optimizadas. En la base de datos *Protein Data Bank* (PDB), existen estructuras tridimensionales (3D) de SARS-CoV PLpro formando complejos con inhibidores no covalentes derivados del naftaleno, localizados en los subsitios S3-S4 de la proteasa. En base a estas estructuras, se seleccionaron aquellas con identificadores 3MJ5, 3E9S, 4OVZ y 4OW0, considerando la diversidad de conformaciones presentes en el sitio de unión de PLpro. La optimización de las estructuras proteicas se realizó a través del módulo *Protein Preparation Wizard* de Maestro (Protein Preparation Wizard, Schrodinger LLC, New York, NY, USA, 2016), incluyendo una serie de pasos para la optimización. El primer paso, corresponde a un preprocesamiento de la macromolécula en estudio, en donde para cada uno de los casos se emplearon los parámetros por defecto otorgados por la herramienta. De esta forma se asignaron átomos faltantes y se corrigieron las orientaciones de los aminoácidos que componen la proteína. Dado que las estructuras extraídas desde PDB carecen de átomos de hidrógeno, se realizó la adición de estos, con la finalidad de tener todos los átomos representados y posicionados explícitamente. Sumado a lo anterior, se asignaron órdenes de enlace utilizando la base de datos de compuestos químicos (CCD), así como también se llevó a cabo el refinamiento de cadenas laterales y bucles faltantes utilizando *Prime* (Prime, Schrödinger, LLC, New York, NY, 2021). Teniendo en cuenta la variedad de compuestos presentes en la serie congénica, se decidió remover todas las moléculas de agua presentes en los cristales. Una vez que el preprocesamiento finalizó, se eliminaron los grupos con la etiqueta 'HET' correspondientes a iones de Cl, Na, Zn, moléculas de disolvente y grupos que no forman parte de la proteína. Posteriormente, las estructuras proteicas se sometieron a

una etapa de refinamiento, donde se realizó una optimización de la red de enlaces de hidrógeno mediante la reorientación de los grupos hidroxilo y tiol, las moléculas de agua, los grupos amida de la asparagina (Asn), la glutamina (Gln), y el anillo de imidazol de la histidina (His). Además, se efectuó una predicción de los estados de protonación de grupos ionizables de ácido aspártico (Asp), ácido glutámico (Glu) e histidina (His), en conjunto con sus estados tautoméricos. Para finalizar, se minimizaron las estructuras utilizando el campo de fuerza OPLS3.

Una vez finalizado el protocolo de preparación de proteínas para cada uno de los cristales y dada la diversidad conformacional del sitio de unión de los derivados del naftaleno, se realizó un análisis de las estructuras con la finalidad de conocer con más detalle el sitio preferencial de unión de este tipo de compuestos. De esta forma, se obtuvo información relevante, para así emplear un análisis más exhaustivo en torno a un menor número de estructuras. La proteína correspondiente a 3E9S, presenta un estado monomérico, mientras que las entradas con los códigos PDB 3MJ5, 4OVZ y 4OW0 corresponden a dímeros con ligandos de la serie congénica co-cristalizados. Primeramente, las estructuras se alinearon en relación a la cadena A del monómero utilizando el plugin “*Protein Structure Alignment*” de Maestro (Maestro, Schrödinger, LLC, New York, NY, 2021.), resultando valores de RMSD de 0.844 Å, 1.008 Å y 1.045 Å para 3MJ5, 4OVZ y 4OW0, respectivamente. Una vez alineadas las estructuras, se llevó a cabo una comparación por medio del RMSD (Å) de los residuos del sitio de unión de los cristales mencionados, considerando los aminoácidos que se encuentran a una distancia corte de 5 Å desde los átomos del ligando. Los resultados obtenidos a partir del script “rmsd_per_residue” de Schrödinger (Kevin et al., 2006) se exponen en la [tabla 4.5](#).

Residuo	3MJ5	4OVZ	4OW0
Tyr113	0.22	0.29	0.34
Lys158	1.23	1.27	1.27
Leu163	1.41	1.36	1.44
Val166	0.27	0.22	0.23
Asp165	0.45	0.41	0.29
Arg167	1.20	1.33	1.16

Glu168	1.38	1.68	1.77
Met209	1.52	1.28	0.86
Ala247	0.58	0.81	0.73
Pro248	0.99	1.30	1.27
Pro249	1.13	1.12	1.15
Tyr265	0.60	0.33	0.40
Gly267	0.52	0.80	0.83
Asn268	1.68	1.69	1.62
Tyr269	1.04	1.19	1.04
Gln270	5.53	5.06	5.15
Cys271	2.41	-	2.48
Gly272	0.92	0.82	0.79
Tyr274	0.26	0.39	0.38
Thr302	1.97	1.97	1.94

Tabla 4.5: Valores de RMSD (Å) de los residuos que se encuentran a un radio de 5Å de los átomos del ligando. El cristal 3E9S se utilizó como referencia. Residuos causantes de interacciones con los ligandos co-cristalizados se destacan en negrita. El residuo Cys271 corresponde a CSO271 (S-hidroxicisteína) en el cristal 4OVZ, razón por la cual no se obtuvo el valor comparativo.

De un total de 20 residuos comprometidos en el sitio de unión de la proteína SARS-CoV PLpro, el 90% presenta valores de RMSD inferiores a 2.0 Å. Entre estos, se encuentran los residuos Asp165, Tyr265 y Tyr269, esenciales para la formación del complejo proteína-ligando. Sin embargo, el residuo Gln270, de igual importancia, existe en conformaciones distintas al cristal de referencia 3E9S, presentando un RMSD > 5.0 Å en 3MJ5, 4OVZ y 4OW0.

Sumado a esto, los datos presentes en la [tabla 4.5](#) para las estructuras diméricas revelan una similitud entre los residuos del sitio de unión de estas. Los valores para la Glutamina en la posición 270 son de 5.53 Å (3MJ5), 5.06 Å (4OVZ) y 5.15 Å (4OW0) con respecto a su posición en la referencia. Este mismo residuo se comparó entre las estructuras diméricas, utilizando como referencia el cristal 3MJ5, obteniéndose valores de RMSD de 1.58 Å (4OVZ) y 1.53 Å (4OW0). De esta forma se pudo observar que existe una diferencia no importante en la posición de Gln270. Asimismo, se verificó visualmente que el resto de los residuos involucrados en el sitio de unión presentan una misma orientación entre las estructuras en un mismo estado. Esta información se utilizó como material para el desarrollo de los análisis

posteriores, los cuales se emplearon en torno a la estructura cristalográfica monomérica 3E9S y al dímero 4OW0, con una menor resolución (2.10 Å) que 3MJ5 (2.63 Å) y 4OVZ (2.50 Å).

4.2 Docking Molecular

Los estudios de acoplamiento molecular se realizaron utilizando *Glide* de la suite de Schrödinger (*Glide*, Schrödinger, LLC, New York, NY, 2021), permitiendo crear un modelo de interacción proteína-ligando para la serie de inhibidores de naftaleno. *Glide* diseña una serie de filtros jerárquicos para buscar las posibles poses y orientaciones del ligando dentro del sitio de unión del receptor, utilizando una grilla definida previamente. Los filtros implican un enfoque de búsqueda sistemática que muestrea el espacio posicional, conformacional y de orientación del ligando antes de evaluar las interacciones de energía entre el ligando y la proteína (Friesner et al., 2004).

Glide ofrece tres filtros jerárquicos: cribado virtual de alto rendimiento (HTVS), precisión estándar (SP) y precisión extra (XP). Para este caso en particular, los compuestos se sometieron a los módulos SP y XP. *Glide* SP, es una función más indulgente y permite la identificación de ligandos que tienen una tendencia razonable a unirse. Por otra parte, el módulo de precisión extra (XP), es una función más estricta, la cual penaliza a las poses que violan los principios de la fisicoquímica (Friesner et al., 2004). La utilización de estos dos módulos de manera conjunta ayudó a acceder a una mayor cantidad de soluciones con buena calidad. *Glide* SP permitió evaluar la capacidad del protocolo para encontrar poses que posean interacciones con los residuos presentes en las estructuras cristalográficas reportadas; mientras que la función menos indulgente, XP, se utilizó para obtener las poses finales de docking, con las cuales se dio inicio al análisis.

La mejor pose de cada ligando fue escogida por medio de dos criterios. El primero de estos corresponde a la puntuación otorgada por *Glide*; mientras mayor sea la puntuación del módulo, mejor será la afinidad calculada del hit en la unión a la proteína objetivo. Sumado a lo anterior, se consideraron aquellas poses que tuvieron interacciones con residuos claves reportados en los cristales de PLpro. Mediante un análisis previo sobre las estructuras de SARS-CoV PLpro en complejo con moléculas derivadas del naftaleno (PDB ID: 3MJ5, 3E9S,

4OVZ y 4OW0), se determinaron los residuos D165, Y265, Y269 y Q270 como importantes para la formación de enlaces de hidrógeno, *pi-pi stacking*, puente salido y pi-cación (Figura 4.2). Tres de estos residuos (Y269, Y265 y D165) forman interacciones que están presentes en las entradas PDB 3MJ5, 4OVZ y 4OW0 (figura 4.2B). Las diferencias estructurales del compuesto presente en 3E9S conllevan a una nueva interacción de tipo enlace de hidrógeno con el residuo Gln270 (figura 4.2A).

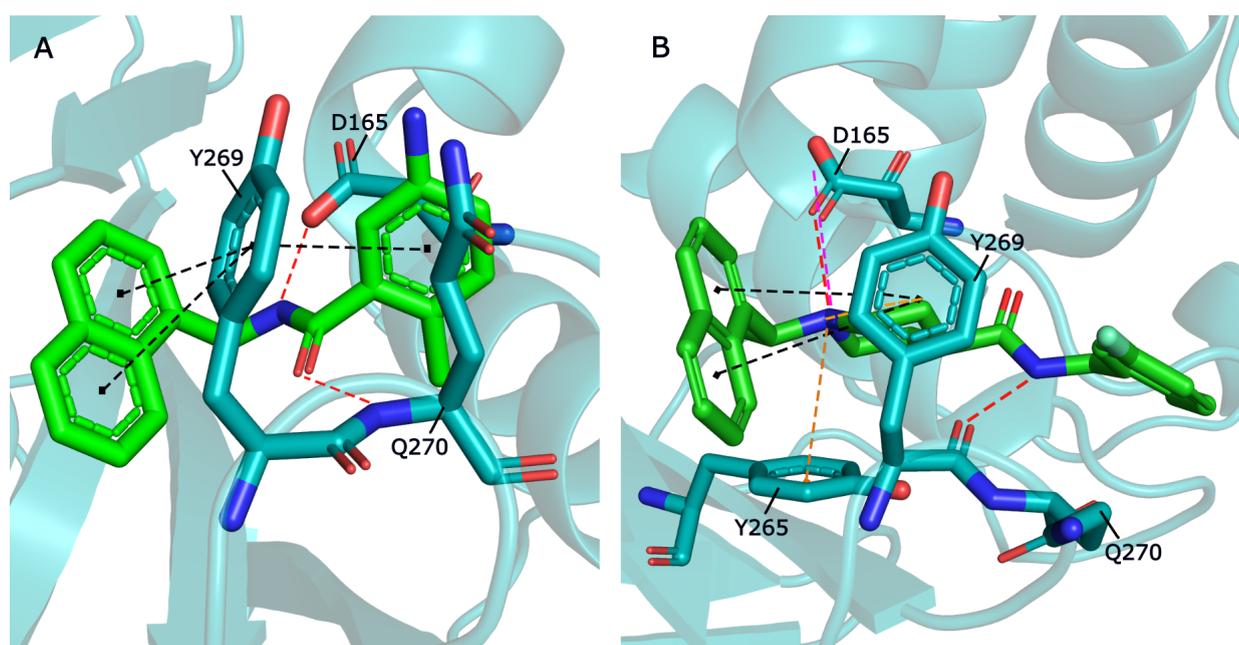


Figura 4.2: Residuos participantes en la interacción de los compuestos derivados del naftaleno en los cristales 3E9S, 4OW0, 4OVZ, 3MJ5. (A) Interacciones en el cristal 3E9S. (B) Interacciones presentes en los cristales 3MJ5, 4OVZ y 4OW0. Las interacciones se indican mediante líneas discontinuas y diferentes coloraciones. Enlaces del tipo *pi-pi stacking* se colorean negro, mientras que interacciones de tipo enlace de hidrógeno, pi-cación y puente salido se rotulan de rojo, naranja y magenta, respectivamente.

