



Facultad de Ingeniería

Escuela de Ingeniería Civil en Bioinformática

---

# Identificación de patrones estructurales que participan en la función de la enzima PLpro de SARS-CoV y SARS-CoV2.

---

**Estudiante:**

Claudio Illanes Solis

**Profesor Tutor:**

Dr. Julio Caballero

**Co-Tutor:**

Dr. José Luis Velázquez-Libera

**Matricula:**

2017430017

**Profesora Informante:**

Dra. Ariela Vergara Jaque

## 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, 2023

## Resumen

Las infecciones zoonóticas por coronavirus (CoV) representan una gran amenaza para los humanos, ya que este tipo de virus posee la capacidad de adaptarse a nuevos entornos mediante la mutación y la recombinación genética. La pandemia de COVID-19 ha entregado indicios de la letalidad que puede alcanzar este tipo de agentes patógenos, impulsando así una gran cantidad de estudios e investigaciones orientadas a la obtención de diversas terapias, ya sea con fármacos o vacunas para combatir los efectos que provocan los coronavirus. Las principales investigaciones apuntan a atacar directamente al virus, buscando interrumpir pasos claves del mecanismo viral, como lo son la escisión de proteínas no estructurales (NSP) y la deubiquitinación. Estas etapas son mediadas por PLpro, que es una de las dos proteasas que se encargan de la replicación del coronavirus. A pesar de que se ha reportado una gran cantidad de compuestos inhibitorios de la PLpro, la información que se posee actualmente, de los mejores candidatos, es bastante limitada, ya que en su gran mayoría se desprende de los datos obtenidos de las estructuras cristalizadas. Para poder contribuir a enriquecer la información estructural existente, se utilizaron métodos de modelado y simulación molecular, que permitieron observar la evolución dinámica de complejos moleculares enfocados al estudio de las proteasas PLpro pertenecientes a SARS-CoV-1 y SARS-CoV-2. Ambas proteínas fueron analizadas frente a un conjunto de inhibidores diversos estructuralmente mediante análisis comparativos, como RMSD, RMSF, Interaction Fingerprint y Clustering, todo esto con el objetivo de encontrar patrones estructurales que muestran similitudes y diferencias entre las interacciones proteína-ligando. Se ha encontrado una estructura representativa de la proteasa, donde los residuos que aportan la similitud conformacional para SARS-CoV-1 y SARS-CoV-2 resultan ser Asp165/164, Pro249/248, Tyr265/264 y Thr302/301. Por otro lado, quienes aportan el grado de variabilidad estructural en el sitio de unión para la interacción con inhibidores son los residuos Tyr269/268, Gln270/269 y Leu163/162.

## Abstract

Zoonotic coronavirus (CoV) infections represent a great threat to humans, as this type of virus has the ability to adapt to new environments through mutation and genetic recombination. The COVID-19 pandemic has demonstrated the lethality that this type of pathogen can reach, thus promoting a large number of studies and research aimed at obtaining various therapies, either with drugs or vaccines to combat the effects caused by coronaviruses. The main research aims at directly attacking the virus, seeking to interrupt key steps of the viral mechanism, such as the cleavage of non-structural proteins (NSP) and deubiquitination. These steps are mediated by PLpro, which is one of the two proteases responsible for coronavirus replication. Although a large number of PLpro inhibitory compounds have been reported, the information currently available on the best candidates is rather limited, as it is mostly derived from data obtained from crystallized structures. To contribute to enrich the existing structural information, molecular modeling and simulation methods were used to observe the dynamic evolution of molecular complexes focused on the study of PLpro proteases belonging to SARS-CoV-1 and SARS-CoV-2. Both proteins were analyzed against a set of structurally diverse inhibitors by comparative analyses, such as RMSD, RMSF, interaction fingerprinting and clustering, with the aim of finding structural patterns that show similarities and differences between protein-ligand interactions. A representative protease structure has been found, where the residues providing conformational similarity for SARS-CoV-1 and SARS-CoV-2 are Asp165/164, Pro249/248, Tyr265/264 and Thr302/301. On the other hand, residues Tyr269/268, Gln270/269 and Leu163/162 provide the degree of structural variability in the binding site for interaction with inhibitors.

# Índice de contenidos

<b>1. Introducción</b>	<b>7</b>
1.1 Síndrome respiratorio agudo grave causado por coronavirus	7
1.2 Estructura y actividad de los coronavirus	8
1.3 Aspectos estructurales y actividad de la proteasa PLpro	11
1.4 Inhibidores de SARS-CoV-1-PLpro y SARS-CoV-2-PLpro	14
1.5 Alcance de esta tesis	18
<b>2. Hipótesis</b>	<b>19</b>
<b>3. Objetivos</b>	<b>19</b>
<b>4. Materiales y métodos</b>	<b>20</b>
4.1 Selección y preparación de inhibidores y estructura de la proteína	21
4.2 Docking molecular	23
4.2.1 LigRMSD	24
4.3 Dinámica molecular	24
4.4 Análisis post-dinámica molecular	27
4.4.1 Análisis RMSD	27
4.4.2 Análisis de RMSF	27
4.4.3 Interaction fingerprint	28
4.4.4 Clustering	28
<b>5. Resultados y discusión</b>	<b>30</b>
5.1 Docking molecular	30
5.1.1 Docking de 15G	31
5.1.2 Docking de 3K	32
5.1.3 Docking GRL0617	32
5.1.4 Docking de Aloidin A / Aloidin B	33
5.1.5 Docking de Sitagliptin	33
5.1.6 Docking de Daclatasvir	33
5.1.7 Docking de XR8-23	34
5.1.8 Docking de XR8-24	34
5.1.9 LigRMSD	35
5.2 Estabilidad y dinámica conformacional del complejo proteína-ligando	38
5.2.1 RMSD	38
5.2.2 RMSF	39
5.2.3 Interaction fingerprint	41
5.2.3.1 Puente salino	41
5.2.3.2 Enlace de hidrogeno	41
5.2.3.3 Hidrofobicas	42
5.2.3.3.1 Pi-stacking	42
5.2.3.3.2 VdW	42
5.2.4 Fase 1: Clustering de las dinámicas	44
5.2.5 Fase 2: Clustering de los centroides del backbone	46
5.2.6 Fase 3: Clustering de los centroides	47
5.2.6.1 Centroides de las cadenas laterales SARS-CoV-1	47
5.2.6.2 Centroides de las cadenas laterales SARS-CoV-2	48
5.2.6.3 Centroides de las cadenas laterales SARS-CoV-1/SARS-CoV-2	50
<b>6. Conclusiones</b>	<b>52</b>
<b>7. Referencias</b>	<b>53</b>

## Índice de figuras

<b>Figura 1.</b> Representación de las proteínas estructurales del coronavirus	9
<b>Figura 2.</b> Escisiones proteolíticas por PLpro y 3CLpro	10
<b>Figura 3.</b> Organización de los dominios estructurales de SARS-CoV PLpro	11
<b>Figura 4.</b> Representación del sitio activo de PLpro	12
<b>Figura 5.</b> Alineamiento de secuencia entre los sitios de unión PLpro SARS-CoV-1/SARS-CoV-2	13
<b>Figura 6.</b> Esquema de los subsitios de unión de PLpro	14
<b>Figura 7.</b> Representación 2D de Inhibidores de la PLpro y valor de IC50	15
<b>Figura 8.</b> Representación 2D de Inhibidores de la PLpro. (A) Aloin A. (B) Aloin B	16
<b>Figura 9.</b> Representación 2D de Inhibidores de la PLpro. (A) Sitagliptin. (B) Daclatasvir	17
<b>Figura 10.</b> Representación 2D del Inhibidor XR8-24 de la PLpro	17
<b>Figura 11.</b> Diagrama de la metodología	20
<b>Figura 12.</b> Representación esquemática del método de dinámica molecular acelerada	25
<b>Figura 13.</b> Flujo de trabajo general en Amber	26
<b>Figura 14.</b> Resultados para el protocolo de acoplamiento molecular	31
<b>Figura 15.</b> Comparación estructural entre las poses de acoplamiento molecular	35
<b>Figura 16.</b> Resultados del análisis RMSF	40
<b>Figura 17.</b> Resultados del clustering de las dinámicas	45
<b>Figura 18.</b> Clustering de los centroides del backbone	46
<b>Figura 19.</b> Centroides pertenecientes al clustering de la fase 3 para SARS-CoV-1	48
<b>Figura 20.</b> Centroides pertenecientes al proceso de clustering de la fase 3 para SARS-CoV-2.	49
<b>Figura 21.</b> Centroides pertenecientes al clustering de la fase 3 para SARS-CoV-1/SARS-CoV-2	51

## Índice de tablas

<b>Tabla 1.</b> Conjunto de compuestos seleccionados	22
<b>Tabla 2.</b> Definiciones de interacciones utilizadas para calcular los IFP en el análisis MD-IFP	28
<b>Tabla 3.</b> Análisis de LigRMSD de los compuestos utilizados con respecto a su referencia	37
<b>Tabla 4.</b> Resultados del análisis Interaction Fingerprint	43

# 1. Introducción

## 1.1 Síndrome respiratorio agudo grave causado por coronavirus

Una pandemia se define como una enfermedad epidémica que se extiende a muchos países, o que ataca a casi todos los individuos de una localidad o región (ASALE & RAE, s. f.); sin embargo, dicho significado no resulta ser un consenso para los medios de comunicación, organismos oficiales de la salud y publicaciones científicas, dado que resulta muy general y poco específico. Para solucionar este inconveniente se han establecido ciertas características claves para determinar una pandemia, las cuales son: amplia extensión geográfica, movimiento de enfermedades, altas tasas de ataque y explosividad, inmunidad mínima de la población, novedad, infecciosidad, contagio y gravedad (Morens et al., 2009).

A lo largo de la historia, han surgido diversas pandemias, que han causado una gran cantidad de muertes humanas. Algunas de las más trascendentales corresponden a: la peste negra (Benedictow, 2011), la gripe española (Gallo, 2020), el VIH-sida (Cardona-Arias & Higueta-Gutiérrez, 2014) y el SARS, la cual resulta ser la primera pandemia del siglo XXI (Cherry & Krogstad, 2004).

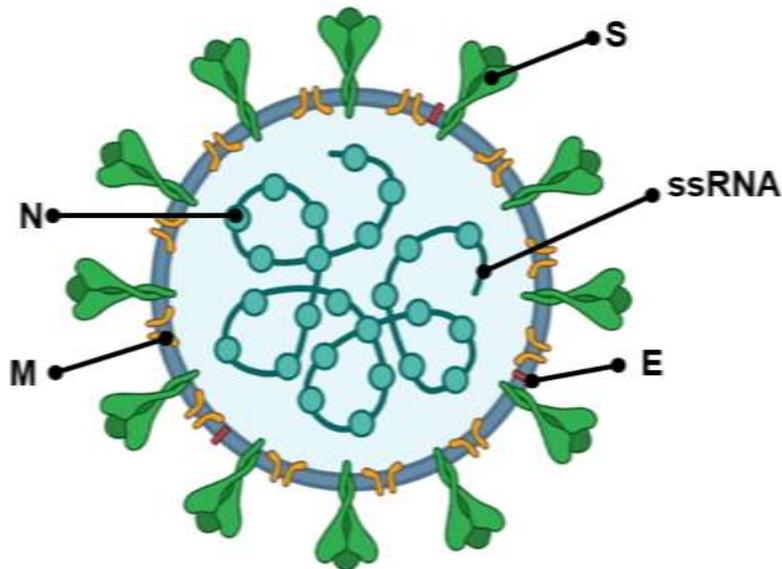
El Síndrome Respiratorio Agudo Severo (SARS, del inglés Severe Acute Respiratory Syndrome ) es una enfermedad que aparece el año 2002, en la provincia de Guangdong, China. Este síndrome se propagó por 29 países generando 8.422 casos y 916 muertes (Cherry & Krogstad, 2004). Para poder diagnosticar dicho síndrome, se establecieron ciertos parámetros por parte de los Centros para el Control y la prevención de Enfermedades (CDC, del inglés Centers for Disease Control and Prevention), incluyendo criterios clínicos como: tener temperatura superior a 38°C y evidencia de alguna enfermedad respiratoria, criterios epidemiológicos, donde se incluye el viaje a una zona con brotes de SARS, contacto estrecho con algún caso positivo de SARS y por último se encuentran los criterios de laboratorio, tales como la detección de anticuerpos contra el SARS-CoV o la detección de ARN del SARS-CoV por PCR. Además, dicho síndrome presentó diversos factores que provocaron una alerta mundial por parte de la Organización Mundial de la Salud (OMS), entre los más relevantes se encuentran que la enfermedad se había extendido desde su foco inicial en Asia a América del Norte y Europa, además ninguno de los diferentes antibióticos y antivirales que se habían probado empíricamente parecía tener efecto para combatir dicha enfermedad (Drosten et al., 2003).