Recientemente, Pitsillou y col (2020) realizaron estudios basados en compuestos derivados del naftaleno, logrando destacar por medio de docking molecular una alta afinidad del compuesto GRL0617 por la proteína SARS-CoV PLpro. A través de este estudio se identificaron residuos claves en la formación del complejo receptor-ligando (Pitsillou et al, 2020). Basándose en los protocolos de este trabajo, el conjunto de ligandos utilizado se acopló a la proteína mediante una grilla de 20 Å x 20 Å x 20 Å centrada en los subsitios S3 y

S4 de PLpro. Se utilizaron parámetros de docking predeterminados, escalando los radios de Van der Waals en un factor de 0,80 y considerando como átomos no polares a aquellos con carga atómica parcial $\leq 0,15$. Se especificó además un docking con ligando flexible y se aplicaron limitaciones de posición asociadas al grupo naftaleno de los compuestos. La minimización del ligando en el sitio del receptor se realizó utilizando la función de energía de mecánica molecular estándar, que incorpora el campo de fuerza OPLS junto con un dieléctrico dependiente de la distancia (Friesner et al., 2004). Posteriormente, los ligandos con menor energía se refinaron mediante un muestreo de Monte Carlo, lo que es crucial para obtener una pose de docking precisa. La selección de la mejor pose se determinó por medio de la puntuación *Emodel*. Por otra parte, se utilizó *GlideScore* para clasificar los compuestos en base a su unión al receptor.

4.3 Análisis post-docking

4.3.1 LigRMSD

Al efectuar el docking de compuestos congénéricos, se esperaría obtener poses similares de estos, y a la vez similares a las de un compuesto análogo co-cristalizado en las estructuras de PLpro seleccionadas para el desarrollo de este estudio. La búsqueda de similitudes entre las poses de docking y las poses de los compuestos en los cristales fueron útiles para demostrar la calidad de los resultados obtenidos, y se realizó utilizando el servidor LigRMSD (Velázquez-Libera et al., 2020). LigRMSD permite seleccionar los átomos comunes entre las estructuras que se comparan, estableciendo grafos coincidentes entre ellas, y calcula la desviación cuadrática media (RMSD) entre los átomos equivalentes con respecto a una referencia. Bajo el mismo contexto, la coincidencia se define utilizando los valores “%Ref” y “%Mol”. “%Ref” señala el porcentaje de grafos comunes entre un compuesto acoplado y una referencia seleccionada, en relación al número total de átomos de la referencia seleccionada. Por otra parte, “%Mol” es el porcentaje de grafos comunes entre el compuesto acoplado y la referencia seleccionada, en relación con el número total de átomos del compuesto acoplado. Estos valores entregados por el servidor LigRMSD representan la similitud máxima entre los compuestos que se comparan, de manera que altos valores de %Ref y %Mol se asocia con una gran semejanza entre los compuestos estudiados.

En base a esto, primeramente, se realizó una comparación entre las poses de los inhibidores de SARS-CoV PLpro derivados del naftaleno con identificadores 3MJ5, 3E9S, 4OVZ y 4OW0, con la finalidad de conocer el comportamiento y semejanza de los ligandos cristalizados. Posteriormente, se compararon las poses obtenidas utilizando múltiples referencias. Para compuestos acoplados al cristal monomérico 3E9S, se utilizó la pose del ligando co-cristalizado y la proveniente del inhibidor 6577871, mientras que para la estructura dimérica de SARS-CoV PLpro, se utilizó la pose del compuesto acoplado en el cristal 4OW0 y la del ligando 7724772.

4.3.2 Análisis de interacción molecular

Para hacer una descripción detallada a nivel global de los contactos e interacciones químicas entre el receptor (SARS-CoV PLpro) y los ligandos acoplados, se realizó un análisis de interacción molecular (*interaction fingerprints*, IFP) para las poses obtenidas a través de docking. Para construir los IFP se utilizó el panel de *Interaction Fingerprints* de Maestro (Maestro, Schrodinger LLC, New York, NY, USA, 2016). Este método describe la presencia o ausencia de interacciones químicas entre los ligandos y los residuos del sitio de unión del receptor mediante el uso de bits, para una posterior construcción de una matriz de interacción. Cada bit describe si tiene lugar un tipo de interacción específico entre el ligando y un residuo de la proteína, teniendo en cuenta interacciones hidrófobas (H), polares (P) y Aromáticas (Ar). Es posible detectar también si un residuo proporciona aceptor(es) (A) o donador(es) (D) de enlace de hidrógeno e interacciones electrostáticas con grupos cargados (Ch). Para este estudio, se contabilizó como interacción, cuando un residuo de la proteína en estudio se encuentre dentro de una distancia de corte máxima de 4.0 Å entre átomos pesados, con respecto a los átomos del ligando (Velázquez-Libera et al., 2019).

4.4 3D-QSAR

La relación cuantitativa estructura-actividad (QSAR) es un enfoque basado en ligandos que implica la cuantificación de la correlación entre los rasgos estructurales de una serie de compuestos y la respuesta biológica/toxicológica subyacente (Wang et al., 2020).

Esencialmente, el método QSAR se basa en la hipótesis de que propiedades estructurales o fisicoquímicas similares producen actividades similares. Los enfoques QSAR han evolucionado desde el QSAR clásico (1D y 2D) hasta el QSAR multidimensional (3D-6D), prefiriéndose la metodología 3D-QSAR sobre los otros paradigmas. El objetivo principal de este, es establecer la relación entre la actividad biológica y las propiedades espaciales de los compuestos químicos (estéricas, electrostáticas y lipofílicas). El enfoque 3D-QSAR se aplica principalmente a series de compuestos para descubrir características moleculares farmacóforas de los mismos, llevando a cabo el análisis y la optimización de aquellos compuestos que pueden aumentar su actividad biológica (Verma et al., 2010).

Basado en lo mencionado anteriormente, se realizaron estudios de relación cuantitativa estructura-actividad (QSAR) con la finalidad de correlacionar características estructurales de los compuestos estudiados en su conformación activa con sus actividades inhibitorias experimentales. Las estructuras del conjunto de datos obtenidas del acoplamiento molecular fueron alineadas considerando su orientación en el sitio activo de la proteína. Se utilizó el módulo “*Field-Based QSAR*” de *Phase* (Phase, Schrodinger LLC, Nueva York, NY, EE. UU., 2016) para generar modelos 3D-QSAR, empleando el modelo de regresión de factores por mínimos cuadrados (PLS) para establecer la correlación. Para originar el modelo QSAR, los datos se dividieron en un conjunto de entrenamiento (training) y uno de validación externa (test) con una división de un 80 y 20% de los datos, respectivamente. El modelo se realizó usando solamente el conjunto de entrenamiento. Se especificó un espaciado de malla de la grilla por defecto, correspondiente a 1Å. El programa *Phase* proporciona un conjunto estándar de seis descriptores de campo asociados a las estructuras de los compuestos en estudio: aceptores de enlaces de hidrógeno (A), donantes de enlaces de hidrógeno (D), grupos hidrófobos (H), grupos ionizables negativamente (N), ionizables positivamente (P) y anillos aromáticos (R), pudiendo estas características ser observadas por medio de diferentes colores. (Dixon et al., 2006). Durante el procedimiento PLS, se eliminaron las variables con una desviación estándar de menos de 0.01. Por otra parte, las variables cuyos coeficientes de regresión eran sensibles a pequeños cambios en la composición del conjunto de entrenamiento se eliminaron del modelo QSAR, utilizando un $|t\text{-value}| < 2.00$. El número de componentes principales máximo a analizar en el PLS se

estableció en 6 y se exploraron distintas combinaciones de descriptores de campo para seleccionar la combinación que mejor describe los datos bajo estudio.

Los modelos desarrollados en el conjunto de entrenamiento se validaron con un protocolo de validación cruzada *leave-one-out* mediante el cálculo del valor Q^2 (referencia). Una vez seleccionado el mejor modelo (el de mejor Q^2), se hizo una predicción de los compuestos colocados en el conjunto de validación y se verificó el valor de R^2 del test (R^2_{test}). Un buen modelo QSAR debe cumplir con dos condiciones de validación interna y externa: un valor de $Q^2 > 0.5$ y $R^2_{\text{test}} > 0.6$.

5. RESULTADOS

5.1 Poses de acoplamiento molecular

Se tiene conocimiento que los ligandos derivados del naftaleno comparten una serie de características estructurales, como la presencia de un grupo aromático adyacente a la carboxamida y el grupo naftilo característico de este tipo de compuestos, enlazado a un centro estereogénico de la molécula (Ghosh et al., 2006). Los enlaces intermoleculares formados entre estos compuestos y los residuos de la proteína varían según la estructura cristalina. Para los cristales utilizados en este estudio, los ligandos de esta serie forman interacciones no covalentes entre sus anillos aromáticos y la cadena lateral del residuo Tyr269. Además, la presencia de un anillo de piperidina en los compuestos derivados de 6577871 (Figura 1.3C) favorece la formación de interacciones electrostáticas entre el átomo de nitrógeno cargado en el ligando y residuos como Tyr265, Tyr269 y Asp165. El residuo de Gln270 forma un enlace de hidrógeno con el grupo carbonilo del ligando cocrystalizado en el monómero 3E9S. Esta interacción no tiene lugar entre los compuestos estudiados y los dímeros 3MJ5, 4OVZ, 4OW0. Sumado a esto, existen residuos hidrófobos, tales como Pro249, Tyr274, Pro248, Leu163 y Met209 cercanos a los grupos hidrófobos de los ligandos. En base a esto, se esperaba que los compuestos en estudio tuvieran orientaciones parecidas a las observadas en PDB y establecieran interacciones químicas semejantes a las mencionadas.

Para estudiar la base molecular de las interacciones entre la serie congénica de compuestos derivados del naftaleno y la proteasa PLpro de SARS-CoV y analizar la afinidad de esta unión, se acoplaron 75 ligandos a los residuos que componen los subsitios S3 y S4 de la proteína, utilizando la herramienta *Glide* de la suite de Schrödinger (*Glide*, Schrödinger, LLC, New York, NY, 2021). Los métodos de precisión estándar (SP) y extraprecisión (XP) se emplearon con los mismos parámetros para los dos cristales utilizados con identificadores 3E9S y 4OW0. *Glide* SP permitió evaluar la capacidad del protocolo para encontrar poses que establezcan interacciones con los residuos presentes en las estructuras cristalográficas

reportadas, tales como: Asp165, Gln270, Tyr269, Tyr265. En adición a esto, se utilizó la función Glide XP, la cual es menos indulgente, para obtener las poses finales de docking.

A partir de los resultados del acoplamiento se logra apreciar que todos los ligandos de la serie adoptan el mismo modo de unión, orientando los restos de naftilmetilamina en los subsitios S3 y S4 de la enzima, compuestos por los residuos de L163, Y265 y D165, P249, T302, respectivamente (Figura 5.1). Ratia y col. revelaron la preferencia de estos inhibidores competitivos por el bolsillo adyacente al sitio catalítico de la proteasa a través de cristalografía de rayos X del complejo PLpro:GRL0617. Estos subsitios son críticos para el reconocimiento del residuo de leucina en el motivo LXGG, el cual se encaja en el bolsillo hidrófobo formado por Pro249, Tyr265 y Thr302. Por esta razón, los ligandos unidos a S3 y S4, se catalogan como buenos candidatos para inhibidores de la enzima (Ratia et al., 2006; Osipiuk, 2021).

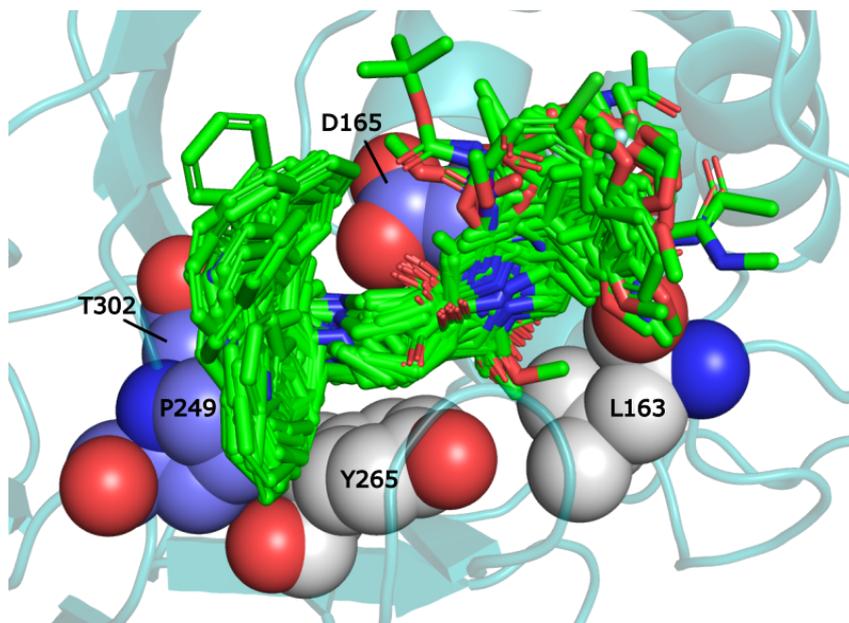


Figura 5.1: Estructuras acopladas dentro del sitio de unión de SARS-CoV PLpro. Los ligandos acoplados se representan por medio de barras verdes. Los residuos del subsitio S3 de la proteasa, se representan por medio de esferas blancas. Los residuos que componen el subsitio S4 de PLpro, se encuentran representados por esferas violetas.

De manera general, se aprecia que los compuestos originados desde 7724772 en el monómero 3E9S, presentan un enlace de hidrógeno entre el carbonilo adyacente al anillo aromático del inhibidor y el residuo Gln270 (Figura 5.2A). Se reporta en la literatura que este enlace, así como otro adicional formado con el residuo Asp165, son de gran importancia para la estabilización de ligandos cortos en el sitio de unión descrito anteriormente (Ratia et al., 2008). Con relación al mismo cristal, la interacción establecida con Gln270 se encontró considerablemente en menor proporción para los ligandos que presentan restos de piperidina. Los aminoácidos con cadenas laterales aromáticas, tales como Tyr269 y Tyr265 se encontraron involucrados en la formación de interacciones *pi-pi stacking* con los grupos aromáticos del ligando; además, se observaron interacciones *cación-pi* con el nitrógeno cargado del anillo de piperidina presente en derivados de 6577871. Esta porción protonada del ligando formó una interacción de puente salino con el carboxilato aniónico del residuo de Asp165, el cual a su vez, se encuentra a una distancia de enlace de hidrógeno con respecto a la misma región del ligando. Para compuestos derivados de 7724772, Asp165 formó un enlace de hidrógeno con el NH de la carboxamida de los inhibidores. Por otra parte, la presencia del aminoácido con carga positiva Lys158 en el sitio de unión de la serie congénica, contribuyó a la formación de interacciones del tipo *cación-pi* con el grupo aromático y sus sustituyentes en compuestos provenientes de 6577871 (Figura 5.2B).

Sumado a esto, se observó que el bolsillo revestido por los residuos Pro248, Pro249, Met209, Thr302, Tyr269 y Tyr265 estuvo en proximidad al grupo naftilo, lo cual contribuye a la formación de interacciones hidrófobas. Así mismo, el residuo Leu163 también contribuye a este tipo de interacciones, dada su cercanía con el grupo benceno próximo al resto de carboxamida en ligandos más grandes. Tanto para compuestos en los cuales se presenta el anillo de piperidina, como para aquellos que carecen de este, el naftaleno se colocó de forma estable entre los residuos de Pro248 y Pro249 (Figura 5.2A-D).

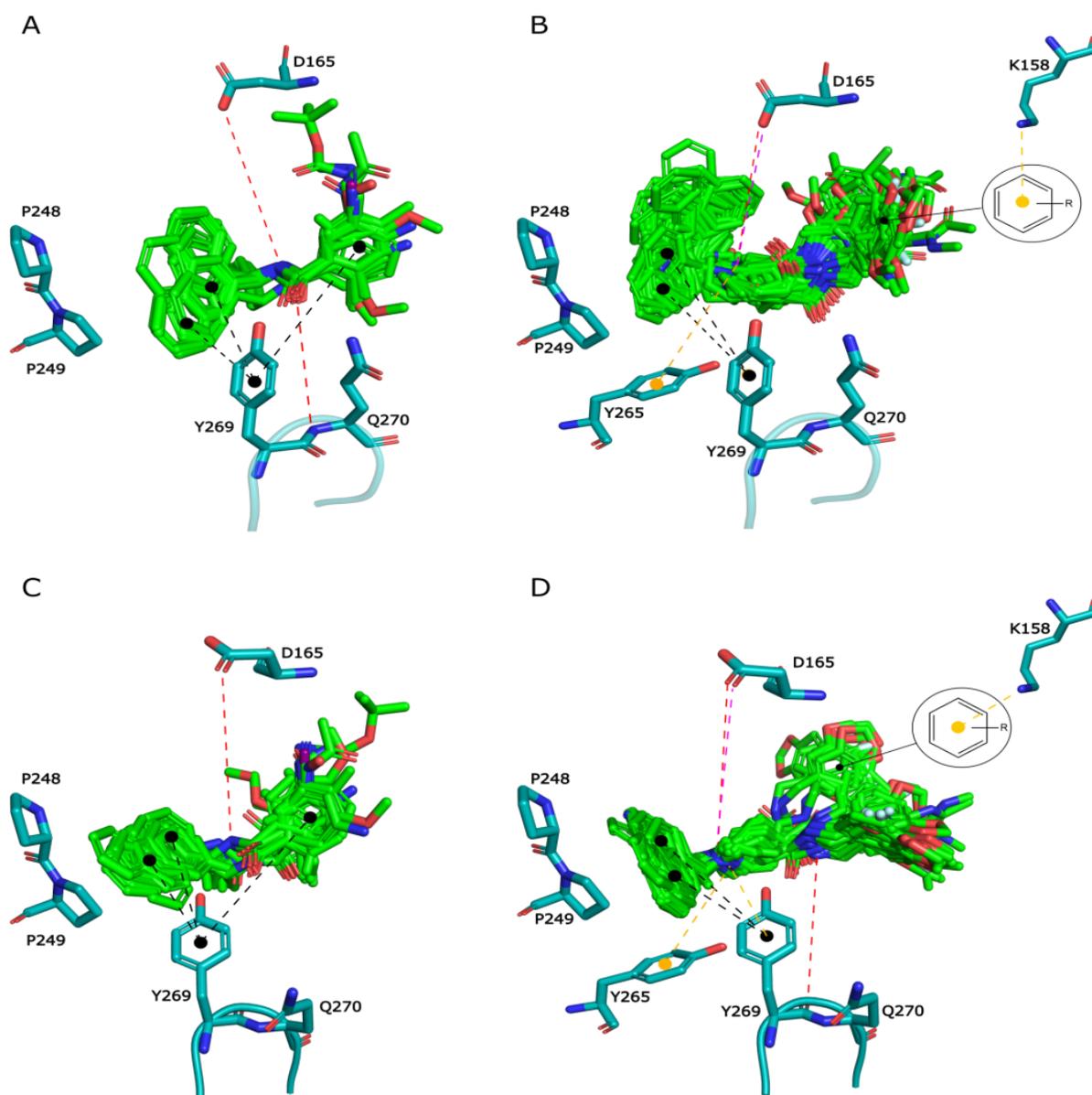


Figura 5.2: Serie de compuestos derivados del naftaleno en el sitio de unión de SARS-CoV PLpro. (A) Compuestos de la serie derivados del ligando 7724772, acopladas al cristal monomérico 3E9S. (B) Compuestos de la serie derivados del ligando 6577871, acoplados al cristal 3E9S. (C) Compuestos de la serie derivados del compuesto 7724772, en el sitio de unión del cristal 4OW0 (dímero). (D) Compuestos derivados de 6577871, acoplados en el sitio de unión de 4OW0. Los ligandos se representan por medio de barras de color verde, en tanto que los residuos involucrados en el sitio de unión se colorearon de azul-verdoso. Las interacciones de tipo *pi-pi stacking* y los enlaces de hidrógeno se representan con líneas negras y rojas segmentadas, respectivamente. Por otra parte, las líneas discontinuas de color rosado hacen referencia a puentes salinos, mientras que las de color naranja denotan interacciones pi-catión.

Las diferencias existentes para estructuras diméricas con respecto al monómero, recae en la ausencia considerable del residuo Gln270 como causante de la formación de enlaces de hidrógeno, siendo reemplazo por Tyr269, el cual establece este tipo de interacción con el grupo NH de la carboxamida ([Figura 5.2D](#)). Este cambio ocurre ya que el bucle articulado BL2 existe en diferentes conformaciones para cada uno de los estados proteicos estudiados, debido a un mecanismo de unión de ajuste inducido de este tipo de compuestos (Ratia et al., 2008). Para el estado dimérico de la proteína PLpro, el grupo donante del residuo Gln270 se aleja de los inhibidores impidiendo la formación de enlaces (ver material complementario, [Figura S1](#)) y permitiendo a su vez la participación de la cadena principal de Tyr269.

La flexibilidad del bucle, no es lo único asociado a la pérdida de interacciones. Ligandos derivados de 6577871 presentan diferentes conformaciones, puntualmente en la porción protonada de estos. Esta región es causante de varias interacciones, involucrando a los residuos Asp165, Tyr269 y Tyr265. El primero de estos, además de formar una interacción del tipo puente salino, es partícipe en la formación de un enlace de hidrógeno, el cual se mantiene solo si el protón de la piperidina se encuentra en orientación al Aspartato. En ligandos como B7b (entre muchos otros), el hidrógeno de la región cargada de la piperidina se voltea ([Figura S2](#)), quedando expuesto a residuos como Tyr265 e incluso, en algunos casos, orientado hacia las Pro248-249, lo cual impide la formación del enlace mencionado. Sin embargo, las orientaciones de las poses de acoplamiento obtenidas en el bolsillo de unión de PLpro son similares a las de los compuestos co-cristalizados en los registros PDB 3E9S (Ratia et al., 2008) y 4OW0 (Baez-Santos et al., 2014).

Las puntuaciones de docking obtenidas para cada uno de los compuestos frente a los estados proteicos utilizados se tabulan en la [tabla complementaria S1](#). En este estudio, se encontró una baja correlación entre las energías de puntuación de Glide XP y los valores del registro experimental (pIC_{50}) para las poses acopladas al cristal monomérico ($R^2 = 0.0048$) y para la estructura dimérica ($R^2 = 0.1428$). La obtención de bajos coeficientes de correlación ha sido un tema de discusión recurrente en diferentes artículos, en los que se reporta en concreto que la puntuación de acoplamiento molecular y las afinidades de unión

determinadas experimentalmente, en general no se correlacionan de buena manera (Mikra et al., 2017).

Los resultados del acoplamiento molecular nos permiten conocer si los ligandos se unen correctamente en el sitio de unión, proporcionando un cálculo de la afinidad de unión (Chen, 2015). En cuanto a los modelos obtenidos para el monómero de PLpro, sugieren que los compuestos más pequeños (derivados de 7724772) son capaces de unirse de mejor manera que los compuestos que presentan el grupo de piperidina en su estructura. La puntuación *Glidescore* para este tipo de compuestos (derivados de 7724772) es más favorable, presentando valores de -7.849, -7.825 y -7.514 kcal/mol en ligandos como D47, D02 y D49, respectivamente. En contraste a lo mencionado, los compuestos derivados de 65778771 como B15f, B7a y C3J poseen puntuaciones menos favorables: -5.594, -5.576 y -5.552 Kcal/mol, respectivamente.

El comportamiento de las puntuaciones obtenidas en el cristal dimérico es totalmente opuesto a los casos anteriores, ya que esta vez presentan una mejor puntuación aquellos ligandos que tienen restos de piperidina en su estructura. Derivados del naftaleno como C4d(S) (-11.456 Kcal/mol), C3e (-10.964 Kcal/mol) y C2b (-10.470 kcal/mol) tienen una mayor afinidad de unión. Si bien para este estado, D47, D02 y D49 no presentan bajas puntuaciones (-8.301, -8.364 y -8.295), se aprecia que los compuestos con un anillo de piperidina tienden a unirse mejor al dímero de SARS-CoV PLpro.

La pose de acoplamiento más favorable para la unión con PLpro en estado monomérico, D47, mantiene las interacciones descritas previamente de manera general y reportadas en la literatura para compuestos derivados de 7724772 (Gosh et al., 2010). La estructura del naftaleno y el benceno presentes en el ligando, así como también la disposición del anillo del residuo de Tyr269 en el bucle BL2, permiten la formación de interacciones *pi-pi stacking* entre ambas partes. A su vez, Gln270 y Asp165 establecen enlaces de hidrógeno con la carboxamida del ligando, puntualmente con los grupos carbonilo y amino, respectivamente, lo cual conduciría al cierre del bucle. Además, el sustituyente presente en la posición 5 del anillo aromático (tert-butoxicarbonil metilamino) contribuye a la formación de enlaces de hidrógeno con los residuos Glu168 y Tyr269 (Figura 5.3A). En cuanto a la pose de mejor

puntuación para la estructura dimérica, C4d(S), esta mantiene las interacciones *pi-pi stacking* con el residuo Tyr269, pese a la sustitución del anillo de naftilo por una isoquinolina. El nitrógeno cargado de la piperidina, es causante de un gran número de interacciones, como: enlace de hidrógeno, puente salino con el residuo Asp165 y una interacción pi-cación con el residuo Tyr265. Como se mencionó con anterioridad, para este estado del receptor, el grupo amino del ligando establece un enlace de hidrógeno con la cadena lateral de la Tyr269 (Figura 5.3B).