Se ha determinado que el factor etiológico perteneciente al SARS corresponde a un coronavirus (CoV), el cual fue aislado mediante estudios en muestras clínicas de pacientes con SARS, utilizando cultivos celulares y técnicas moleculares (Drosten et al., 2003). Las infecciones zoonóticas por coronavirus (CoV) representan grandes amenazas para los humanos. Sin ir más lejos, el virus del Síndrome Respiratorio Agudo Severo (SARS-CoV-1) (Cheng et al., 2007) infectó a 8.000 personas con una tasa de mortalidad de ~10% (Li, 2016). En 2012, el Síndrome Respiratorio de Oriente Medio (MERS-CoV) (de Groot et al., 2013) se esparció a más de 1.700 personas, con una tasa de mortalidad de ~36% (Li, 2016), siendo un primer indicio de lo que pueden causar este tipo de enfermedades. Por último, en el año 2019 el SARS-CoV-2 (Gorbalenya et al., 2020) originó la pandemia denominada COVID-19 con cifras cercanas a 8 millones de infectados y más de 400.000 muertos en todo el mundo, según cifras reportadas hasta junio de 2020 (El Zowalaty & Järhult, 2020).

## **1.2 Estructura y actividad de los coronavirus**

Los coronavirus son una subfamilia perteneciente a la familia coronaviridae, que se conforma de cuatro géneros: Alphacoronavirus, Betacoronavirus, Gammacoronavirus y Deltacoronavirus, los cuales pueden infectar a una diversa cantidad de aves y mamíferos, siendo los géneros Alpha y Beta capaces de afectar a humanos. La mayoría de las cepas de CoV, presentan un rango estrecho de huéspedes, no obstante, los coronavirus zoonóticos, son más propensos a migrar a nuevas especies (Sheahan et al., 2017). Estos agentes patógenos, poseen la capacidad de adaptarse a nuevos entornos mediante la mutación y la recombinación genética, por lo que el rango de huéspedes conocidos puede ampliarse considerablemente. Dicho esto, los coronavirus significan una constante amenaza para la salud, por lo que comprender su virología y controlar su propagación tiene implicaciones importantes para la salud mundial y la estabilidad económica (Li, 2016).

Estructuralmente, los CoV son de forma esférica, con un diámetro que varía entre los 60 a 140 nm, con picos distintivos de aproximadamente 8 a 12 nm de longitud, los cuales, al ser observados mediante microscopía electrónica, se asemejan a una corona solar, lo que dio origen al nombre “coronavirus” (Zheng, 2020). En cuanto al genoma del coronavirus, se encuentra conformado por una molécula de ARN con cadena positiva, de hebra única, con un tamaño que varía entre 26 y 32 kb, la cual contiene seis marcos de lectura abiertos (ORF) (Perlman & Netland, 2009). Estudios recientes han demostrado que el SARS-CoV-2, comparte un ~89 % de similitud de secuencia con SARS-CoV-1, compartiendo la región 5' no traducida, 16 proteínas no estructurales (complejo ORF1a y ORF1b), proteínas estructurales (Figura 1) y otras proteínas presentes hacia el extremo 3' (Mody et al., 2021).

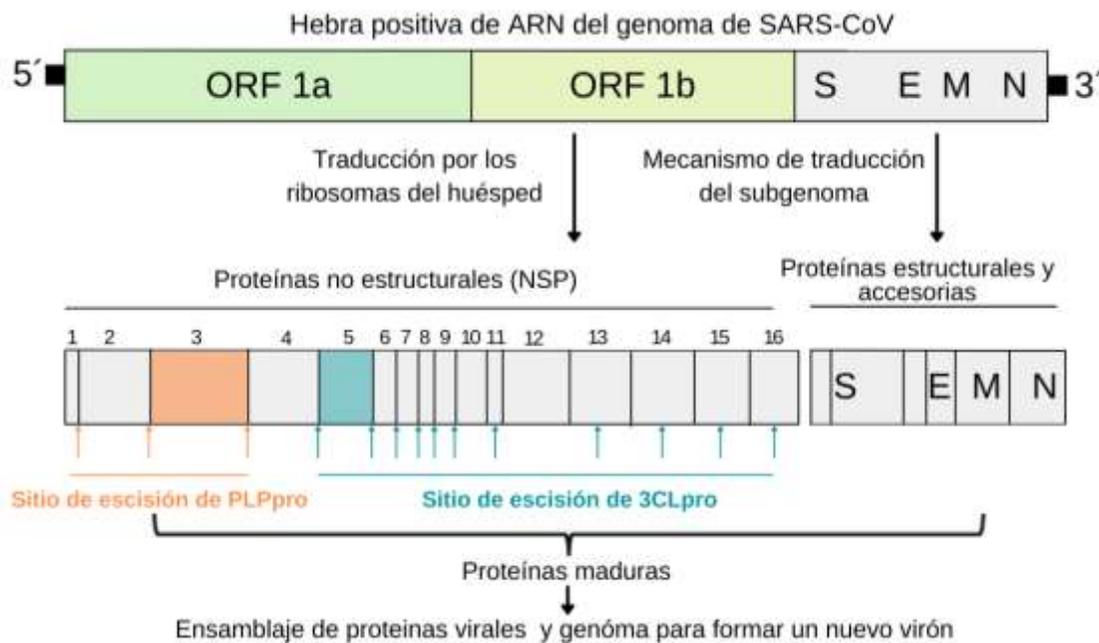


**Figura 1:** Representación de las proteínas estructurales del coronavirus, (S) proteína spike, (E) proteína de envoltura, (M) proteína de membrana y (N) proteína de nucleocápside.

El virus comienza su ciclo de vida en el momento en que la proteína spike (S) se une a su receptor en las células del huésped, con la finalidad de liberar el genoma viral en el citoplasma de la célula infectada. El genoma viral (ssRNA) es traducido mediante los ribosomas del huésped a una cadena polipeptídica (PP), la cual una vez generada es escindida por dos proteasas, la proteasa similar a la papaína (PLpro) y la proteasa similar a 3-quimiotripsina (3CLpro). Ambas proteasas son codificadas por el genoma viral para producir proteínas no estructurales (NSP), las cuales son necesarias para la replicación viral del virus (Mody et al., 2021).

Debido a la importancia funcional que posee 3CLpro en el ciclo viral del coronavirus, se ha propuesto como un objetivo clave para el diseño de fármacos basados en estructuras contra el SARS (Hilgenfeld et al., 2006). Además, estudios sobre los aspectos estructurales y mecánicos de dicha proteasa han brindado múltiples vías para la detección de alto rendimiento y el diseño basado en la estructura de antivirales dirigidos al sitio activo de 3CLpro (Mody et al., 2021). Esta proteasa también identificada como Nsp5, interviene en la escisión de 11 sitios para liberar Nsp4-Nsp16, que es indispensable para la replicación y la infección por CoV (Figura 2) (Huang et al., 2004).

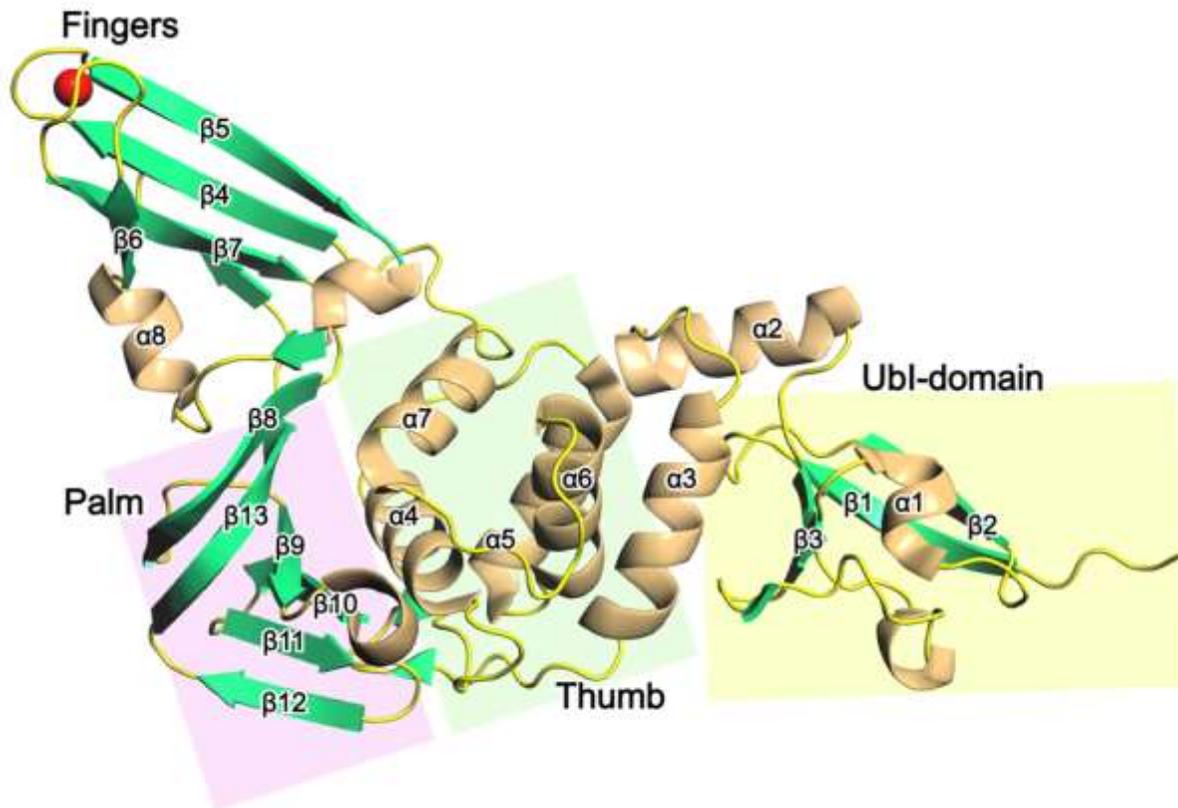
De igual manera, la proteasa PLpro ha sido sugerida como un objetivo clave, para impedir el ciclo viral del coronavirus. Esta enzima forma parte de Nsp3 (Figura 2) y tiene la función de realizar reacciones de escisión para liberar Nsp1, Nsp2 y Nsp3, las cuales intervienen de manera directa en la replicación viral y en los mecanismos antagónicos para evadir el sistema inmunológico innato del huésped (Harcourt et al., 2004). Específicamente, PLpro elimina la ubiquitina (Ub) y las proteínas similares a la ubiquitina (UbL), como ISG15. Esta proteasa reconoce y escinde la secuencia de muchas UbL en su extremo C-terminal, actuando como una deubiquitinasa (DUB) hacia las proteínas modificadas con Ub y UbL. La ubiquitinación tiene un papel importante en la degradación de las proteínas extrañas (como las proteínas virales) tras la infección, para prevenir la propagación viral. Por lo tanto, la actividad de deubiquitinación de PLpro, conduce a la interrupción de la respuesta inmune antiviral del huésped (Lewis et al., 2022). Por consiguiente, modular PLpro es una estrategia atractiva para inhibir la replicación viral y prevenir la interrupción de la respuesta inmune del huésped a la infección viral (Shen et al., 2022).



**Figura 2:** Escisiones proteolíticas por PLpro y 3CLpro. El material genético del SARS-CoV hace ingreso a la célula huésped y utiliza su ribosoma para traducir el ARN en 16 proteínas no estructurales, tras la escisión auto proteolítica por parte de las enzimas PLpro y 3CLpro. Las proteínas estructurales (S, E, M y N) son traducidas mediante el mecanismo de traducción de subgenoma (Mody et al., 2021). Todas las proteínas maduras se ensamblan junto con el genoma de ARN monocatenario de sentido positivo para formar un nuevo virión. Las flechas de color indican los sitios de escisión de cada proteasa. Los sitios coloreados naranja (NSP3) y celeste (NSP5) representan las posiciones donde se encuentran PLpro y 3CLpro, respectivamente.