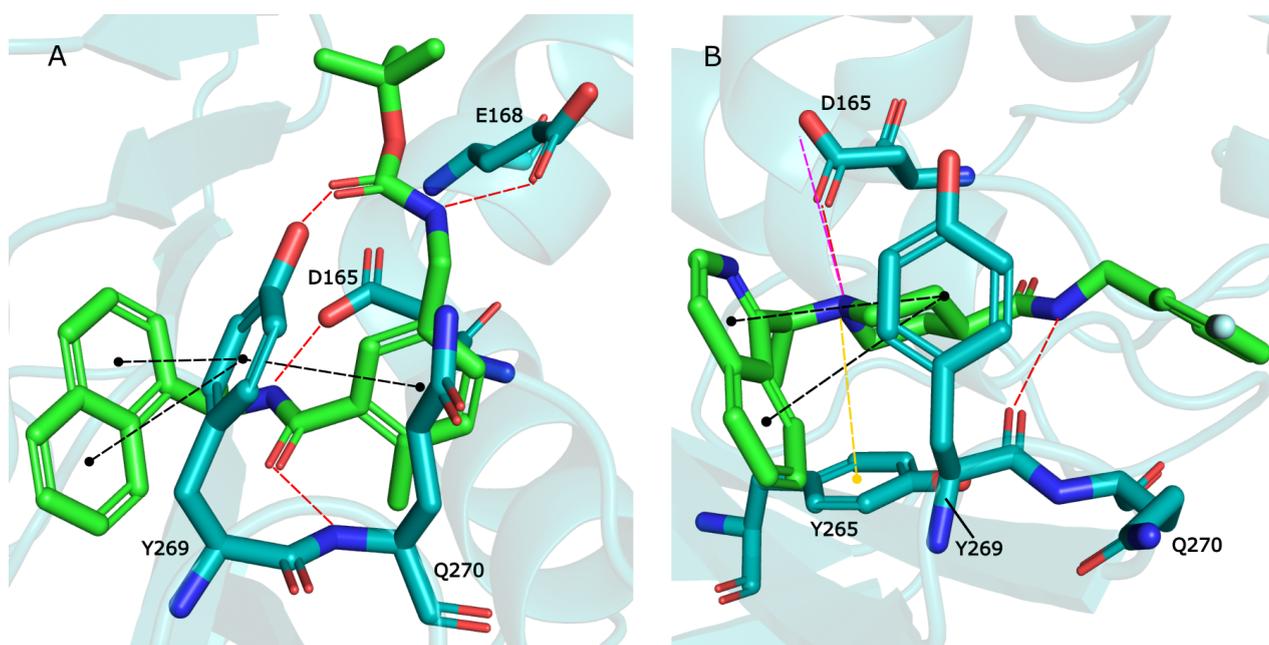


Figura 5.3: Mejores poses obtenidas para la estructura monomérica y dimérica de SARS-CoV PLpro. (A) Compuesto D47 acoplado dentro del sitio de unión de SARS-CoV PLpro (gscore = -7.849 Kcal/mol). (B) Compuesto C4d en su configuración S, acoplado en el sitio de unión de la proteasa (gscore = -11.456 Kcal/mol). Ligandos se representan por medio de barras verdes. Los residuos involucrados en el sitio de unión se colorearon de azul-verdoso. Las interacciones de tipo *pi-pi stacking* y enlaces de tipo hidrógeno se señalan por líneas negras y rojas segmentadas, respectivamente. Por otra parte, las líneas discontinuas de color rosado hacen referencia a enlaces puente salino, mientras que las de color naranja denotan interacciones pi-cación.

5.2 Comparación entre las poses

A partir de los resultados del acoplamiento molecular entre los compuestos de la serie

congenérica y SARS-CoV PLpro, se realizaron cálculos de RMSD para comparar las orientaciones de los ligandos. Para esto, se utilizó el servidor LigRMSD (Velázquez-Libera et al., 2020), el cual nos permitió llevar a cabo comparaciones entre las poses obtenidas para los 75 compuestos, con respecto a los inhibidores cocrystalizados en 3E9S y 4OW0 y con respecto a los ligandos 7724772 y 6577871, identificando los grafos comunes entre las moléculas y calculando la desviación cuadrática media (RMSD) al comparar estos. Con el objetivo de validar el protocolo empleado, primeramente, se obtuvieron los valores de LigRMSD para los compuestos derivados del naftaleno co-cristalizados con los identificadores 3E9S, 3MJ5, 4OVZ y 4OW0, utilizando el inhibidor de la estructura 4OW0 como referencia (Figura 5.4). Los resultados se presentan en la Tabla 5.1, donde es posible apreciar que los compuestos con valores elevados de porcentaje de referencia (%Ref) y valores elevados respecto a la molécula blanco (%Mol) son más parecidos entre ellos. La diferencia para el ligando presente en 3E9S (valores bajos de %Mol y %Ref) se debe a la ausencia del anillo de piperidina adyacente al centro estereogénico, resultando ser un compuesto más corto que los derivados de 6577871.

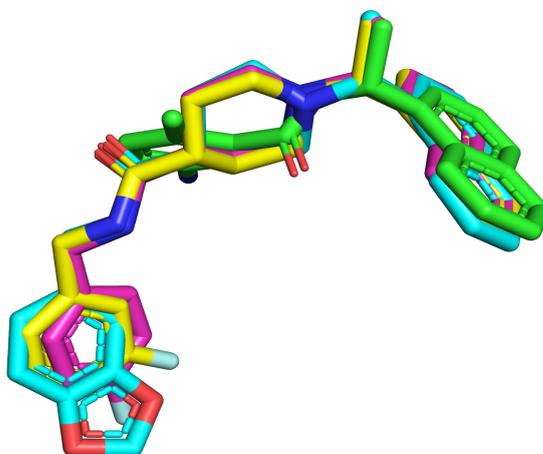


Figura 5.4: Alineamiento estructural de compuestos derivados del naftaleno con identificadores 3E9S (Verde), 3MJ5 (Cian), 4OVZ (Magenta), 4OW0 (Amarilla).

Molécula	RMSD	%Ref	%Mol
4OVZ	0.34	96.55	96.55
3MJ5	0.51	96.55	90.32

3E9S	1.01	41.38	52.17
------	------	-------	-------

Tabla 5.1: Resultado del servidor LigRMSD para compuestos co-cristalizados. Se informan valores de RMSD, %Ref y %Mol para los compuestos cristalizados en 4OVZ, 3MJ5, 3E9S, utilizando como referencia la estructura 4OW0.

Con respecto al RMSD, es aceptado en la literatura que valores menores a 2Å reflejan una relación espacial importante entre las estructuras que se comparan (Ramírez et al., 2018). Por esta razón, a partir de la información tabulada, bajos valores de RMSD con valores altos de %Ref y %Mol, revelan una misma orientación de una serie congénica de inhibidores. Los resultados obtenidos para los 75 inhibidores estudiados, considerando las estructuras 3E9S (estado monomérico) y 4OW0 (SARS-CoV PLpro dimérico), se presentan en la [Tabla complementaria S1](#). Las comparaciones en donde se utilizó el inhibidor co-cristalizado en 3E9S como referencia (compuesto GRL0617), fueron más útiles para caracterizar las orientaciones de los ligandos carentes del anillo de piperidina. El conjunto de compuestos derivados de la estructura 7724772 (compuestos con prefijo A y D, más 7724772 y GRL0617; [Tablas 4.1 y 4.4](#)), exhibió valores de %Ref superiores al 90%, al igual que el %Mol en la mayoría de los casos (exceptuando los compuestos A6, D47 y D49). Estos valores muestran que la composición estructural de estos compuestos es muy similar, lo cual es ideal para hacer la comparación entre las poses. En cuanto a los valores de RMSD para este subconjunto de ligandos, se estuvieron en su mayoría valores en el rango de 0.25 a 1.5 Å, siendo solo 5 compuestos aquellos que presentaron valores de RMSD por encima de 2 Å. Esto se debió a que el grupo naftaleno de estas estructuras se situó de manera opuesta al mismo grupo en la referencia. Pese a esto, el conjunto completo de ligandos de este tipo, se orientó de manera similar al compuesto co-cristalizado GRL0617.

Por otra parte, considerando como referencia el ligando C3k co-cristalizado en la estructura dimérica 4OW0, se tiene que de manera general los valores de RMSD, %Ref y %Mol fueron favorables para ligandos que presentan el sustituyente piperidina en su composición (51 compuestos con prefijos B y C; [Tablas 4.2 y 4.3](#) respectivamente). Los valores de RMSD < 2 Å se exhibieron en 35 estructuras, mientras que las 16 restantes presentaron valores superiores al umbral (RMSD entre 2 y 3Å para 12 compuestos y RMSD > 3Å para 4 de ellos).

En la mayoría de los casos, estos valores elevados en la desviación cuadrática media se deben a flexibilidad de los ligandos en el sitio de unión de la proteasa, en donde el grupo aromático de los compuestos se orienta de manera opuesta a la referencia. Ejemplo de esto es el compuesto B15j. El grupo benceno de esta estructura, a diferencia del inhibidor de referencia C3k, se posicionó en dirección al residuo Lys158 en lugar de la Gln270, induciendo en algunos casos la formación de interacciones pi-cación descritas en la [sección 5.1](#). En adición a los valores de RMSD, la mayoría de los ligandos de este tipo poseen porcentajes superiores al 93% y 84%, con respecto a la referencia (%Ref) y a la molécula blanco (%Mol), respectivamente. En la [Tabla complementaria S1](#), se encontraron además estructuras que exhiben un %Mol de 75.68%, debido principalmente a la diferencia de sustituyentes de la referencia con respecto a estos compuestos (C1d(R) y C1d(S)). Por otro lado, las comparaciones sobre las estructuras C6a y C6b reflejan valores de %Ref de 75.86% y %Mol de 73.33%, como consecuencia de la diferencia de tamaño de la cadena principal. Considerando los antecedentes descritos anteriormente, se puede afirmar que los inhibidores con restos de piperidina en su estructura se orientan de manera similar al ligando C3k co-cristalizado, que es la referencia más cercana para calcular un valor de RMSD para este tipo de compuestos.

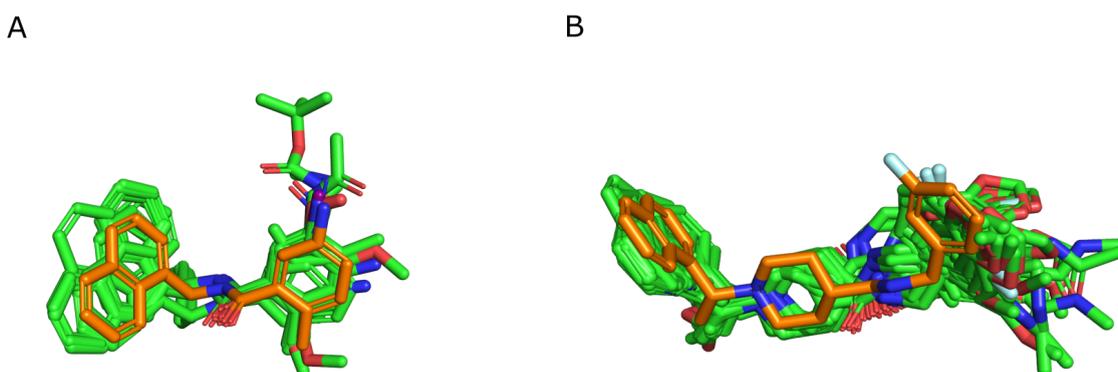


Figura 5.5: Similitud estructural de las poses de docking respecto a las referencias 3E9S y 4OW0. (A) Inhibidores derivados de 7724772 acoplados al cristal monomérico de SARS-CoV PLpro, utilizando el compuesto GRL0617 co-cristalizado en 3E9S como referencia. (B) Inhibidores derivados de 6577871 acoplados al cristal dimérico de la proteasa, utilizando el ligando co-cristalizado 4OW0 (C3k) como

referencia. Para cada uno de los casos, las referencias se representan como estructuras de color naranja, mientras que las poses obtenidas por medio del acoplamiento, se muestran en verde.

En síntesis, conociendo los valores de %Ref, %Mol y RMSD de los compuestos, así como también la causa de sus variaciones, se desprende que los inhibidores co-cristalizados en 3E9S y 4OW0 fueron más apropiados para caracterizar las orientaciones de los compuestos derivados de 7724772 y 6577871, respectivamente (Figura 5.5). El análisis en detalle permite mostrar que se obtuvieron orientaciones similares para todos los compuestos, con algunos grupos colocados en zonas flexibles del sitio de unión ocasionalmente orientados de manera diferente con respecto a su referencia.

5.3 Interacciones con residuos en el sitio de unión de SARS-CoV PLpro

Además de analizar las poses obtenidas a partir del protocolo de acoplamiento, mediante el uso del servidor LigRMSD, se realizó una anotación de las interacciones recurrentes que se observaron entre los ligandos acoplados y el sitio de unión de PLpro. Las interacciones químicas entre la serie de los compuestos derivados del naftaleno en el sitio de unión de SARS-CoV PLpro, se describen en detalle mediante el uso de “Huellas de interacción molecular” (IFP) (Deng et al., 2004). La utilización de esta herramienta permite un análisis detallado del sitio de unión de PLpro que interacciona con la serie congénica.

Los IFP calculados se informan en la Figura 5.6. Por medio del análisis de huellas de interacción molecular aplicado a los 75 complejos entre SARS-CoV PLpro y las poses obtenidas por el acoplamiento de los derivados del naftaleno, se reveló que 20 residuos de la proteasa en estado monomérico, tenían contactos con los inhibidores (Figura 5.6A). Por otra parte los IFP asociados al dímero 4OW0 indicaron la presencia de 21 residuos participantes en los contactos proteína-ligando (Figura 5.6C). Entre estos residuos, se encuentran aquellos responsables del ajuste de los inhibidores al sitio de unión S3 y S4 de la proteasa, como lo son Asp165, Tyr269 y Gln270 (Baez-Santos et al., 2015).

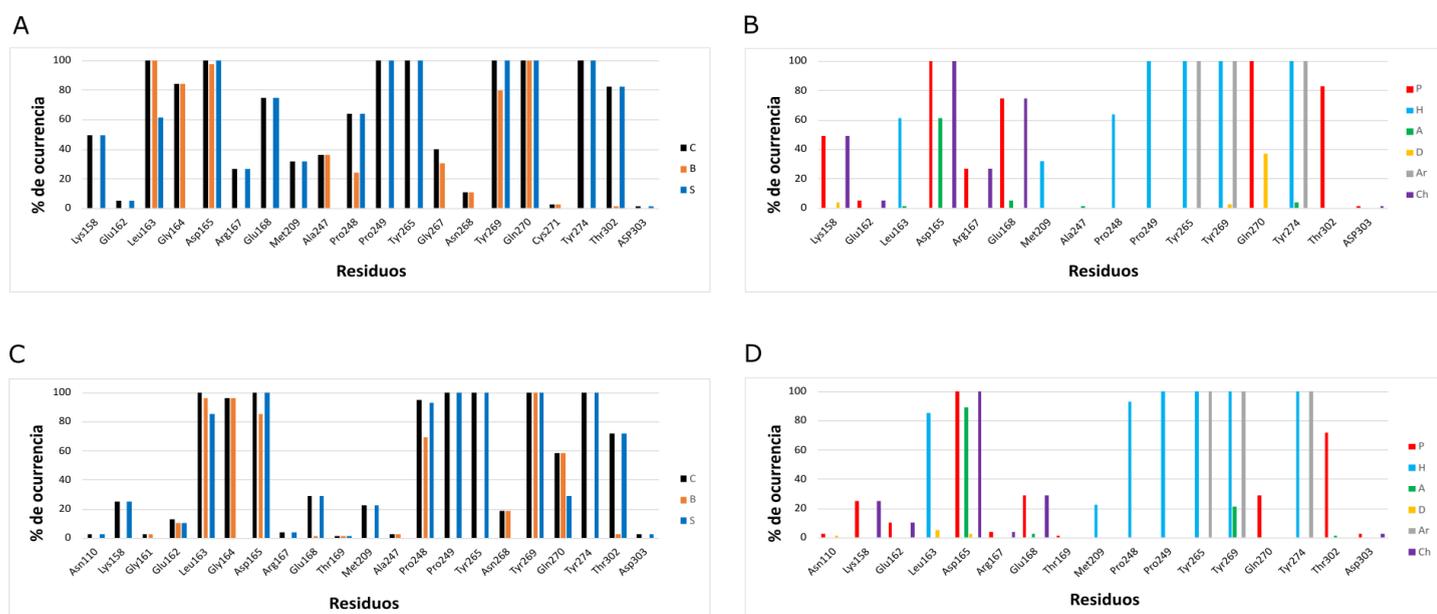


Figura 5.6: Interacciones presentes en los compuestos acoplados a los diferentes cristales de SARS-CoV PLpro. (A-B) Interacciones asociadas al cristal monomérico de PLpro. (C-D) Interacciones correspondientes al estado dimérico de la proteasa. Cada una de las interacciones en la gráfica se presentan como porcentaje de ocurrencia de contactos [C], interacciones con la columna vertebral del residuo [B] e interacciones con la cadena lateral del residuo [S], porcentaje de ocurrencia de interacciones químicas: enlaces de hidrógeno (HBs) donde el residuo es aceptor [A], HBs donde el residuo es donador [D], polar [P], hidrofóbico [H], Aromático [Ar] y electrostático con grupos cargados [Ch].

De manera general, el sitio de unión a ligandos de SARS-CoV PLpro posee regiones polares y apolares. Para ambos estados de la proteína las contribuciones hidrofóbicas ocurren en el 100% de las estructuras acopladas, así como también los contactos que involucran residuos aromáticos. En los complejos de inhibidores derivados del naftaleno con la estructura cristalizada 3E9S, la orientación de los restos de naftaleno contra los residuos apolares Pro248 y Pro249, propiciaron la ocurrencia de contactos hidrofóbicos en la interfaz proteína-ligando. Alrededor del 63% de los ligandos establecieron interacciones de este tipo con el residuo de Prolina 248. Por otra parte, Pro249 presenta un porcentaje de casos del 100%, para el mismo tipo de interacción. Las cadenas laterales de los residuos de Tyr265, Tyr269 y Tyr274 también contribuyeron a los contactos hidrofóbicos en la interfaz

proteína-ligando. Estos tres residuos, no solo fueron causantes de altos porcentajes en el factor de hidrofobicidad, sino que también fueron partícipes en interacciones aromáticas en el 100% de las estructuras totales. Puntualmente, la cercanía del residuo Tyr269 con la porción de naftaleno de los inhibidores, permitió la formación de interacciones del tipo *pi-pi stacking* en el 100% de las estructuras, explicando de esta forma el porcentaje de contribución para este residuo. La interacción anterior resulta fundamental, pues es causante de la conformación cerrada adoptada por el bucle de bloqueo BL2 en presencia de este tipo de compuestos (Baez-Santos et al., 2014). Con mayor detalle, en adición a esta información y en relación con el acoplamiento del estado monomérico de SARS-CoV con los 75 inhibidores recopilados, se observó lo siguiente (Figura 5.6B):

- El residuo Tyr269, al igual que Tyr274, además de formar interacciones hidrófobas y aromáticas en el 100% de las estructuras acopladas, actuó como aceptor y donador de HB respectivamente, en menos del 5% de los casos.
- La proximidad de la Thr302 en relación a los átomos de los compuestos acoplados ocurrió en más del 80% de las estructuras totales.
- El residuo Lys158 tuvo interacciones polares en más del 40% de las estructuras acopladas, además de contribuir en interacciones electrostáticas y ser donador de HB en el 40% y menos del 5% respectivamente.
- El residuo Met209 contribuyó en interacciones polares en menos del 40% de los compuestos analizados, mientras que el residuo de Leu163 lo hizo en el 60% aproximadamente, además de actuar como aceptor de HB en el 2% de los casos.
- El residuo Glu168 tuvo interacciones polares y electrostáticas en más del 70% de las estructuras y fue aceptor de HB solo en un 2%.
- El residuo Gln270 tuvo interacciones polares en el 100% de los casos, además de una contribución de sobre el 35% como donador de HB.
- El residuo Arg167 aportó en interacciones electrostáticas y polares en un porcentaje superior al 20%.
- El residuo Asp165, además de contribuir en un 100% en interacciones polares y electrostáticas, fue aceptor de HB en más del 60% de las estructuras acopladas.

Sumados a estos, los IFP identificaron contribuciones asociadas a interacciones químicas menores al 10%, en los siguientes residuos:

- El residuo Ala247 actuó como aceptor de HB en menos del 5% de las estructuras acopladas, específicamente en el compuesto C1b(R), en donde el grupo carbonilo del backbone del residuo mencionado, estableció un enlace de hidrógeno con el grupo hidroxilo unido al metilo en el centro quiral del ligando.
- Al igual que en el caso anterior, el residuo Asp303 tuvo interacciones electrostáticas y polares en menos del 5% de los casos. Puntualmente, este porcentaje del total corresponde al compuesto C1c(S).
- Por último, el residuo Glu162 contribuyó en interacciones polares y electrostáticas en aproximadamente el 10% de las estructuras. Compuestos como C3d, C3c, C3b y C1a(R) forman parte de este conjunto.

Los análisis de IFP para las estructuras acopladas al estado dimérico de SARS-CoV (Figura 5.6D), revelaron lo siguiente:

- Como se mencionó anteriormente, los residuos de Tyr265, Tyr269 y Tyr274 contribuyeron a interacciones aromáticas e hidrófobas en el 100% de los casos. Además, Tyr269 participó como aceptor de HB en más del 20% de las estructuras acopladas.
- El residuo Thr302 tuvo interacciones polares en más del 70% de las estructuras totales, además de contribuir en menos de un 5% como aceptor de enlace de hidrógeno.
- Los residuos Met209, Pro248 y Pro249 tuvieron interacciones polares en el 100%, 93% y 86% de las estructuras acopladas, respectivamente.
- El residuo Lys158 contribuyó en más del 20% en interacciones electrostáticas y polares.
- El residuo Leu163 contribuyó en más de un 80% en interacciones polares y actuó como donador de HB en un 5% de los casos.
- Los residuos Gln270 y Glu168 establecieron alrededor de un 30% de interacciones polares. En adición a esto, el glutamato contribuyó en interacciones electrostáticas y como aceptor de HB en un 30% y 3% de los casos, respectivamente.
- El residuo Glu162 tuvo interacciones en alrededor del 10% de las estructuras totales.
- El residuo Asp165 contribuyó en interacciones polares y electrostáticas en un 100% de las estructuras acopladas. Además, actuó como aceptor de HB en

aproximadamente el 90% de los casos y como donador de HB en menos del 5% de las estructuras totales.

Al igual que en el estado monomérico, para la estructura dimérica también existieron residuos con contribuciones menores al 10%:

- El residuo Thr169 presentó interacciones polares en menos del 5% de las estructuras analizadas, específicamente en el compuesto B15j.
- El residuo Asp303 tuvo interacciones polares y electrostáticas en menos del 5% de los casos. Ejemplos de esto, son los compuestos C1d(R) y C1d(S).
- El residuo Asn110 presentó interacciones con algunas estructuras acopladas. Contribuyó como donante de HB en menos del 3% de los casos.
- La Arg167 contribuyó en interacciones polares y electrostáticas en alrededor del 5% de las estructuras totales.

La descripción en detalle derivada de las huellas de interacción molecular (IFP) sobre las estructuras acopladas demuestra la presencia de residuos identificados previamente en los complejos cristalizados como causantes de interacciones fundamentales entre PLpro y los compuestos derivados del naftaleno. En este sentido, la descripción detallada de las poses de los compuestos derivados de naftaleno podría considerarse para el diseño de nuevos candidatos con alta capacidad inhibitoria frente a la proteasa.

5.4 Modelos de correlación 3D para describir actividades diferenciales

Para finalizar, se construyeron modelos de relación cuantitativa estructura-actividad (QSAR) para identificar las características estructurales de los compuestos derivados del naftaleno que afectan su actividad inhibitoria contra SARS-CoV PLpro. El paso más importante para la construcción de modelos 3D-QSAR corresponde a la alineación de las moléculas (Guasch et al., 2012). Para este proceso, se utilizaron diferentes estrategias (Figura 5.7): (1) alineamiento de moléculas considerando la cadena más larga en un mismo plano, (2) alineamiento de moléculas considerando las cadenas más largas con los átomos pesados en zig-zag y los sustituyentes en conformación anti, y (3) estrategias basadas en la alineación obtenida a través del acoplamiento molecular (Muñoz-Gutiérrez, et al., 2017), considerando

de manera efectiva la disposición del ligando en el sitio de unión de SARS-CoV PLpro. Los resultados para cada uno de los casos se exponen a continuación.