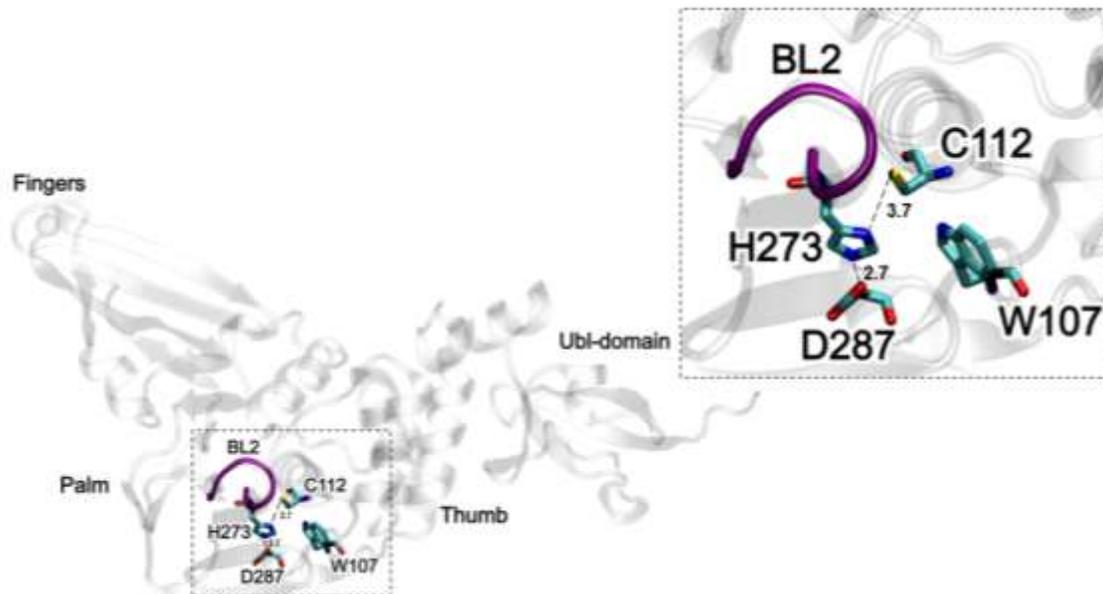
### 1.3 Aspectos estructurales y actividad de la proteasa PLpro

La estructura molecular del dominio catalítico PLpro, se basa en una arquitectura de pulgar-palma-dedos de la mano derecha (Figura 3), la cual está flanqueada en la región N-terminal por un dominio similar a la ubiquitina (Ubl) de función desconocida (Báez-Santos et al., 2015). Los primeros 62 aminoácidos forman un dominio N-terminal independiente, denominado dominio Ubl, el cual se encuentra distante de los otros tres dominios y adopta un pliegue de agarre  $\beta$  similar a la ubiquitina. El dominio del pulgar está formado por cuatro hélices ( $\alpha 4-7$ ) y la palma se conforma por una lámina  $\beta$  de seis hebras ( $\beta 8-13$ ) que se inclina hacia el sitio activo, el cual se encuentra en una hendidura expuesta al solvente entre los dominios del pulgar y la palma. Una hoja  $\beta$  antiparalela, retorcida y de cuatro hebras ( $\beta 4-7$ ), constituye el dominio de los “dedos”. Dos horquillas  $\beta$ , en la región de las yemas de los dedos, contienen cuatro residuos de cisteína, que coordinan un ion zinc con geometría tetraédrica (Ratia et al., 2006).



**Figura 3:** Organización de los dominios estructurales de SARS-CoV PLpro. Los recuadros de colores representan los tres dominios, en primera instancia el dominio de ubiquitina (Ubl) se muestra de color amarillo, luego la palma y el pulgar se presentan de color rosado y verde, respectivamente. Las hélices  $\alpha$  (naranja) y las láminas  $\beta$  (verde) se encuentran enumeradas. En el extremo superior izquierdo se muestra el átomo de zinc (rojo) presente en la estructura.

En cuanto al sitio activo de PLpro, posee una tríada catalítica, compuesta por Cys–His–Asp, que se encuentra ubicada en la interfaz de los subdominios del pulgar y la palma (Ratia et al., 2006). La cisteína catalítica está ubicada en el extremo N terminal de la alfa hélice,  $\alpha 4$ , en el dominio del pulgar. El átomo de azufre de la cadena lateral de Cys está posicionado alrededor de  $\sim 3,7$  Å del átomo del nitrógeno  $\delta$  de la histidina catalítica, que está ubicada en el dominio de la palma y adyacente al bucle BL2 (giro  $\beta$ ). Uno de los átomos de oxígeno de la cadena lateral del ácido aspártico catalítico, forma un enlace de hidrógeno con el nitrógeno  $\epsilon$  de la histidina catalítica, en el dominio de la palma. Cercano a la tríada catalítica, se encuentra un triptófano, que ha sido identificado como esencial para la función enzimática de SARS-CoV-PLpro. La cadena lateral del triptófano se encuentra dentro del agujero de oxianión y se propone que el nitrógeno del anillo participa en la estabilización de los intermedios tetraédricos cargados negativamente producidos durante la catálisis de la proteína (Figura 4) (Báez-Santos et al., 2015).



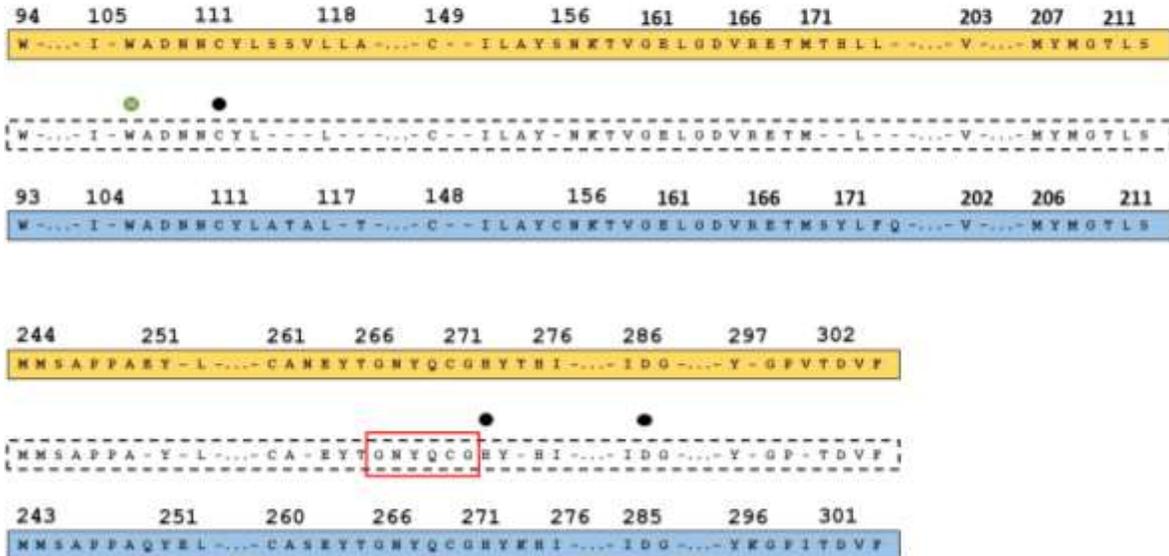
**Figura 4:** Representación del sitio activo de PLpro (SARS-CoV-2, PDB: 2FE8). Los residuos de la tríada catalítica C112, H273 y D287 y el residuo estabilizador del agujero de oxianión W107 han sido coloreados según sus átomos y el bucle BL2 de color morado.

En particular, SARS-CoV-2-PLpro y SARS-CoV-1-PLpro solo difieren en 54 residuos, presentando un alto porcentaje de similitud en el sitio de unión (Figura 5). Además del sitio activo, conformado por la tríada catalítica, la PLpro posee cinco subsitios de reconocimiento a sustratos (S1-S5). Entre estos subsitios, S3 es el único que se encuentra parcialmente expuesto a solvente, ya que el resto de subsitios se ubican hacia el interior de la estructura. En cuanto a S2 y S4, cabe mencionar que son los sitios de unión para los inhibidores competitivos tanto para SARS-CoV-2 como para SARS-CoV-1, debido a que presentan una arquitectura similar, lo que sugiere posibles semejanzas en las interacciones con sustratos (Amin et al., 2021; Bagherzadeh et al., 2020).

## Identidad 73/90 (81%)

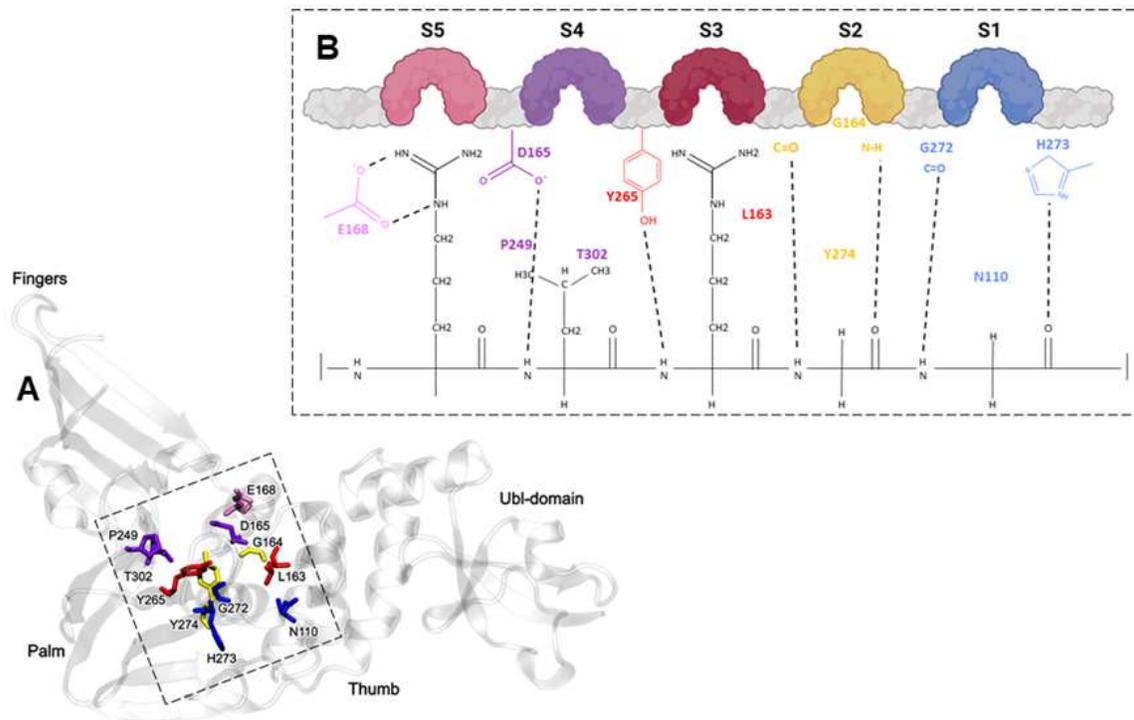
SARS-CoV-1 PLpro

SARS-CoV-2 PLpro



**Figura 5:** Alineamiento de secuencia entre los sitios de unión PLpro SARS-CoV-1 y SARS-CoV-2. Las secuencias alineadas se componen de los residuos a una distancia de 15 Å de sus respectivos ligandos, presentes en los cristales recuperados de PDB, 4OW0 (SARS-CoV-1-PLpro) y 7LBS (SARS-CoV-2-PLpro). La secuencia representada en blanco y líneas punteadas corresponde a los residuos que presentan similitud entre ambos sitios de unión. Los círculos negros muestran la triada catalítica. Los círculos de color verde representan al triptófano identificado como esencial para la función enzimática y el recuadro de color rojo muestra el bucle BL2.

Los subsitios mencionados anteriormente, han sido reportados en estudios comparativos entre PLpro y enzimas deubiquitinantes humanas, donde se han dilucidado requisitos estructurales para el reconocimiento de ubiquitina por parte de la proteasa. Se ha establecido que las interacciones entre los subsitios de PLpro y ubiquitina (Figura 6), se forman luego de la apertura del bucle BL2, que ocurre cuando se une el extremo C-terminal de ubiquitina y el sitio activo de la proteasa. La mayor parte de las interacciones involucran a 5 residuos de ubiquitina (RLRGG) (Ratia et al., 2014).

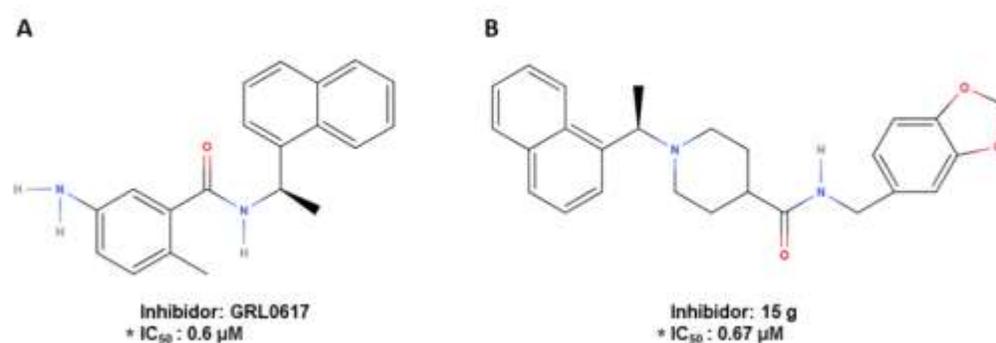


**Figura 6:** Esquema de los subsitios de unión de PLpro (SARS-CoV-1, PDB: 7LFU). **(A)** Residuos claves presentes en los subsitios S5 (Rosa), S4 (Purpura), S3 (Rojo), S2 (Amarillo) y S1 (Azul). **(B)** Se representan los bolsillos S5, S4, S3, S2 y S1 con varios residuos claves involucrados en el enlace de hidrógeno del sustrato RLRGG (terminal C de ubiquitina).

#### 1.4 Inhibidores de SARS-CoV-1-PLpro y SARS-CoV-2-PLpro

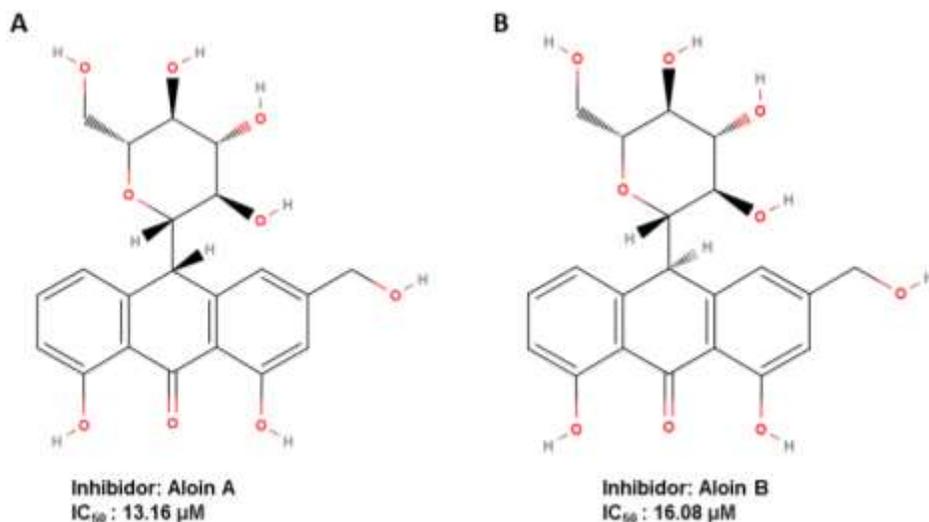
A la fecha, se han reportado y sintetizado distintas series de inhibidores no covalentes para bloquear la acción de PLpro (Shen et al., 2022). Tales moléculas han sido obtenidas mediante el uso de métodos orientados al diseño de fármacos, como el cribado de alto rendimiento (HTS), para interactuar con los subsitios S3, S4, S5 y el bucle BL2 (Ratia et al., 2008).

Entre los inhibidores más relevantes reportados, se encuentra el compuesto GRL0617 (Figura 7A) (Ratia et al., 2008), el cual se obtuvo luego de evaluar una librería con 50.080 compuestos diversos estructuralmente y una posterior optimización sintética. Dicho inhibidor no mostró citotoxicidad asociada y logró inhibir a PLpro uniéndose en los subsitios S4-S3 de la enzima, provocando un cierre de bucle que detiene la catálisis en el sitio activo. Dentro de la misma investigación, se encontró otro compuesto de interés, el cual fue potenciado mediante sustituciones estructurales, logrando así la creación del inhibidor 15g (Figura 7B). Estudios cinéticos y bioquímicos adicionales demostraron que 15g es un potente inhibidor competitivo no covalente de PLpro y que, adicionalmente, presenta bajos valores citotóxicos (Ghosh et al., 2010).



**Figura 7:** Representación 2D de Inhibidores de la PLpro y valor de IC<sub>50</sub>. **(A)** GRL0617. **(B)** 15 g. (\*) El valor de IC<sub>50</sub> indica la cantidad (μM) de una sustancia necesaria para inhibir, un proceso o componente biológico determinado, en un 50 %.

Años más tarde, se demostró que varios productos de enjuague bucal disponibles en el mercado son efectivos para inactivar el SARS-CoV-2 en un entorno *in vitro*. En base a esto, se ha evaluado la actividad inhibitoria enzimática *in vitro* de los principios activos de los principales enjuagues bucales, donde los isómeros de Aloin (A y B) (Figura 8) consiguieron inhibir más del 70% de la actividad de PLpro, por lo que se sugiere que estas moléculas muestran una capacidad inhibitoria específica frente al SARS-CoV-2-PLpro (Lewis et al., 2022). Mediante modelado estructural computacional, se reveló que ambos isómeros establecen enlaces de hidrógeno con Tyr 268 de PLpro (SARS-CoV-2). Estas interacciones son fundamentales para la actividad proteolítica de la proteasa, además, se predijo que ambos inhibidores (Aloin A y B) presentan una fuerte interacción con Glu167 (SARS-CoV-2), el cual es un residuo clave para la actividad de deubiquitinación de PLpro.

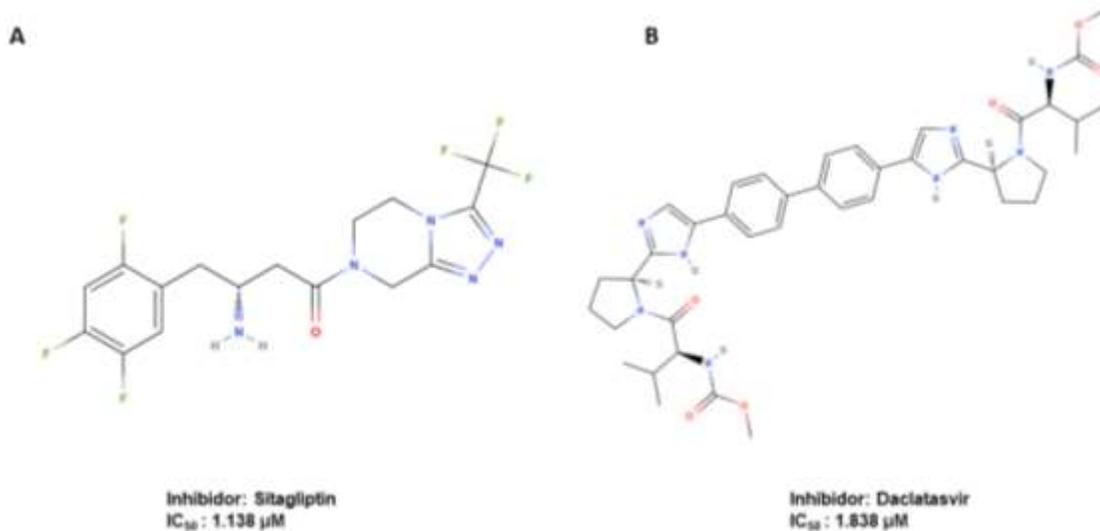


**Figura 8:** Representación 2D de Inhibidores de la PLpro. **(A)** Aloin A. **(B)** Aloin B.

Por otra parte, un estudio realizado en personas hospitalizadas con COVID-19, que padecían diabetes tipo II y que en su tratamiento estándar se les suministraba sitagliptin (Figura 9A), mostró que la tasa de mortalidad en dichos pacientes se redujo considerablemente, lo que sugirió que sitagliptin podría ser eficaz contra el SARS-CoV-2. Mediante técnicas de acoplamiento molecular, se logró determinar que sitagliptin se une al sitio activo de PLpro, bloqueando la entrada a la tríada catalítica (Cys111, His272 y Asp286 (SARS-CoV-2), demostrando que este compuesto logra inhibir efectivamente la actividad de PLpro y disminuir la replicación de SARS-CoV-2.

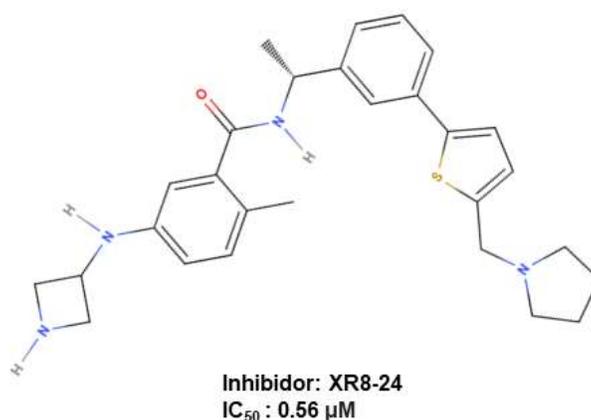
Adicionalmente, daclatasvir (Figura 9B), un compuesto utilizado para inhibir el virus que causa la propagación de la hepatitis C, fue sometido a diversos ensayos en pacientes que estaban infectados por COVID. En estos pacientes se observó que a quienes se les proporcionaba daclatasvir presentaban cifras significativamente reducidas en mortalidad y tiempos de recuperación, en comparación a pacientes a los que no se les suministraba el compuesto. Al igual que el compuesto anterior (sitagliptin), por medio de técnicas de acoplamiento molecular, se observó que daclatasvir tiene bajas afinidades de unión al sitio activo de PLpro debido a su gran tamaño, sugiriendo que el efecto inhibitorio se debe al unirse a un sitio alostérico en lugar del sitio activo de la enzima.

Cabe destacar que ambos inhibidores, sitagliptin y daclatasvir, se encuentran aprobados por la FDA (Administración de Alimentos y Medicamentos de los Estados Unidos) y se reportó que ambos inhibieron la propagación del virus en un 75% y 70%, respectivamente. (Narayanan et al., 2022).



**Figura 9:** Representación 2D de Inhibidores de la PLpro. **(A)** Sitagliptin. **(B)** Daclatasvir.

Por último, una investigación de Shen y col. (2022) ha reportado un nuevo y potente inhibidor denominado, XR8-24 (Figura 10), el cual no presenta citotoxicidad en concentraciones menores a 100 μM. Se pudieron observar dos sitios de unión ligando-proteína, uno incluyendo el residuo Glu167 (SARS-CoV-2), que ejerce un importante rol en la deubiquitinación; y un segundo sitio novedoso que se denominó "surco BL2", el cual se sitúa entre las hebras β8 y β9, adyacente al bucle BL2. Sumado a esto, se identificó una interacción de enlace de hidrógeno entre XR8-24 y el residuo Tyr264 (SARS-CoV-2), lo que puede explicar la potencia de inhibición superior de XR8-24, con respecto al resto de compuestos reportados en el estudio (Shen et al., 2022).



**Figura 10:** Representación 2D del Inhibidor XR8-24 de la PLpro.

## 1.5 Alcance de esta tesis

A través de la presente tesis se identificaron patrones estructurales entre sistemas proteína-ligando, modelados a partir de las estructuras cristalinas de PLpro pertenecientes a SARS-CoV-1 y SARS-CoV-2, contra un conjunto de inhibidores que han sido seleccionados utilizando como parámetros la variabilidad estructural y el valor de  $IC_{50}$ . Para lograr este objetivo, se empleó un protocolo compuesto por tres etapas sistemáticas. La primera de estas corresponde a la técnica de acoplamiento molecular, la cual proporcionó diversas conformaciones asociadas al conjunto de ligandos preparados. A partir de esto, se escogió las mejores poses para cada uno de los inhibidores, considerando la puntuación asignada en relación con las interacciones reportadas presentes en los cristales que poseen algunos de los compuestos en estudio cocrystalizados (Tabla 1). Posterior a esto, los complejos obtenidos fueron sometidos a dinámica molecular acelerada, con la finalidad de ampliar el muestreo conformacional, para cada uno de los casos. Por último, se evaluaron las trayectorias obtenidas mediante análisis comparativos: RMSD, RMSF, Interaction Fingerprint y Clustering. El contemplar cada uno de los resultados obtenidos, de manera conjunta, permitió observar similitudes y/o diferencias entre la respuesta de ambas proteínas en presencia de los inhibidores estudiados.

Actualmente, la información disponible de los inhibidores tanto de SARS-CoV-1-PLpro, como de SARS-CoV-2-PLpro, se desprende, casi en su totalidad, de las estructuras cristalinas reportadas. Sin embargo, mediante el uso de técnicas de acoplamiento molecular, simulación molecular y análisis comparativo, es posible proporcionar nueva información estructural que colabore a profundizar el conocimiento existente. Por lo tanto, el considerar la información entregada en este estudio, puede ser clave para afrontar nuevas enfermedades causadas por coronavirus.

## **2. Hipótesis**

La proteasa PLpro de los virus SARS-CoV-1 y SARS-CoV-2, presenta patrones estructurales similares, asociados a los sitios de unión frente a diversos inhibidores.

## **3. Objetivos**

- Modelar los sistemas entre las PLpro (SARS-CoV-1 y SARS-CoV-2) y los inhibidores seleccionados.
- Evaluar la estabilidad estructural de los complejos proteína-ligando modelados mediante dinámicas moleculares.
- Efectuar análisis comparativos entre los sistemas simulados, para encontrar patrones estructurales que presenten similitudes y diferencias entre las interacciones proteína-ligando.

## 4. Materiales y métodos

Para poder desarrollar los objetivos que se han propuesto, se siguieron los pasos que se muestran en la figura 11. En este estudio se pretendió encontrar patrones estructurales entre las proteasas PLpro de los virus SARS-CoV-1 y SARS-CoV-2 que determinen su asociación con los inhibidores presentes en la tabla 1. Entre las tareas que se realizaron se encuentra el acoplamiento molecular que sirvió para evaluar el modo de unión de los inhibidores con las proteínas. Posteriormente, se llevaron a cabo las dinámicas moleculares, con el propósito de minimizar y evaluar la estabilidad estructural en cada uno de los complejos construidos. Por último, se analizaron las trayectorias de los sistemas simulados, mediante análisis RMSD, RMSF, Interaction Fingerprint y Clustering.