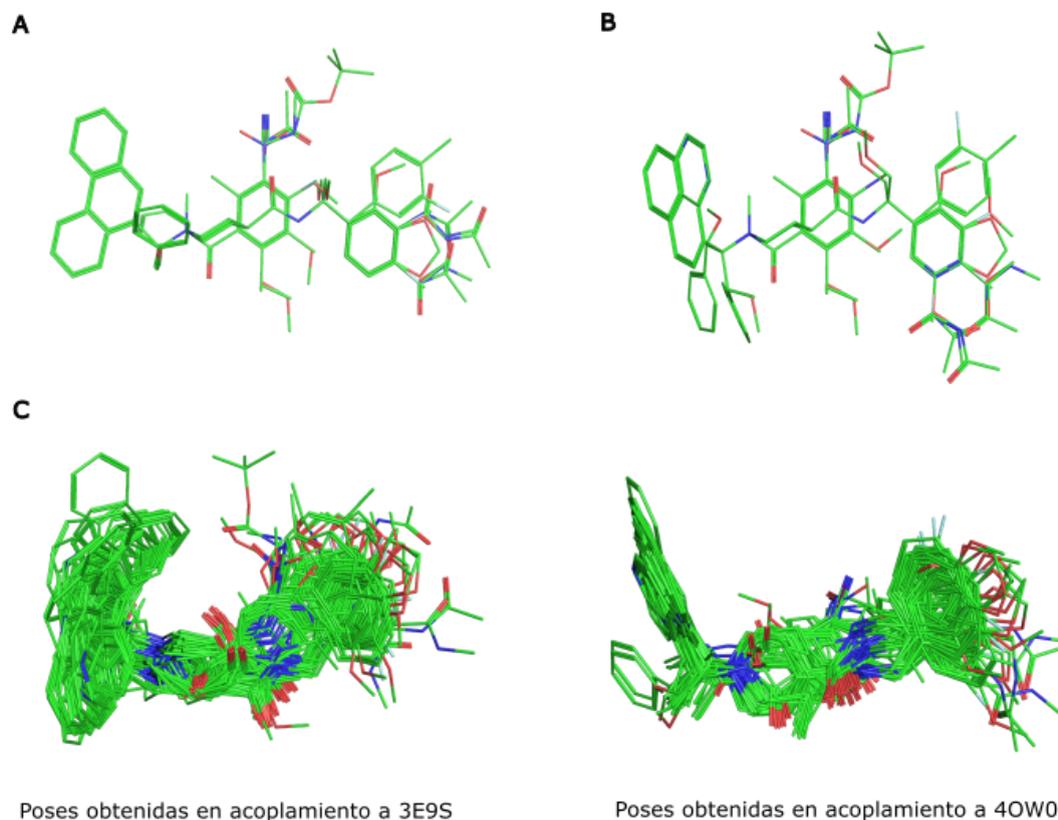


Figura 5.7: Alineamientos de los compuestos derivados del naftaleno mediante diferentes estrategias. (A) Estrategia basada en la disposición de la cadena más larga en un mismo plano. (B) Alineamiento considerando las cadenas más largas con los átomos pesados en zig-zag y los sustituyentes en anti. (C) Estrategia basada en la alineación de acoplamiento: Poses obtenidas para el estado monomérico 3E9S y el estado dimérico 4OW0.

Para las primeras dos estrategias establecidas (1 y 2), los 75 compuestos derivados del naftaleno se alinearon manualmente utilizando el programa Maestro, de la suite de Schrödinger (Maestro, Schrodinger, LLC, Nueva York, NY, EE. UU.). Los modelos generados empleando el módulo Field-Based QSAR, se construyeron considerando cinco descriptores de campo Gaussianos: estérico (S), electrostático (E), hidrófobo (H), donante de enlace de hidrógeno (D), aceptor de enlace de hidrógeno (A) y anillos aromáticos (R). Además, se

utilizaron hasta seis factores PLS, teniendo en cuenta las variables descritas en la [sección 4.4](#). El modelo generado para la primera estrategia, presentó valores de $R^2=0.526$, $Q^2=0.107$, $R^2_{\text{test}}=0.463$, desviación estándar (SD) de 0.6 aproximadamente ($SD = 0.5923$) y una estabilidad de 0.74. Para la segunda estrategia, el modelo presentó valores de $Q^2=0.149$ y $R^2_{\text{test}}=0.528$, mientras que la última estrategia exhibió valores de $Q^2= -0.054$ y $R^2_{\text{test}}=0.600$, correspondiente a las poses acopladas al monómero 3E9S y un Q^2 y R^2_{test} de -0.060 y 0.6057 , respectivamente, para la estructura dimérica 4OW0. Los detalles de los modelos se exponen a continuación en la [Tabla 5.2](#).

	Factor PLS	SD	R^2	Q^2	Estabilidad	R^2_{test}
Estrategia 1	5	0.5923	0.5263	0.1071	0.74	0.4632
Estrategia 2	6	0.5776	0.5777	0.1492	0.71	0.5284
Estrategia 3 (3E9S)	5	0.4571	0.7218	-0.0538	0.48	0.6002
Estrategia 3 (4OW0)	6	0.5046	0.6709	-0.0597	0.34	0.6057

Tabla 5.2: Información estadística de los modelos 3D-QSAR para las estrategias 1, 2 y 3. SD corresponde a la desviación estándar de la actividad del conjunto de entrenamiento. R^2 y Q^2 son coeficientes de correlación del set de entrenamiento y set de prueba respectivamente.

Considerando los valores estadísticos presentados en la [Tabla 5.2](#) se desprende que ninguno de los modelos construidos a partir de las diferentes estrategias, cumplió en su totalidad con las condiciones establecidas para ser seleccionado como óptimo ($Q^2 > 0.5$ y $R^2_{\text{test}} > 0.5$). Comúnmente, los bajos valores de Q^2 están asociados a una menor capacidad predictiva de los modelos QSAR (Böhm et al., 1999). Sin embargo, autores como Golbraikh y Tropsha (2002) y Doweiko (2004) sostienen que un valor alto de Q^2 no implica automáticamente una alta capacidad de predicción del modelo, si no que puede evidenciar la redundancia en el conjunto de entrenamiento. En relación a esto, un valor bajo de Q^2 refleja simplemente la importancia de cada miembro del conjunto de entrenamiento para el modelo (Doweiko, 2004; Golbraikh et al., 2002). No obstante, si bien la estadística no es suficiente para clasificar un modelo como predictivo, marca una condición necesaria. El valor de Q^2 es la

única forma de conocer si el modelo está sobreajustado, por lo tanto, el incumplimiento de esta condición proporcionaría modelos predictivos inadecuados.

Con el esfuerzo por dilucidar el comportamiento de Q^2 en los QSAR obtenidos, se generaron modelos utilizando distintas combinaciones de campo. De esta forma, se pudo apreciar como la inclusión de estos, incidía en esta variable estadística. Los modelos construidos con cada uno de los campos gaussianos estudiados, de manera independiente entregaron valores de validación cruzada (Q^2) menores a 0.18. Las diferentes combinaciones de descriptores estérico, electrostático, hidrofóbico, de donador y aceptor de enlaces de hidrógeno no generaron grandes variaciones en el valor de Q^2 . La incorporación sistemática de los campos electrostático, hidrofóbico y aceptor de enlace de hidrógeno provocó un aumento en el valor de la validación cruzada, resultando de 0.2526 con tres componentes. El aumento de este valor, fue el mayor obtenido para las combinaciones empleadas, sin embargo no favoreció significativamente al modelo en términos estadísticos y en consecuencia, no permitió la construcción de un modelo predictivo fiable.

Bajo el mismo paradigma, se realizó una identificación de los compuestos que más afectan al valor de Q^2 . La presencia de valores atípicos en el conjunto de datos puede deformar las predicciones y afectar la precisión del modelo (Kim, 2021). Se consideraron las variaciones entre los valores experimentales de pIC_{50} y los valores predichos para la construcción de gráficas que permitieran revelar compuestos “problemáticos” en términos predictivos. Los resultados se exponen en la [Figura 5.8](#).

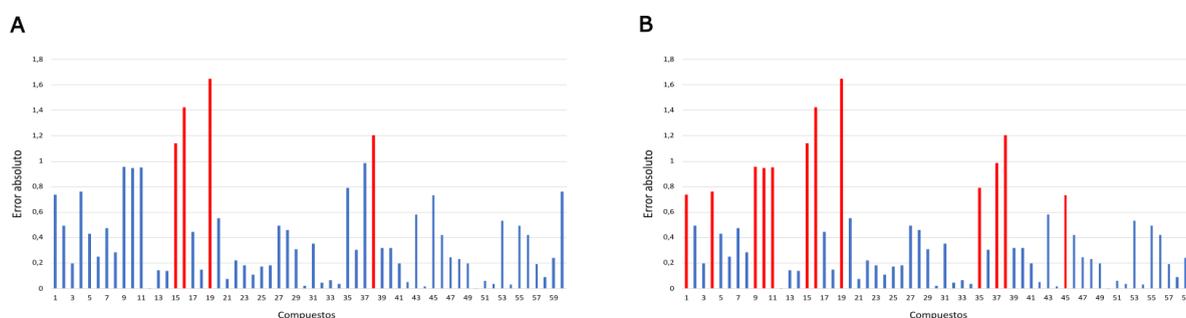


Figura 5.8: Errores en la predicción de la actividad de los compuestos derivados del naftaleno. El error absoluto para cada uno de los inhibidores se presenta por medio de barras en la gráfica. Barras

rojas señalan compuestos con un mayor error en la predicción de actividad. A) Presencia de compuestos con errores superiores a 1. B) Presencia de compuestos con errores superiores a 0.6.

En relación a la gráfica anterior (Figura 5.8A), los inhibidores 15, 16, 19 y 38, correspondientes a B15g, B15h, B15k y C3k, respectivamente, obtuvieron variaciones significativas en las actividades predichas con respecto a lo experimental. En base a esto, se construyeron nuevos modelos, excluyendo los compuestos mencionados y se evaluó la variación de la validación cruzada del conjunto de entrenamiento en ausencia de cada uno de ellos. La tabla 5.3 muestra los resultados obtenidos por la exclusión de cada uno de los compuestos de manera individual, así como también a través de diferentes combinaciones.

Compuesto	Factor PLS	Q ²
B15g	3	0.1964
B15h	3	0.2300
B15k	3	0.2723
C3k	3	0.2216
B15g - B15h	2	0.2709
B15g - B15h - B15k	3	0.2819
B15g - B15h - B15k - C3k	3	0.2843

Tabla 5.3: Variación del parámetro estadístico Q², en ausencia de diferentes compuestos.

En la Tabla 5.3 se observa el aumento de los valores asociados a la validación cruzada, conforme aumenta el número de compuestos excluidos del conjunto de entrenamiento. Como se puede observar, existen pequeñas mejoras en el valor de Q², sin embargo no son suficientes para permitir la construcción de modelos predictivos. Por otra parte, considerar además la exclusión de aquellos compuestos con variaciones entre la actividad experimental y la actividad predicha superiores a 0.6 (Figura 5.8B), permitió finalmente la obtención de un modelo con un valor aceptable de Q²=0.5470. Los compuestos excluidos fueron 65778871, A4, B15a, B15b, B15c, B15g, B15h, B15k, C3g, C3i, C3k, C6a y GRL0617. De estos 13 inhibidores, solo 2 forman parte del subconjunto de ligandos cortos derivados de 7724772

(A4 y GRL0617), mientras que los 11 restantes presentan andamios más largos, con restos de piperidina en su estructura (derivados de 6577871). Considerando estos compuestos designados como “outliers” y su papel en el ajuste del modelo, se plantearon las siguientes inferencias:

- La razón por la que compuestos como A4 y B15K presentan grandes errores en su predicción, estaría asociada a las escasas muestras cercanas a sus niveles de actividad en el conjunto de entrenamiento. Estos inhibidores presentan valores de $IC_{50} = >200 \mu M$, por lo que la presencia de pocos compuestos con valores similares, tendría como consecuencia un bajo aprendizaje por parte del modelo.
- En contraste al punto anterior, los compuestos 65778871, B15a, B15b, B15c, B15g, B15h, C3g, C3i, C3k, C6a y GRL0617 están distribuidos en todo el rango de actividades. Sin embargo, el problema asociado a las actividades no es el único motivo por el cual se pueden producir valores atípicos.