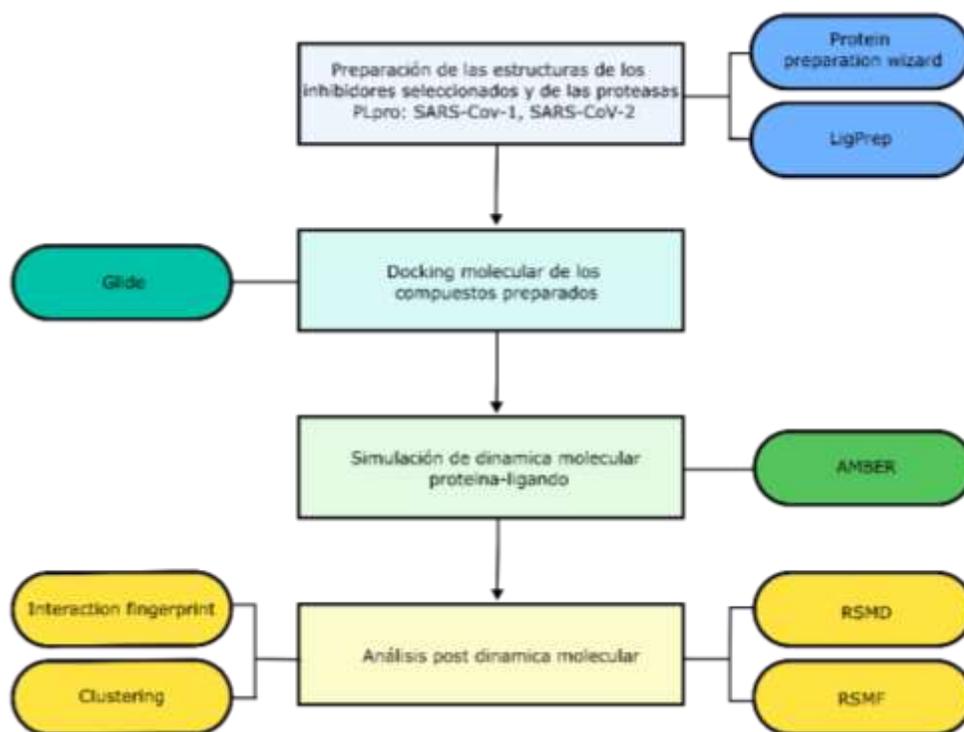
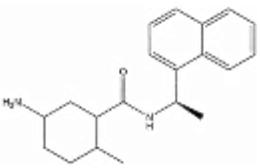
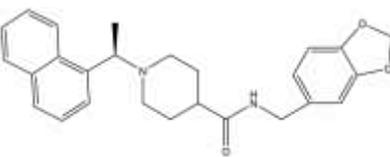
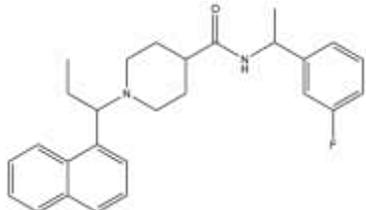
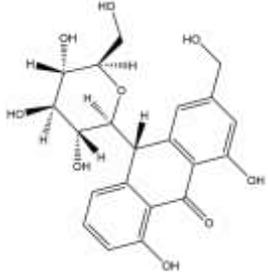
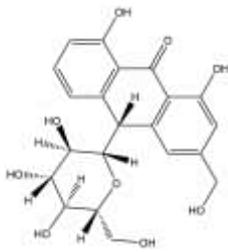
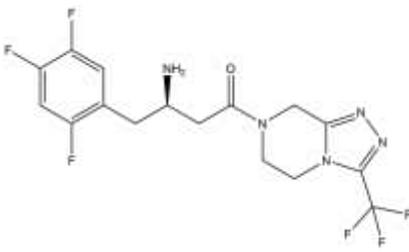
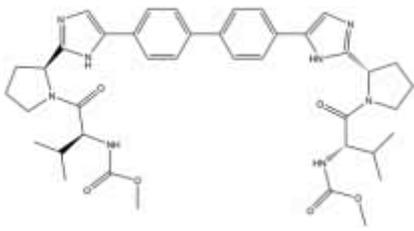
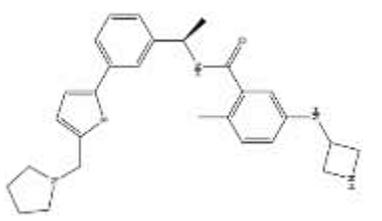
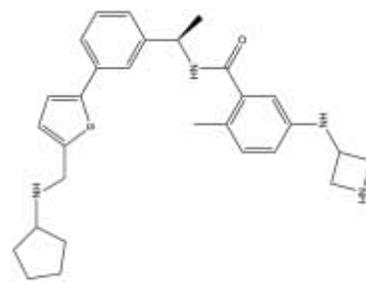


Figura 11: Diagrama de la metodología.

#### 4.1 Selección y preparación de inhibidores y estructura de la proteína

Para el estudio de los complejos proteína-ligando, se recopilaron nueve inhibidores que presentan diversidad estructural y han sido reportados como relevantes contra SARS-CoV-1-PLpro y SARS-CoV-2-PLpro (Tabla 1). Las estructuras de los compuestos fueron dibujadas utilizando el editor molecular de *Maestro* (Maestro, Schrödinger, LLC, New York, NY, 2021). Posteriormente, se realizó el pre-procesamiento de los ligandos mediante la herramienta *LigPrep* (LigPrep, Schrödinger, LLC, New York, NY, 2021). El estado de protonación que presento cada una de las estructuras a un pH fisiológico se estimó utilizando *Epik* (Epik, Schrödinger, LLC, New York, NY, 2021). De este modo, se obtuvieron las estructuras tridimensionales (3D) óptimas de los compuestos para evaluar su asociación con las proteasas PLpro de los virus SARS-CoV-1 y SARS-CoV-2.

Nombre	Estructura	IC50 (µM)	Cristal/PLpro	Referencia
GRL0617		<b>0.6</b>	Reportado para PLpro-SARS-CoV-2 (7CJM)	(Fu et al., 2021)
15 g		<b>0.67</b>	Reportado para PLpro-SARS-CoV-1 (3MJ5)	(Ghosh et al., 2010)
3k		<b>0.15</b>	Reportado para PLpro-SARS-CoV-1 (4OW0)	(Báez-Santos et al., 2015)
Aloin A		<b>13.16</b>	-	(Lewis et al., 2022)

Aloin B		16.08	-	(Lewis et al., 2022)
Sitagliptin		1.138	-	(Narayanan et al., 2022)
Daclatasvir		1.838	-	(Narayanan et al., 2022)
XR8-24		0.56	Reportado para PLpro-SARS-CoV-2 (7LBS)	(Shen et al., 2022)
XR8-23		0.39	-	(Shen et al., 2022)

**Tabla 1:** Conjunto de compuestos seleccionados.

Las estructuras de PLpro de SARS-CoV-1 y SARS-CoV-2 se obtuvieron desde los cristales correspondientes a cada uno de los inhibidores en cuestión, de la base de datos de proteínas PDB (Protein Data Bank) (Berman et al., 2000). Las proteasas, se prepararon para el acoplamiento molecular utilizando la herramienta Protein Preparation Wizard (Madhavi Sastry et al., 2013) de Maestro. Este complemento posee un protocolo para la optimización de proteínas. En primer lugar, se encuentra el preprocesamiento de la proteína, donde se añadieron los átomos de hidrógeno, se corrigieron las orientaciones que presentan los residuos que conforman la proteína, se asignaron cargas y se eliminaron moléculas no deseadas que se adhirieron a la estructura de las proteasas durante la cristalografía. Por último, las estructuras se optimizaron y minimizaron con los parámetros predeterminados bajo el campo de fuerza OPLS3.

## 4.2 Docking Molecular

El procedimiento del acoplamiento molecular (Docking) se realizó utilizando *Glide* de la suite de *Schrödinger* (Friesner et al., 2006), permitiendo modelar los sistemas de interacción proteína-ligando para el conjunto de inhibidores y las proteínas PLpro pertenecientes a SARS-CoV-1 y SARS-CoV-2. *Glide* utiliza una serie de filtros para buscar las posibles poses y orientaciones del ligando dentro del sitio de unión del receptor, utilizando una grilla de un tamaño definido previamente. Este software permite una búsqueda sistemática completa del espacio conformacional, orientacional y posicional del ligando antes de evaluar las interacciones de energía entre el ligando y la proteína (Friesner et al., 2004).

*Glide* ofrece una gama completa de algoritmos que combinan velocidad de búsqueda frente a precisión, incluyendo cribado virtual de alto rendimiento (HTVS), precisión estándar (SP) y precisión extra (XP). Para este estudio, los compuestos se sometieron inicialmente al algoritmo SP, que permite la identificación de ligandos que tienen una tendencia razonable a unirse, y luego se aplicó el método XP, el cual logra la eliminación de falsos positivos mediante un muestreo más extenso y una puntuación avanzada. De manera general, el utilizar ambos métodos aportaron mayor robustez en el proceso. De todas las conformaciones que se obtuvieron, se escogió la mejor pose para cada ligando mediante dos criterios. El primero fue la puntuación que entrega *Glide* (Glide Score), la cual representa la afinidad de unión estimada entre el inhibidor y la proteína objetivo. En adición al primer criterio, se consideraron las conformaciones que presentan la mayor cantidad de interacciones con los residuos presentes en los sitios de unión previamente reportados de ambas PLpro.

Durante el proceso de docking, los ligandos se acoplaron a la proteína mediante una grilla con tamaño 20 Å × 20 Å × 20 Å (Pitsillou et al., 2020). Además, se consideró la posición de los ligandos provenientes de cristales en el sitio de unión en la proteína, se determinó la ubicación de la grilla utilizando la configuración predeterminada “*Receptor Grid Generation*” (Halgren et al., 2004). Por último, se configuró la precisión del acoplamiento como “precisión extra (XP)”, el muestreo de ligandos como “flexible” y se definieron un total de 10 poses por ligando utilizado. 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 (Friesner et al., 2004). En cuanto a la selección de la mejor pose, se determinó por medio de la puntuación “Emodel” en conjunto con la estimación GlideScore que clasifica los compuestos en base a su unión al receptor.

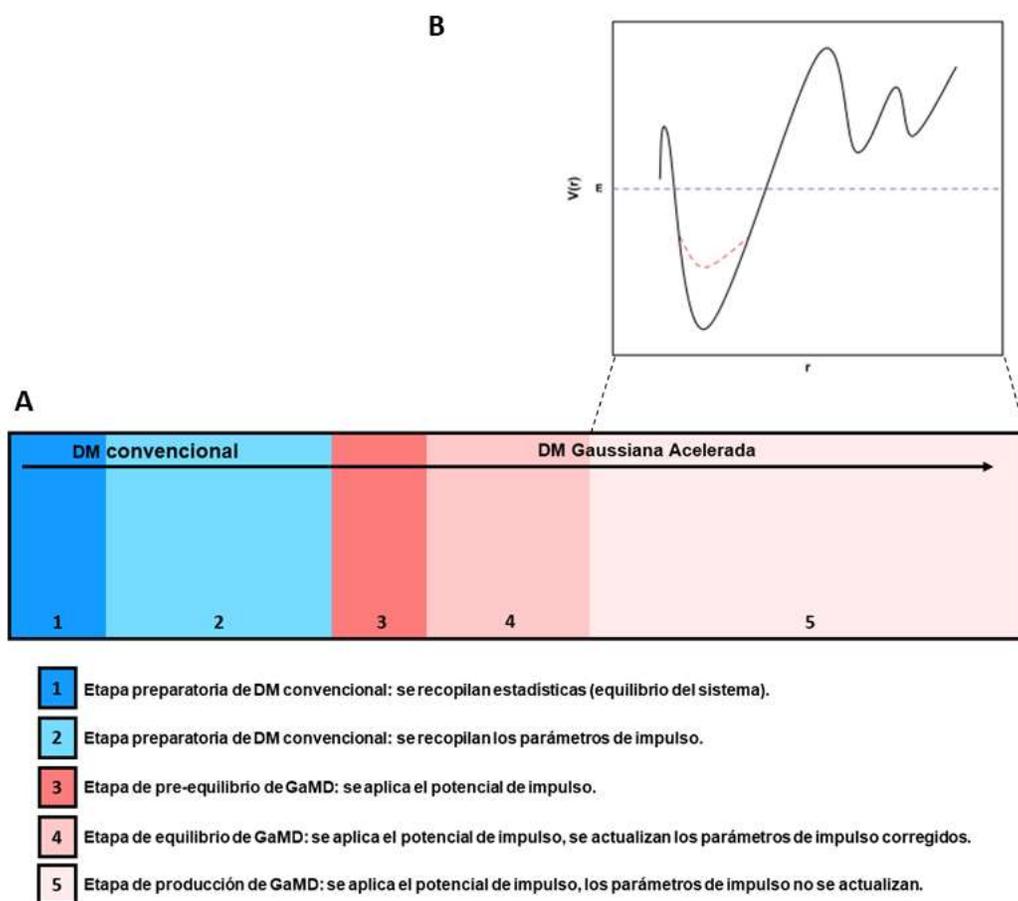
### 4.2.1 LigRMSD

Se utilizó el software LigRMSD (Velázquez-Libera et al., 2020), para la búsqueda de similitudes entre las poses de docking y las poses de los compuestos en los cristales. Este programa fue útil para demostrar la calidad de los resultados obtenidos tanto de los inhibidores que presentaban una estructura cristalizada, como de quienes poseían un homólogo en SARS-CoV-1 o en SARS-CoV-2. LigRMSD selecciona los átomos comunes entre las estructuras comparadas, estableciendo grafos coincidentes entre ellas, calculando la desviación cuadrática media (RMSD) entre átomos equivalentes con respecto a una referencia. Dicho esto, 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 se realizaron dos etapas de estimación mediante LigRMSD; primero se compararon las estructuras de los inhibidores que poseían un cristal, versus sus poses obtenidas en docking para ambos SARS-CoV. Posteriormente, se analizaron las poses de los inhibidores que no poseían una estructura cristalográfica, donde se compararon las poses obtenidas del ligando en cuestión perteneciente a SARS-CoV-1, versus su homólogo en SARS-CoV-2.