Zhao y col. plantean que la calidad de los datos puede ser diferente según la naturaleza de los protocolos experimentales, por lo que la precisión de los modelos QSAR se ve afectada en gran medida por una posible imprecisión de los datos biológicos (Zhao et al., 2017); en base a esto, se sugiere que los datos sean medidos mediante un único protocolo e idealmente en el mismo laboratorio. Por otra parte, se ha propuesto que los valores atípicos podrían estar asociados a un sitio de unión flexible. Como se demostró en la [sección 5.1](#), los compuestos derivados del naftaleno son moléculas análogas, que se unen al mismo sitio de unión en una forma esencialmente idéntica. Sin embargo, las proteínas son estructuras dinámicas y experimentan cambios conformacionales al unirse al ligando. Esta flexibilidad podría ser un punto fundamental para comprender los modelos QSAR, puntualmente los outliers que se presentan en este escrito. Las actividades biológicas observadas en los compuestos, se deben a las interacciones proteína-ligando. Los descriptores para los grupos funcionales de inhibidores estructuralmente análogos, pero con modos de unión inusuales, no se calculan para reflejar las mismas interacciones proteína-ligando si el sitio de unión se vuelve diferente para cada uno de ellos (Kim, 2007). Estudiar el mecanismo de unión de estos compuestos en un sitio de unión flexible, podría entregar respuestas asociadas a la exclusión de estos inhibidores y la repercusión en el ajuste del modelo QSAR, así como

también permitirnos aclarar por qué los modelos 3D-QSAR generados no son capaces de describir la relación estructura-actividad.

En síntesis y en relación a lo discutido a lo largo de esta sección, se destaca la expulsión de 13 compuestos denominados outliers, para lograr un aumento en el parámetro estadístico $Q^2=0.547$. Pese a esto, los modelos con un gran número de valores atípicos en el conjunto de datos, no son considerados buenos modelos (Verma et al., 2005). La utilización de los 75 compuestos de la serie congénica para describir la relación estructura-actividad no entregó resultados satisfactorios, ya que no se cumple con todos los requisitos estadísticos necesarios para correlacionar las estructuras de los compuestos con su actividad biológica. No obstante, se continuará con los esfuerzos por dilucidar el rol de los sustituyentes en la actividad biológica de esta serie de ligandos.

6. CONCLUSIONES Y PROPUESTAS FUTURAS

6.1 Conclusiones

De manera general, se realizó un estudio de 75 inhibidores derivados del naftaleno de la serie reportada por Ratia y col. Este trabajo nos permitió conocer con mayor detalle las características esenciales de esta clase de inhibidores y el comportamiento que estos presentan en el sitio de unión de los diferentes estados de SARS-CoV PLpro. Las interacciones moleculares, así como también la orientación de los compuestos de esta serie congénica se demostraron mediante un protocolo de docking molecular, análisis de huellas de interacción molecular y uso del servidor LigRMSD. Estas herramientas nos entregaron una descripción exhaustiva de la interfaz proteína-ligando, así como también de los inhibidores como tal. Sumado a esto, se generó un modelo QSAR predictivo con la finalidad de conocer las características que explican la actividad diferencial de los derivados del naftaleno. Al considerar cada uno de los métodos implementados en este trabajo, se puede concluir que:

- Existe homogeneidad en la orientación de los compuestos que forman parte de la serie congénica. Esto se comprobó por medio de los resultados del acoplamiento molecular, además del uso del servidor LigRMSD.
- A través de la estrategia de docking puntualmente, se desprende que el grupo naftilo se dispone en regiones adyacentes a los residuos de Pro248, Pro249 y Tyr269, estableciendo interacciones *pi-pi stacking* importantes para el ajuste de los inhibidores.
- La región de la benzamida neutra de los derivados cortos, así como también la piperidina protonada de los de los compuestos originados de 6577871, establece principalmente enlaces de hidrógeno con el residuo Asp165.
- Las poses obtenidas por el acoplamiento fueron utilizadas para el análisis de las huellas de interacción molecular (*interaction fingerprinting*), en donde se demostró que las interacciones críticas para los estados monoméricos y diméricos de SARS-CoV PLpro involucran a los residuos Asp165, Tyr269, Tyr265, Tyr274, Thr302, Pro248, Pro249 y Leu163.

- Los modelos 3D QSAR contruidos y presentados en este escrito no poseen variables estadísticas satisfactorias para ser considerados modelos predictivos. La exclusión de compuestos desde el conjunto de entrenamiento, con mayores variaciones en la predicción de su actividad biológica con respecto a la experimental, permitió identificar posibles outliers de los modelos.

De manera general, proporcionar información adicional a los antecedentes estructurales disponibles en literatura utilizando una diversidad de herramientas computacionales, nos permite contribuir positivamente en el campo investigativo orientado a la SARS-CoV PLpro. Como se menciona en este escrito, la similitud estructural de la PLpro de SARS-CoV con la de SARS-CoV2 enfatiza aún más la importancia de extraer interacciones cruciales y la inspección exhaustiva del sitio de unión de la proteasa.

6.2 Proyecciones futuras

Existen diferentes direcciones que se podrían tomar desde los resultados de esta tesis. A partir de esta investigación y con el objetivo de revelar el impacto de los sustituyentes en la actividad de los compuestos de esta serie congénica, se propone emplear un método de *cross-docking*. Dado que los datos cristalográficos representan la adaptación de una proteína a un ligando en particular, esta metodología nos permitirá aumentar la diversidad de conformaciones disponibles de la estructura proteica en estudio, además de ofrecernos información relevante de cómo la serie de inhibidores derivados del naftaleno se ajustan en forma y tamaño a estas estructuras de PLpro. En base a esto, se podría correlacionar los valores de pIC_{50} con las energías libres de unión estimadas para cada una de las poses seleccionadas. Contar con diversidad conformacional para el acoplamiento molecular puede ser clave para describir eventos vinculantes con mayor precisión y en consecuencia explicar por qué algunos inhibidores son más potentes que otros.

REFERENCIAS

- Amin, S. A., Ghosh, K., Gayen, S., & Jha, T. (2021). Chemical-informatics approach to COVID-19 drug discovery: Monte Carlo based QSAR, virtual screening and molecular docking study of some in-house molecules as papain-like protease (PLpro) inhibitors. *Journal of biomolecular structure & dynamics*, 39(13), 4764–4773. <https://doi.org/10.1080/07391102.2020.1780946>
- Amin, S.A., Ghosh, K., Singh, S. et al. Exploring naphthyl derivatives as SARS-CoV papain-like protease (PLpro) inhibitors and its implications in COVID-19 drug discovery. *Mol Divers* (2021). <https://doi.org/10.1007/s11030-021-10198-3>
- Baez-Santos, Y. M., Mielech, A. M., Deng, X., Baker, S., & Mesecar, A. D. (2014). Catalytic Function and Substrate Specificity of the Papain-Like Protease Domain of nsp3 from the Middle East Respiratory Syndrome Coronavirus. *Journal of Virology*, 88(21), 12511–12527. <https://doi.org/10.1128/jvi.01294-14>.
- Báez-Santos, Y. M., St. John, S. E., & Mesecar, A. D. (2015). The SARS-coronavirus papain-like protease: Structure, function and inhibition by designed antiviral compounds. *Antiviral Research*, 115, 21–38. [doi:10.1016/j.antiviral.2014.12.0](https://doi.org/10.1016/j.antiviral.2014.12.0)
- Baez-Santos, Y. M., Barraza, S. J., Wilson, M. W., Agius, M. P., Mielech, A. M., Davis, N. M., Baker, S. C., Larsen, S. D., & Mesecar, A. D. (2014). X-ray structural and biological evaluation of a series of potent and highly selective inhibitors of human coronavirus papain-like proteases. *Journal of medicinal chemistry*, 57(6), 2393–2412. <https://doi.org/10.1021/jm401712t>.
- Baker, S. C. (2008). Coronaviruses: Molecular Biology. *Encyclopedia of Virology*, 554–562. <https://doi.org/10.1016/B978-012374410-4.00769-X>.
- Barretto, N., Jukneliene, D., Ratia, K., Chen, Z., Mesecar, A. D., & Baker, S. C. (2005). The Papain-Like Protease of Severe Acute Respiratory Syndrome Coronavirus Has Deubiquitinating Activity. *Journal of Virology*, 79(24), 15189–15198. <https://doi.org/10.1128/jvi.79.24.15189-15198.2005>
- Böhm, M., Stürzebecher, J., & Klebe, G. (1999). Three-Dimensional Quantitative Structure–Activity Relationship Analyses Using Comparative Molecular Field Analysis and Comparative Molecular Similarity Indices Analysis To Elucidate Selectivity

- Differences of Inhibitors Binding to Trypsin, Thrombin, and Factor Xa. *Journal of Medicinal Chemistry*, 42(3), 458–477. [doi:10.1021/jm981062r](https://doi.org/10.1021/jm981062r)
- Bosken, Y. K., Cholko, T., Lou, Y.-C., Wu, K.-P., & Chang, C. A. (2020). Insights Into Dynamics of Inhibitor and Ubiquitin-Like Protein Binding in SARS-CoV-2 Papain-Like Protease. *Frontiers in Molecular Biosciences*, 7. [doi:10.3389/fmolb.2020.00174](https://doi.org/10.3389/fmolb.2020.00174)
 - Centers for Disease Control and Prevention (CDC) (2003). Update: Outbreak of severe acute respiratory syndrome--worldwide, 2003. *MMWR. Morbidity and mortality weekly report*, 52(12), 241–248.
 - Cherry, J. D., & Krogstad, P. (2004). SARS: The first pandemic of the 21st century. *Pediatric Research*, 56(1), 1–5. <https://doi.org/10.1203/01.PDR.0000129184.87042.FC>
 - Chen, Y.-C. (2015). Beware of docking! *Trends in Pharmacological Sciences*, 36(2), 78–95. [doi:10.1016/j.tips.2014.12.001](https://doi.org/10.1016/j.tips.2014.12.001)
 - Clemente, V., D'arcy, P., & Bazzaro, M. (2020). Deubiquitinating enzymes in coronaviruses and possible therapeutic opportunities for COVID-19. *International Journal of Molecular Sciences*, 21(10), 1–17. <https://doi.org/10.3390/ijms21103492>
 - Cockburn, W. C., Delon, P. J., & Ferreira, W. (1969). Origin and progress of the 1968-69 Hong Kong influenza epidemic. *Bulletin of the World Health Organization*, 41(3), 345–348.
 - Cui, J., Li, F., & Shi, Z. L. (2019). Origin and evolution of pathogenic coronaviruses. *Nature Reviews Microbiology*, 17(3), 181–192. <https://doi.org/10.1038/s41579-018-0118-9>.
 - Daviet, L., & Colland, F. (2008). Targeting ubiquitin specific proteases for drug discovery. *Biochimie*, 90(2), 270–283. <https://doi.org/10.1016/j.biochi.2007.09.013>.
 - Del Rio, C., & Guarner, J. (2015). Ebola: Implications and Perspectives. *Transactions of the American Clinical and Climatological Association*, 126, 93–112.
 - Deng, Z., Chuaqui, C., & Singh, J. (2004). Structural Interaction Fingerprint (SIFt): A Novel Method for Analyzing Three-Dimensional Protein–Ligand Binding Interactions. *Journal of Medicinal Chemistry*, 47(2), 337–344. [doi:10.1021/jm030331x](https://doi.org/10.1021/jm030331x).
 - Dixon, S. L., Smondryev, A. M., Knoll, E. H., Rao, S. N., Shaw, D. E., & Friesner, R. A. (2006). PHASE: A new engine for pharmacophore perception, 3D QSAR model

development, and 3D database screening: 1. Methodology and preliminary results. *Journal of Computer-Aided Molecular Design*, 20(10–11), 647–671. <https://doi.org/10.1007/s10822-006-9087-6>

- Demmler, G. J., & Ligon, B. L. (2003). Severe acute respiratory syndrome (SARS): A review of the history, epidemiology, prevention, and concerns for the future. *Seminars in Pediatric Infectious Diseases*, 14(3), 240–244. [https://doi.org/10.1016/S1045-1870\(03\)00056-6](https://doi.org/10.1016/S1045-1870(03)00056-6).
- Drosten, C., Günther, S., Preiser, W., van der Werf, S., Brodt, H.-R., Becker, S., ... Doerr, H. W. (2003). Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome. *New England Journal of Medicine*, 348(20), 1967–1976. <https://doi.org/10.1056/nejmoa030747>.
- Freitas, B. T., Durie, I. A., Murray, J., Longo, J. E., Miller, H. C., Crich, D., ... Pegan, S. D. (2020). Characterization and Noncovalent Inhibition of the Deubiquitinase and deISGylase Activity of SARS-CoV-2 Papain-Like Protease. *ACS Infectious Diseases*, 6(8), 2099–2109. <https://doi.org/10.1021/acsinfecdis.0c00168>.
- Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., ... Shenkin, P. S. (2004). Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medicinal Chemistry*, 47(7), 1739–1749. <https://doi.org/10.1021/jm0306430>.
- Ghosh, K., Amin, S. A., Gayen, S., & Jha, T. (2021). Unmasking of crucial structural fragments for coronavirus protease inhibitors and its implications in COVID-19 drug discovery. *Journal of molecular structure*, 1237, 130366. <https://doi.org/10.1016/j.molstruc.2021.130366>
- Ghosh, A. K., Takayama, J., Aubin, Y., Ratia, K., Chaudhuri, R., Baez, Y., Sleeman, K., Coughlin, M., Nichols, D. B., Mulhearn, D. C., Prabhakar, B. S., Baker, S. C., Johnson, M. E., & Mesecar, A. D. (2009). Structure-based design, synthesis, and biological evaluation of a series of novel and reversible inhibitors for the severe acute respiratory syndrome - Coronavirus papain-like protease. *Journal of Medicinal Chemistry*, 52(16), 5228–5240. <https://doi.org/10.1021/jm900611t>.
- Ghosh, A. K., Takayama, J., Rao, K. V., Ratia, K., Chaudhuri, R., Mulhearn, D. C., Lee, H., Nichols, D. B., Baliji, S., Baker, S. C., Johnson, M. E., & Mesecar, A. D. (2010). Severe acute respiratory syndrome coronavirus papain-like novel protease inhibitors:

Design, synthesis, protein-ligand X-ray structure and biological evaluation. *Journal of Medicinal Chemistry*, 53(13), 4968–4979. <https://doi.org/10.1021/jm1004489>.