### 4.3 Dinámica molecular

La dinámica molecular (MD) es una técnica computacional que simula el comportamiento dinámico de los sistemas moleculares en función del tiempo, tratando todas las entidades del espacio de simulación como flexibles, lo que genera diversas conformaciones estructurales tanto de las proteínas como de los ligandos presentes. Además, MD permite individualizar moléculas de agua relevantes, y estimar su contribución en la modulación de la unión de ligandos. La posibilidad de investigar el proceso de unión proteína-ligando brinda información sobre los estados más estables alcanzados por el ligando durante la simulación, sitios de unión alternativos, el papel del agua durante la unión y los reordenamientos conformacionales de las estructuras presentes en el sistema. Desafortunadamente, todo el proceso de reconocimiento, desde el estado no ligado al ligado, es un evento muy complejo de describir a nivel molecular, incluso con los recursos informáticos actuales, debido a que se requiere una gran cantidad de tiempo de MD (Henzler-Wildman y Kern, 2007). En particular, mayores escalas de tiempo están vinculadas a procesos que requieren la superación de una barrera de alta energía, lo que produce en ocasiones que el sistema simulado quede atrapado en un mínimo local, por lo que la MD clásica es inadecuada para explorar en gran medida el espacio conformacional (Salmaso & Moro, 2018).

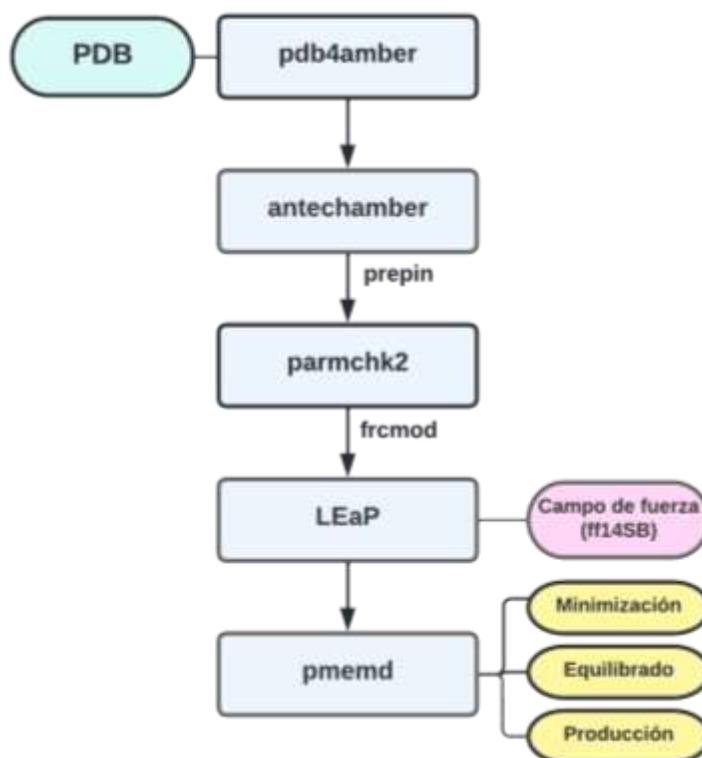
Por otro lado, la dinámica molecular acelerada (DMA) es un método que mejora el muestreo del espacio conformacional al reducir las barreras de energía que separan los diferentes estados de un sistema (Figura 12), modificando el espectro de la energía potencial, elevando los pozos de energía que están por debajo de un cierto nivel de umbral, mientras que los que están por encima de este nivel no se ven afectados. Como resultado, se reducen las barreras que separan las cuencas de energía adyacentes, lo que permite que el sistema pueda muestrear el espacio conformacional al que no se puede acceder fácilmente en una simulación de dinámica molecular clásica (Hamelberg et al., 2004).



**Figura 12:** **A** Representación esquemática del método de dinámica molecular acelerada. **B** La función hipotética de energía potencial verdadera,  $V(r)$  se muestra mediante la línea negra. Para una energía de refuerzo fija,  $E$  (línea violeta discontinua), se representan una función de energía potencial modificada, de color rojo.

En cuanto a la DMA, esta se llevó a cabo utilizando el motor de simulación *Amber18* (Case et al., 2005), mediante los pasos expuestos en la figura 13. Se realizaron 54 dinámicas, correspondientes a los 18 complejos representados por los 9 inhibidores acoplados en cada una de las 2 proteasas para las tres replicas generadas por cada uno de los sistemas mencionados en un tiempo total de 210 (ns). En primera instancia, se generaron los archivos necesarios para la dinámica, utilizando *pdb4amber*. De esta forma, se convirtieron los archivos resultantes del proceso de acoplamiento molecular a un formato PDB

compatible con *Amber18*. La preparación de los ligandos para la DMA se llevó a cabo en dos etapas. La primera de estas fue el uso de *antechamber*, a partir del cual se construyó el archivo de topología y se encontraron los parámetros de campo de fuerza faltantes. En la segunda etapa, se generó un archivo de parámetros mediante la herramienta *parmchk2*. Posterior a lo descrito, se utilizó el programa *LEaP*, para añadir las cargas necesarias para neutralizar el sistema. Sumado a esto, se utilizó el campo de fuerza *Amber ff19SB* (Tian et al., 2020) y se solvató el sistema utilizando TIP3P como modelo de agua. Una vez concluidos los pasos anteriores, se desarrolló una minimización que comprende 10.000 pasos de ejecución y dos equilibrados bajo condiciones NVT y NPT. En cuanto a la simulación de producción, esta se llevó a cabo hasta que el sistema alcanzó el equilibrio. Para cada una de las etapas mencionadas anteriormente se usó *pmemd* (Particle Mesh Ewald Molecular Dynamics), el cual es el motor principal para ejecutar simulaciones DMA con *Amber*.



**Figura 13:** Flujo de trabajo general en Amber.

## **4.4 Análisis post-dinámica molecular**

Luego de realizar el proceso de dinámica molecular, se llevaron a cabo análisis comparativos considerando RMSD, RMSF, interaction finger-print y clustering. A través de estos análisis se encontraron similitudes y diferencias estructurales en el modo de unión de los compuestos frente a las dos proteasas PLpro provenientes de los virus SARS-CoV-1 y SARS-CoV-2, las cuales presentan un alto porcentaje de identidad en su sitio de unión a ligandos.

### **4.4.1 Análisis RMSD**

Mediante el software CPPTRAJ, que destaca por ser el programa principal de Amber para procesar trayectorias de coordenadas y archivos de datos (Roe & Cheatham, 2013), se estimaron los valores de desviación cuadrática media (RMSD) que corresponde a una característica comúnmente utilizada para comparar diferentes conformaciones de sistemas moleculares. En los enfoques de dinámica molecular y acoplamiento proteína-ligando (PLD), esta medida es esencial para el análisis y la validación de los resultados (Velázquez-Libera et al., 2020).

En base a esto, se hicieron estimaciones de los valores de RMSD del backbone de las distintas proteínas pertenecientes a cada uno de los sistemas simulados, para poder observar las diferencias de estabilidad estructural de las proteínas entre los complejos simulados tanto para SARS-CoV-1 como para SARS-CoV-2. El cálculo de RMSD se realizó considerando dos requisitos importantes; el primero consiste en que el número de átomos en el objetivo debe coincidir con el número de átomos en la referencia y segundo que el orden de los átomos en el objetivo debe coincidir con el orden de los átomos en la referencia. Además, CPPTRAJ necesita de un archivo de topología y un archivo de trayectoria para poder realizar los cálculos de RMSD para todos los átomos que no corresponden a hidrógenos.

### **4.4.2 Análisis de RMSF**

Con la finalidad de observar la fluctuación de los residuos que conforman el sitio de unión a ligandos en ambas proteínas (PLpro-SARS-CoV-1, PLpro-SARS-CoV-2), se evaluaron dichos residuos en presencia de los distintos inhibidores a estudiar. Para ello, se efectuó un análisis RMSF (Root Mean Square Fluctuation), donde se observaron residuos con un alto valor de fluctuación a lo largo de la dinámica, los cuales fueron un indicador de la diversidad conformacional presente en los sistemas analizados. RMSF corresponde a una métrica que muestra el promedio de cuánto oscilan los residuos de una proteína, alrededor de una estructura de referencia a lo largo del tiempo de simulación. Para este análisis se utilizó el programa CPPTRAJ (Roe & Cheatham, 2013). Este software tiene la capacidad de realizar diversos cálculos, entre los cuales se encuentra RMSF. Se realizó este cálculo para todos los residuos que se encuentren a 5.0 Å del inhibidor de mayor tamaño, destacando los residuos que componen los subsitios S3 y S4 en conjunto del bucle BL2.

### 4.4.3 Interaction fingerprint

Con la finalidad de analizar el modo de unión de los diversos ligandos presentes en cada uno de los sistemas simulados, se calculó el porcentaje de ocurrencia de cada una de las interacciones presentes en la tabla 2. Para esto se utilizó el script de Prolif Interaction fingerprint (IFP) (Bouysset & Fiorucci, 2021). La identificación de los IFP receptor-ligando se realizó utilizando el software RDKit45, que identifica las propiedades químicas del ligando utilizando un archivo de formato mol2 de ligando de entrada, y la librería de Python MDAnalysis46. Esta última permite la lectura de trayectorias MD, independientemente de su formato o del programa utilizado para generarlas, así como operaciones frame por frame en las trayectorias, como calcular distancias y detectar enlaces de hidrógeno entre grupos seleccionados de átomos (Kokh et al., 2020).

Tipo de interacción	Tipo de átomo del ligando	Átomos de la proteína	Distancia máxima (Å) o ángulo (°)
catiónica	Pos. ionizable	OE* OD* of ASP GLU	4.5
Enlace hidrogeno	Donor	O OC1 OC2 OH2 OW OD1 OD2 SG OE1 OE1 OE2 ND1 NE2 SD OG OG1 OHe	3.3, 100
Enlace hidrogeno	Aceptor	N OH2 OW NE NH1 NH2 ND2 SG NE2 ND1 NZ OG OG1 NE1 OHe	3.3, 100
Hidrofóbica	Hidrofóbico o fluorina	S de MET o CYS, carbonos excepto para CG en ASN ASP; CD en GLU, GLN ARG; CZ en TYR ARG; CB en SER THR; CE en LYS; C en backbone	4.0
Pi-stacking	Aromático	Átomos CZ* CD* CE* CG* CH* NE* ND* in PHE TRP TYR HIS HI2 HIE HIDc	5.5
Van der Waals	Todos	Todos	Suma de radios de VdW

**Tabla 2:** Definiciones de interacciones utilizadas para calcular los IFP en el análisis MD-IFP

### 4.4.4 Clustering

Con el objetivo de encontrar conformaciones estructurales representativas de los sistemas proteína-ligando (centroides) en presencia de un ligando en cuestión, se realizó un análisis de agrupamiento estructural (clustering), donde las conformaciones de la DM similares geométrica o cinéticamente se concentran en el mismo grupo. Este procedimiento de clustering puede reducir una gran cantidad de conformaciones de MD en un pequeño número de estados conformacionales, lo que nos permite investigar más a fondo varias propiedades termodinámicas o cinéticas de las macromoléculas estudiadas. Este análisis se llevó a cabo utilizando Bitclust (González-Alemán et al., 2020), la cual es una herramienta que utiliza una interfaz de línea de comandos de Python concebida para el agrupamiento rápido de trayectorias de dinámica molecular relativamente largas. Para establecer los cálculos de agrupamiento se utilizaron los valores por defecto que integra Bitclust. Sin embargo, este software presenta una gran cantidad de opciones, donde el

único parámetro que es obligatorio incluir es -traj, el cual hace referencia al archivo de la trayectoria de la dinámica molecular. En cuanto al resto, de no ser añadidos en el comando, se utilizan los valores por defecto. Además, cabe destacar que la opción -top (archivo de topología), acepta los formatos de entrada psf o pdb y de no añadir este archivo, se obtendrá la información desde el archivo de trayectoria. También se puede ajustar el agrupamiento a preferencia del usuario mediante un amplio abanico de parámetros, dentro de los más relevantes se encuentran -first (primer frame) y -last (ultimo frame), los cuales permiten establecer los frames de la trayectoria que se quieren agrupar, -sel permite realizar una selección de los átomos de interés presentes en el sistema de la dinámica.

El clustering de las conformaciones estructurales de las proteínas pertenecientes a cada uno de los sistemas simulados se dividió en tres fases principales las cuales son:

### **1) Clustering de las dinámicas:**

- a) Se clusterizó el backbone de cada una de las proteínas presentes en los complejos simulados, utilizando como parámetros un cutoff de RMSD igual o menor a 1.5 Å y un número máximo de 5 clusters.
- b) Se clusterizaron las cadenas laterales de los residuos pertenecientes a los subsitios de unión S3 y S4 en conjunto del bucle BL2 para cada una de las proteínas presentes en los complejos simulados. Se utilizaron como parámetros un cutoff de RMSD igual o menor a 1.5 Å y un número máximo de 5 clusters.

### **2) Clustering de los centroides del clustering del backbone (fase 1.a):**

- a) Se clusterizó el backbone de cada uno de los centroides obtenidos de los cinco clusters resultantes en la fase 1.a de las proteínas presentes en los complejos simulados para SARS-CoV-1, utilizando como parámetros un cutoff de RMSD igual o menor a 1.5 Å y un número máximo de 5 clusters.
- b) Se clusterizó el backbone de cada uno de los centroides obtenidos de los cinco clusters resultantes en la fase 1.a de las proteínas presentes en los complejos simulados para SARS-CoV-2, utilizando como parámetros un cutoff de RMSD igual o menor a 1.5 Å y un número máximo de 5 clusters.
- c) Se clusterizó el backbone de cada uno de los centroides obtenidos de los cinco clusters resultantes en la fase 1.a de las proteínas presentes en los complejos simulados para SARS-CoV-1 y SARS-CoV-2, utilizando como parámetros un cutoff de RMSD igual o menor a 1.5 Å y un número máximo de 5 clusters.

### **3) Clustering de los centroides del clustering de las cadenas laterales (fase 1.b):**

- a) Se clusterizaron las cadenas laterales de los residuos pertenecientes a los subsitios de unión S3 y S4 en conjunto del bucle BL2 pertenecientes a los centroides obtenidos de los cinco clusters resultantes de la fase 1.b, para cada una de las proteínas presentes en los complejos simulados para SARS-CoV-1. Se utilizaron como parámetros un cutoff de RMSD igual o menor a 1.5 Å y un número máximo de 5 clusters.
- b) Se clusterizaron las cadenas laterales de los residuos pertenecientes a los subsitios de unión S3 y S4 en conjunto del bucle BL2 pertenecientes a los centroides obtenidos de los cinco clusters resultantes de la fase 1.b, para cada

una de las proteínas presentes en los complejos simulados para SARS-CoV-2. Se utilizaron como parámetros un cutoff de RMSD igual o menor a 1.5 Å y un número máximo de 5 clusters.

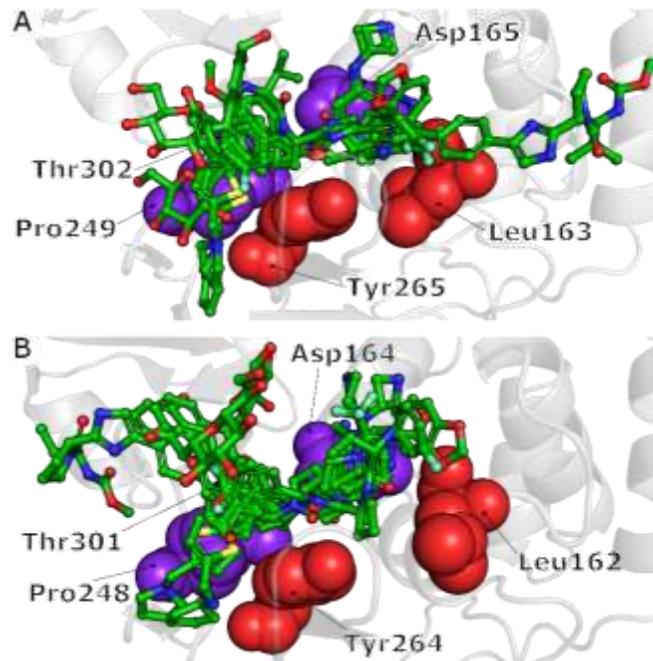
- c) Se clusterizaron las cadenas laterales de los residuos pertenecientes a los subsitios de unión S3 y S4 en conjunto del bucle BL2 pertenecientes a los centroides obtenidos de los cinco clusters resultantes de la fase 1.b, para cada una de las proteínas presentes en los complejos simulados para SARS-CoV-1 y SARS-CoV-2. Se utilizaron como parámetros un cutoff de RMSD igual o menor a 1.5 Å y un número máximo de 5 clústers.

## 5. Resultados y discusión

Para mayor comprensión de este escrito, se establece una nomenclatura para hacer alusión a los residuos correspondientes a las proteasas SARS-CoV-1-PLpro y SARS-CoV-2-PLpro. Esta se ha definido como: “**ResiduoN°residuo\_SARS-CoV-1-PLpro/ N°residuo\_SARS-CoV-2-PLpro**” (ejemplo: “Asp165/164”).

### 5.1 Docking molecular

A partir del método de acoplamiento molecular se crearon diversos modelos de interacción proteína-ligando. Se utilizó el conjunto de inhibidores de SARS-CoV-1/SARS-CoV-2-PLpro descritos en la Tabla 1, los cuales fueron situados entre los subsitios de unión S3 y S4 de la proteasa (Figura 14).



**Figura 14:** Resultados para el protocolo de acoplamiento molecular. **A)** Conjunto de inhibidores acoplados a la proteasa PLpro de SARS-CoV-1. **B)** Conjunto de inhibidores acoplados a la proteasa PLpro de SARS-CoV-2. Los inhibidores se presentan con sticks de color verde, mientras que algunos residuos de los subsitios S3 y S4 se presentan con esferas rojas y moradas, respectivamente.

### 5.1.1 Docking de 15G

Los resultados del acoplamiento molecular para el inhibidor 15G revelan una orientación similar a la estructura cristalográfica de 15G unida a PLpro, reportada por Gosh et al. (Ghosh et al., 2010). Se observa un enlace de hidrógeno entre el nitrógeno de la carboxamida del ligando, con el carbonilo del *backbone* del residuo Tyr269, correspondiente a SARS-CoV-1. La cadena lateral del mismo residuo (Tyr269) se encarga de mantener interacciones del tipo pi-stacking con el grupo naftaleno del compuesto. Esta interacción se conserva para el residuo análogo en SARS-CoV-2, Tyr268. Otra de las interacciones presentes en el complejo SARS-CoV-2-PLpro-15g corresponde a aquella formada por el nitrógeno protonado de la piperidina y el carboxilato de la cadena lateral del Asp165. La presencia de estas interacciones dentro de la hendidura de unión de PLpro es probablemente responsable de anclar al ligando y potencia su capacidad inhibitoria (Báez-Santos et al., 2015). Además, la piperidina protonada de este compuesto, no solo favorece la formación de enlaces de hidrógeno, sino que también participa en la formación de interacciones del tipo catión-pi por el anillo aromático del residuo Tyr265/264. Esta interacción no solo ocurre por parte del residuo Tyr265/264, el residuo cargado positivamente Lys157 (SARS-CoV-2) establece este tipo de interacción con el grupo aromático que enlaza al benzodioxol en el extremo del inhibidor (Figura 15.A).

### 5.1.2 Docking de 3K

3K es uno de los compuestos proveniente del naftaleno de segunda generación, obtenido a partir de la optimización del inhibidor 15g (Báez-Santos et al., 2015). Las interacciones observadas para el complejo proteína-ligando son similares a aquellas obtenidas en el caso anterior (compuesto 15g). Se observa la formación de un enlace de hidrógeno entre el carbonilo principal de Tyr269 y el nitrógeno de la carboxiamida del inhibidor, para el caso de SARS-CoV-1. Sumado a esto, se presenta un enlace de hidrógeno adicional entre el grupo piperidina del inhibidor y el carboxilato de la cadena lateral de Asp165/164. Ambos enlaces de hidrógeno son los encargados de unir de manera estable al inhibidor entre el sitio de unión (Báez-Santos et al., 2015). Además, la carga negativa otorgada por la cadena lateral de Asp165/164, produce un puente salino con la piperidina protonada. El mismo nitrógeno en la piperidina, mantiene una interacción de pi-cation con la cadena lateral del residuo Tyr265-264. Este último residuo establece además dos interacciones tipo pi-stacking con el grupo naftaleno del compuesto, al igual que los residuos Tyr269/268 (Figura 15.B, Figura S1.B, Figura S2.B).

### 5.1.3 Docking de GRL0617

El acoplamiento de GRL0617 en el sitio de unión de SARS-CoV-1/SARS-CoV-2 PLpro muestra interacciones ya descritas para este tipo de inhibidores (derivados de naftaleno). La amida neutra del inhibidor forma un enlace de hidrógeno con la cadena lateral del Asp165/164. A diferencia de los inhibidores anteriores que provienen de la misma serie congénica, este carece de una piperidina protonada, lo cual resulta en la no formación del puente salino establecido por 3K. Además, en el caso de la PLpro de SARS-CoV-2, este inhibidor, interactúa con Glu167 formando un enlace de hidrogeno con el grupo amino que se encuentra presente en GRL0617. El resto de las interacciones son consistentes a aquellas reportadas en los casos anteriores. Por otra parte, la cadena lateral de Tyr268/269 mantiene las interacciones del tipo pi-stacking con el grupo naftaleno del ligando. Las interacciones mencionadas anteriormente han sido reportadas como relevantes en la estabilización entre GRL0617 y SARS-CoV-1-PLpro, como para SARS-CoV-2-PLpro (Ratia et al., 2008) (Figura 15.C, Figura S1.C, Figura S2.C).

La orientación de estos compuestos derivados del naftaleno (15g, 3K, GRL0617) puntualmente sus restos de naftilmetilamina, se ubican en la misma posición dentro del sitio activo de la enzima. Además, cabe destacar que el resto de naftilo propio de estos inhibidores, se sitúa en un bolsillo hidrofóbico rodeado de los residuos Tyr269, Tyr265, Pro248, Pro249 y Thr302 (Calleja et al., 2022).

### 5.1.4 Docking de Aloin A / Aloin B

Las interacciones presentes en el compuesto Aloin A se generan entre los fenoles del grupo emodina del inhibidor, formando interacciones pi-stacking con los grupos aromáticos de Tyr269/268 y Tyr265 solo en SARS-CoV-1. Además, se observa la formación de un enlace de hidrógeno entre el carbonilo del *backbone* de la Tyr268 y el grupo 6-hidroxilo presente en la glucosa de Aloin A (Figura 15.D, Figura S1.D, Figura S2.D).

Por otro lado, en Aloin B, los hidroxilos presentes en los fenoles del grupo emodina generan enlaces de hidrógeno con las cadenas laterales de los residuos que se encuentran cercanos, como Asp165/164. De igual forma los hidroxilos presentes en la glucosa del inhibidor forman enlaces de hidrógeno con Pro249 y Gly267 en SARS-CoV-1 y en SARS-CoV-2 con el hidroxilo presente en la cadena lateral de Tyr268. Sumado a esto, al igual que en Aloin A, el grupo aromático de Tyr269/268, establece una interacción del tipo pi-stacking con uno de los fenoles del grupo emodina del inhibidor (Figura 15.E, Figura S1.E, Figura S2.E). Finalmente, el acoplamiento molecular de Aloin A y B con PLpro reveló que ambos isómeros de aloína forman enlaces de hidrógeno con el residuo Tyr269/268 de PLpro (Lewis et al., 2022).

### 5.1.5 Docking de Sitagliptin

El anclaje del Sitagliptin a la PLpro se caracteriza por la interacción pi-stacking del anillo trifluorobencilo con el residuo Tyr269. Sumado a esto, se pueden evidenciar interacciones de enlace de hidrógeno y puentes salinos entre el grupo amonio del inhibidor y el carboxilato de cadena lateral de Asp165/164. Dicho grupo también tiene una interacción pi-cación con el grupo aromático del residuo Tyr265/264. Mediante inspección visual, se determinó que el modo de unión de Sitagliptin es similar al de los compuestos derivados del naftaleno, puntualmente los grupos trifluorobencilo y el naftaleno ocupan la misma posición (Figura 15.F, Figura S1.F, Figura S2.F).