- Ghosh, A. K., Xi, K., Johnson, M. E., Baker, S. C., & Mesecar, A. D. (2006). Progress in Anti-SARS Coronavirus Chemistry, Biology and Chemotherapy. *Annual Reports in Medicinal Chemistry*, 41(06), 183–196. [https://doi.org/10.1016/S0065-7743\(06\)41011-3](https://doi.org/10.1016/S0065-7743(06)41011-3).
- Golbraikh, A., & Tropsha, A. (2002). Beware of q2! *Journal of Molecular Graphics and Modelling*, 20(4), 269–276. [doi: 10.1016/S1093-3263\(01\)00123-1](https://doi.org/10.1016/S1093-3263(01)00123-1)
- Guasch, L., Sala, E., Valls, C., Mulero, M., Pujadas, G., and Garcia-Vallvé, S. (2012). Development of docking-based 3D-QSAR models for PPARgamma full agonists. *J. Mol. Graph. Model.* 36, 1–9. [doi: 10.1016/j.jmgm.2012.03.001](https://doi.org/10.1016/j.jmgm.2012.03.001)
- Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., & Banks, J. L. (2004). Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *Journal of Medicinal Chemistry*, 47(7), 1750–1759. <https://doi.org/10.1021/jm030644s>.
- Harcourt, B. H., Jukneliene, D., Kanjanahaluethai, A., Bechill, J., Severson, K. M., Smith, C. M., ... 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>.
- Horton, R. (2020). Offline: 2019-nCoV outbreak—early lessons. *The Lancet*, 395(10221), 322. [https://doi.org/10.1016/S0140-6736\(20\)30212-9](https://doi.org/10.1016/S0140-6736(20)30212-9).
- Hu, M., Li, P., Li, M., Li, W., Yao, T., Wu, J. W., ... Shi, Y. (2002). Crystal structure of a UBP-family deubiquitinating enzyme in isolation and in complex with ubiquitin aldehyde. *Cell*, 111(7), 1041–1054. [https://doi.org/10.1016/S0092-8674\(02\)01199-6](https://doi.org/10.1016/S0092-8674(02)01199-6).
- Kevin J. Bowers, Edmond Chow, Huafeng Xu, Ron O. Dror, Michael P. Eastwood, Brent A. Gregersen, John L. Klepeis, Istvan Kolossvary, Mark A. Moraes, Federico D. Sacerdoti, John K. Salmon, Yibing Shan, and David E. Shaw, "Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters," Proceedings of the ACM/IEEE Conference on Supercomputing (SC06), Tampa, Florida, 2006, November 11-17.

- Kim, K.H. Outliers in SAR and QSAR: 3. Importance of considering the role of water molecules in protein–ligand interactions and quantitative structure–activity relationship studies. *J Comput Aided Mol Des* 35, 371–396 (2021). <https://doi.org/10.1007/s10822-021-00377-7>
- Kim, K.H. Outliers in SAR and QSAR: 2. Is a flexible binding site a possible source of outliers?. *J Comput Aided Mol Des* 21, 421–435 (2007). <https://doi.org/10.1007/s10822-007-9126-y>
- Lam, W. K., Tsang, K., Ooi, C., Ip, M., & Chan-Yeung, M. (2005). Severe acute respiratory syndrome. *Respiratory Medicine: An Asian Perspective*, 307–323.
- Lee, H., Lei, H., Santarsiero, B. D., Gatz, J. L., Cao, S., Rice, A. J., ... Johnson, M. E. (2015). Inhibitor Recognition Specificity of MERS-CoV Papain-like Protease May Differ from That of SARS-CoV. *ACS Chemical Biology*, 10(6), 1456–1465. <https://doi.org/10.1021/cb500917m>.
- Lindner, H. A., Lytvyn, V., Qi, H., Lachance, P., Ziomek, E., & Ménard, R. (2007). Selectivity in ISG15 and ubiquitin recognition by the SARS coronavirus papain-like protease. *Archives of Biochemistry and Biophysics*, 466(1), 8–14. <https://doi.org/10.1016/j.abb.2007.07.006>.
- Loeb, M. B. (2006). Severe acute respiratory syndrome. *Infection Management for Geriatrics in Long-Term Care Facilities, Second Edition, 2004*(December 2003), 465–472. <https://doi.org/10.1142/9789812565662>.
- Madhavi Sastry, G., Adzhigirey, M., Day, T., Annabhimoju, R., & Sherman, W. (2013). Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer-Aided Molecular Design*, 27(3), 221–234. <https://doi.org/10.1007/s10822-013-9644-8>.
- Mikra, Chrysoula & Rossos, Georgios & Hadjikakou, Sotiris & Kourkoumelis, Nikolaos. (2017). Molecular Docking and Structure Activity Relationship Studies of NSAIDs. What do they Reveal about IC50?. *Letters in Drug Design & Discovery*. 14. 10.2174/1570180814666161207143231.
- Mody, V., Ho, J., Wills, S., Mawri, A., Lawson, L., Ebert, M. C. C. J. C., Fortin, G. M., Rayalam, S., & Taval, S. (2021). Identification of 3-chymotrypsin like protease (3CLPro) inhibitors as potential anti-SARS-CoV-2 agents. *Communications Biology*, 4(1). <https://doi.org/10.1038/s42003-020-01577-x>.

- Muñoz-Gutiérrez, C., Cáceres-Rojas, D., Adasme-Carreño, F., Palomo, I., Fuentes, E., & Caballero, J. (2017). Docking and quantitative structure–activity relationship of bi-cyclic heteroaromatic pyridazinone and pyrazolone derivatives as phosphodiesterase 3A (PDE3A) inhibitors. *PLOS ONE*, 12(12), e0189213. [doi:10.1371/journal.pone.0189213](https://doi.org/10.1371/journal.pone.0189213)
- Osipiuk, J., Azizi, S.A., Dvorkin, S. et al. Structure of papain-like protease from SARS-CoV-2 and its complexes with non-covalent inhibitors. *Nat Commun* 12, 743 (2021). <https://doi.org/10.1038/s41467-021-21060-3>
- Perlman, S., & Netland, J. (2009). Coronaviruses post-SARS: Update on replication and pathogenesis. *Nature Reviews Microbiology*, 7(6), 439–450. <https://doi.org/10.1038/nrmicro2147>.
- Pitsillou, E., Liang, J., Ververis, K., Lim, K. W., Hung, A., & Karagiannis, T. C. (2020). Identification of Small Molecule Inhibitors of the Deubiquitinating Activity of the SARS-CoV-2 Papain-Like Protease: in silico Molecular Docking Studies and in vitro Enzymatic Activity Assay. *Frontiers in Chemistry*, 8(December), 1–15. <https://doi.org/10.3389/fchem.2020.623971>.
- Qiu, W., Rutherford, S., Mao, A., & Chu, C. (2017). The Pandemic and its Impacts. *Health, Culture and Society*, 9, 1–11. <https://doi.org/10.5195/hcs.2017.22>.
- Ramírez, D., & Caballero, J. (2018). Is It Reliable to Take the Molecular Docking Top Scoring Position as the Best Solution without Considering Available Structural Data? *Molecules*, 23(5), 1038. [doi:10.3390/molecules23051038](https://doi.org/10.3390/molecules23051038)
- Rao, P., Patel, R., Shukla, A., Parmar, P., Rawal, R. M., Saraf, M., & Goswami, D. (2021). Identifying structural–functional analogue of GRL0617, the only well-established inhibitor for papain-like protease (PLpro) of SARS-CoV2 from the pool of fungal metabolites using docking and molecular dynamics simulation. *Molecular Diversity*, (0123456789). <https://doi.org/10.1007/s11030-021-10220-8>.
- Ratia, K., Pegan, S., Takayama, J., Sleeman, K., Coughlin, M., Baliji, S., ... Mesecar, A. D. (2008). A noncovalent class of papain-like protease/deubiquitinase inhibitors blocks SARS virus replication. *Proceedings of the National Academy of Sciences of the United States of America*, 105(42), 16119–16124. <https://doi.org/10.1073/pnas.0805240105>.

- Ratia, K., Saikatendu, K. S., Santarsiero, B. D., Barreto, N., Baker, S. C., Stevens, R. C., & Mesecar, A. D. (2006). Severe acute respiratory syndrome coronavirus papain-like-protease: Structure of a viral deubiquitinating enzyme. *Proceedings of the National Academy of Sciences of the United States of America*, 103(15), 5717–5722. <https://doi.org/10.1073/pnas.0510851103>.
- Velázquez-Libera, J. L., Caballero, J., Murillo-López, J. A., & de la Torre, A. F. (2019). Structural requirements of n-alpha-mercaptoacetyl dipeptide (Namdp) inhibitors of pseudomonas aeruginosa virulence factor lasb: 3d-qsar, molecular docking, and interaction fingerprint studies. *International Journal of Molecular Sciences*, 20(24), 1–18. <https://doi.org/10.3390/ijms20246133>.
- Velázquez-Libera, J. L., Durán-Verdugo, F., Valdés-Jiménez, A., Valdés-Jiménez, A., Núñez-Vivanco, G., & Caballero, J. (2020). LigRMSD: A web server for automatic structure matching and RMSD calculations among identical and similar compounds in protein-ligand docking. *Bioinformatics*, 36(9), 2912–2914. <https://doi.org/10.1093/bioinformatics/btaa018>.
- Velázquez-Libera, J. L., Rossino, G., Navarro-Retamal, C., Collina, S., & Caballero, J. (2019). Docking, Interaction Fingerprint, and Three-Dimensional Quantitative Structure–Activity Relationship (3D-QSAR) of Sigma1 Receptor Ligands, Analogs of the Neuroprotective Agent RC-33. *Frontiers in Chemistry*, 7(July), 1–20. <https://doi.org/10.3389/fchem.2019.00496>.
- Verma, R. P., & Hansch, C. (2005). An approach toward the problem of outliers in QSAR. *Bioorganic & Medicinal Chemistry*, 13(15), 4597–4621. [doi:10.1016/j.bmc.2005.05.002](https://doi.org/10.1016/j.bmc.2005.05.002).
- Verma, J., Khedkar, V., & Coutinho, E. (2010). 3D-QSAR in Drug Design - A Review. *Current Topics in Medicinal Chemistry*, 10(1), 95–115. <https://doi.org/10.2174/156802610790232260>.
- Taubenberger, J. K., & Morens, D. M. (2006). 1918 Influenza: the Mother of All Pandemics. *Emerging Infectious Diseases*, 12(1), 15–22. <https://doi.org/10.3201/eid1209.050979>.
- Wang, Y.-L., Wang, F., Shi, X.-X., Jia, C.-Y., Wu, F.-X., Hao, G.-F., & Yang, G.-F. (2020). Cloud 3D-QSAR: a web tool for the development of quantitative structure–activity

relationship models in drug discovery. *Briefings in Bioinformatics*, 00(September), 1–8. <https://doi.org/10.1093/bib/bbaa276>.

- Watanabe, T., Watanabe, S., Maher, E. A., Neumann, G., & Kawaoka, Y. (2014). Pandemic potential of avian influenza A (H7N9) viruses. *Trends in Microbiology*, 22(11), 623–631. <https://doi.org/10.1016/j.tim.2014.08.008>.
- Yang, H., Bartlam, M., & Rao, Z. (2006). Drug Design Targeting the Main Protease, the Achilles Heel of Coronaviruses. *Current Pharmaceutical Design*, 12(35), 4573–4590. <https://doi.org/10.2174/138161206779010369>.
- Zhao, L., Wang, W., Sedykh, A., & Zhu, H. (2017). Experimental Errors in QSAR Modeling Sets: What We Can Do and What We Cannot Do. *ACS omega*, 2(6), 2805–2812. <https://doi.org/10.1021/acsomega.7b00274>.