### 5.1.6 Docking de Daclatasvir

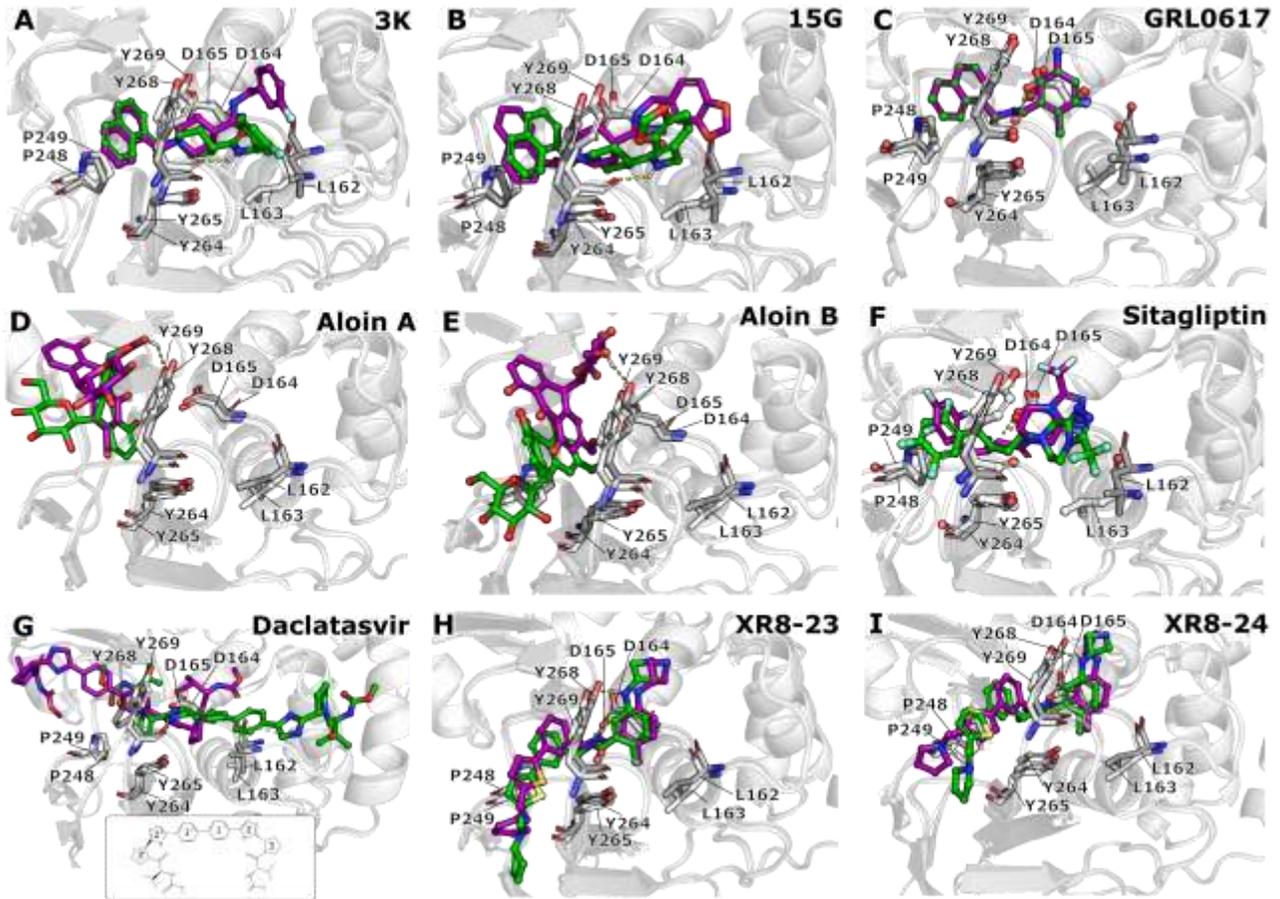
Los resultados de docking para Daclatasvir tanto en SARS-CoV-1 como en SARS-CoV-2, muestran diferencias en cuanto a su disposición estructural dentro del sitio de unión. En el caso de SARS-CoV-1, el extremo del anillo 3' se ubica entre los subsitios S3 y S4, formando enlaces de hidrógeno entre el imidazol (anillo 2') perteneciente al inhibidor y el carbonilo del *backbone* del Asp165 y la Tyr269. Por otro lado en SARS-CoV-2, el extremo 3' se orienta hacia el dominio de fingers, provocando que el imidazol perteneciente al extremo 3 (anillo 2) se sitúe en el sitio de unión, originando interacciones con Gly163 y Tyr268 (Figura 15.G). Los resultados del acoplamiento muestran que Daclatasvir no puede insertarse completamente en el sitio de unión de la proteasa, y la mayor parte del compuesto está expuesto al solvente. Esto se debe a que la geometría de PLpro se caracteriza por un sitio de unión muy estrecho, requiriendo un compuesto altamente flexible, lo que limita la unión de una molécula más rígida y grande, como Daclatasvir (Narayanan et al., 2022) (Figura S1.G, Figura S2.G).

### **5.1.7 Docking de XR8-23**

El compuesto XR8-23 presenta interacciones pi-stacking entre los grupos arilos con el anillo aromático de la cadena lateral de Tyr269/268. Además, se pueden observar enlaces de hidrógeno, los cuales se producen entre los distintos grupos nitrogenados presentes en la estructura del inhibidor y los residuos cercanos como Tyr269, Asp165/164, Glu168/167, Gly 267/266 (Figura 15.H, Figura S1.H, Figura S2.H).

### **5.1.8 Docking de XR8-24**

El inhibidor XR8-24 presenta interacciones pi-stacking entre los grupos arilos con el anillo aromático de la cadena lateral de Tyr269/268. Además, se pueden observar enlaces de hidrogeno, que se producen entre los distintos grupos nitrogenados presentes en la estructura del inhibidor y los residuos cercanos como lo son Asp165, Glu168, Tyr269 y Gly267 para la PLpro de SARS-CoV-1, y con Asp164, Glu167 y Pro248 para la PLpro de SARS-Cov-2 (Figura 15.I, Figura S1.I, Figura S2.I).



**Figura 15:** Comparación estructural entre las poses de acoplamiento molecular. Se presentan en verde y magenta las poses de los inhibidores correspondientes a SARS-CoV-1 y SARS-CoV-2, respectivamente. Los identificadores PDB para SARS-CoV-1/SARS-CoV-2 utilizados en cada caso, son: A) 4OW0/7CJM, B) 3MJ5/7CJM, C) 3E9S/7CJM, D) 3E9S/7CMD, E) 3E9S/7CMD F) 4OW0/7CJM, G) 7LFU/6WX4, H) 3E9S/7LBS, I) 3E9S/7LBS. Los residuos pertenecientes a SARS-CoV-1-PLpro se colorean en blanco y los de SARS-CoV-2 en gris. En el recuadro de G) se muestran las numeraciones correspondientes a los anillos del inhibidor.

### 5.1.9 LigRMSD

Posterior al acoplamiento molecular, se realizó una comparación estructural de las poses obtenidas para cada uno de los inhibidores en el sitio de unión de los 18 sistemas, mediante el servidor LigRMSD (Velázquez-Libera et al., 2020). Este se encarga de identificar los grafos comunes entre moléculas que se comparan y realiza un cálculo de la desviación cuadrática media (RMSD) entre estas.

Para el cálculo de LigRMSD se utilizaron como referencia aquellos compuestos que presentan cristales en la base de datos PDB, donde se encuentran los inhibidores 3K, 15G, GRL0617y XR8-24. Por otro lado, los inhibidores carentes de una estructura cocrystalizada, fueron comparados con sus respectivas poses obtenidas en el proceso de docking

molecular en la proteasa PLpro de SARS-CoV-1. Cabe mencionar que el inhibidor XR8-23 se consideró dentro del conjunto que posee cristal, debido a la similitud estructural con XR8-24, variando únicamente en una sustitución de la pirrolidina por un ciclopentilamina. En literatura se ha descrito que un valor RMSD = 2.0 Å puede ser considerado como umbral para discriminar entre las soluciones de acoplamiento correctas e incorrectas para compuestos idénticos (el valor puede ser mayor para compuestos no idénticos) (Velázquez-Libera et al., 2019). En base a esto, se consideró este valor como umbral, para evaluar la orientación de los compuestos utilizados en este estudio expuestos en la Tabla 1.

Los resultados del análisis LigRMSD se muestran en la Tabla 3. A partir de esta, se desprende que al analizar el conjunto de inhibidores que poseen como referencia un cristal, los valores de RMSD obtenidos son menores a 2.0 Å en el 80% de los casos. Esto sugiere una correspondencia entre la disposición estructural de las poses obtenidas mediante el acoplamiento molecular y el compuesto co-cristalizado. Sin embargo, en algunos casos, las poses obtenidas para 3K y 15G en donde se utilizó SARS-CoV-2 PLpro para el acoplamiento, resultó en valores de RMSD de 2.60 y 3.71, respectivamente. La causa de la desviación en la estructura del inhibidor se debe a la diferencia en la disposición estructural de los residuos presentes en el sitio de unión de la proteína PDB:7CJM, generando un leve desplazamiento del ligando, en ambos casos (Figura 15.A-B).

Por otro lado, los valores de RMSD para aquellos inhibidores que fueron comparado con su pose homóloga en SARS-CoV-1, fueron superiores a 2.0 Å. Las causas de los valores superiores al umbral RMSD establecido, han sido descritos en tres puntos: 1) En el caso de Aoin A y Aoin B, la cantidad de anillos e hidroxilos que posee su estructura interactúa de distinta forma con los residuos aromáticos presentes en el sitio de unión (TYR269-268 o TYR265-264), generando diferentes orientaciones (Figura 15.D-E). 2) En cuanto a Daclatasvir, el alto valor en el RMSD se origina debido a que a pesar de que ambas estructuras presentan una misma orientación, la pose obtenida para SARS-CoV-2 se encuentra desplazada con respecto a su referencia (Figura 15.G). 3) El último compuesto corresponde a Sitagliptin. Pese a que tanto la referencia situada en 4OW0 y la pose comparada ubicada en 7CJM presentan una igual disposición en el sitio de unión, la conformación de los residuos pertenecientes a los cristales de PLpro para ambas proteasas no es similar. Esto genera un leve desplazamiento del ligando y en consecuencia un valor de RMSD superior a 2.0 Å (Figura 15.F).

LIGANDO	Virus	PDB del cristal	Glide Score (kcal/mol)	RMSD (Å)	%RefMatch	%MolMatch
<b>A</b> PDB: SARS-CoV-1 3K 4OW0						
3K	SARS-CoV-1	4OW0	-7,874	1,21	100	100
	SARS-CoV-2	7CJM	-5,808	2,60	100	100
<b>B</b> PDB: SARS-CoV-1 15G 3MJ5						
15G	SARS-CoV-1	3MJ5	-5,962	1,38	100	100
	SARS-CoV-2	7CJM	-4,775	3,71	100	100
<b>C</b> PDB: SARS-CoV-1 GRL0617 3E9S						
GRL0617	SARS-CoV-1	3E9S	-7,907	0,85	100	100
	SARS-CoV-2	7CJM	-6,791	0,85	100	100
<b>D</b> PDB: SARS-CoV-2 XR8-24 7LBS						
Xr8_23	SARS-CoV-1	3E9S	-8,332	0,94	85	83
	SARS-CoV-2	7LBS	-8,121	1,00	85	83
Xr8_24	SARS-CoV-1	3E9S	-7,965	1,32	100	100
	SARS-CoV-2	7LBS	-7,053	1,09	100	100
<b>E</b> DOCKING SARS-CoV-1 ALOIN A 3E9S						
Aloin A	SARS-CoV-1	3E9S	-3,726	0,00	100	100
	SARS-CoV-2	7CMD	-2,812	5,87	100	100
<b>F</b> DOCKING SARS-CoV-1 ALOIN B 3E9S						
Aloin B	SARS-CoV-1	3E9S	-4,988	0,00	100	100
	SARS-CoV-2	7CMD	-3,583	7,49	100	100
<b>G</b> DOCKING SARS-CoV-1 DACLATASVIR 7LFU						
Daclatasvir	SARS-CoV-1	7LFU	-4,458	0,00	100	100
	SARS-CoV-2	6WX4	-4,953	19,51	100	100
<b>H</b> DOCKING SARS-CoV-1 SITAGLIPTIN 4OW0						
Sitagliptin	SARS-CoV-1	4OW0	-6,822	0,00	100	100
	SARS-CoV-2	7CJM	-5,697	3,44	100	100

**Tabla 3:** Análisis de LigRMSD de los compuestos utilizados con respecto a su referencia. **A)** Las poses se comparan con el cristal del inhibidor 3K cocrystalizado en el PDB 4OW0. **B)** Las poses se comparan con el cristal del inhibidor 15G cocrystalizado en el PDB 3MJ5. **C)** Las poses se comparan con el cristal del inhibidor GRL0617 cocrystalizado en el PDB 3E9S. **D)** Las poses de XR8-23 y XR8-24 se comparan con el cristal del inhibidor XR8-24 cocrystalizado en el PDB 7LBS. **E)** Comparación de la pose de Aloin A en el cristal 7CMD con respecto a la referencia Aloin A acoplado al cristal de SARS-CoV-1: 3E9S. **F)** Comparación de la pose de Aloin B en el cristal 7CMD con respecto a la referencia Aloin B acoplado al cristal de SARS-CoV-1: 3E9S. **G)** Comparación de la pose de Daclatasvir en el cristal 6WX4 con respecto a la referencia Daclatasvir acoplado al cristal de SARS-CoV-1: 7LFU. **H)** Comparación de la pose de Sitagliptin en el cristal 7CJM con respecto a la referencia Sitagliptin acoplado al cristal de SARS-CoV-1: 4OW0.

## 5.2 Estabilidad y dinámica conformacional del complejo proteína-ligando

La naturaleza dinámica de complejos proteína-ligando ha sido ampliamente estudiada por simulaciones MD para examinar la estabilidad y cambios en la conformación de los complejos, los movimientos internos y las interacciones proteína-ligando. En este estudio, las simulaciones MD se realizaron para los complejos SARS-CoV-1-PLpro y SARS-CoV-2-PLpro, con el conjunto de inhibidores presentado en la tabla 1. Utilizando las trayectorias de MD generadas, se realizaron análisis RMSD, RMSF, Interaction Fingerprint y Clustering. Cabe mencionar que, en 14 de los 18 sistemas simulados, los inhibidores permanecen unidos al sitio de unión a lo largo de todo el tiempo de simulación para las tres réplicas realizadas. Por otro lado, se presentan 4 sistemas donde el ligando se libera del sitio de unión, Aoin A y Aoin B, tanto para SARS-CoV-1 como SARS-CoV-2. Cabe destacar que el complejo SARS-CoV-1 Aoin A, no logra permanecer en ningún momento unido al sitio de unión, por lo que fue omitido para los análisis siguientes, obteniendo así 17 sistemas.

### 5.2.1 RMSD

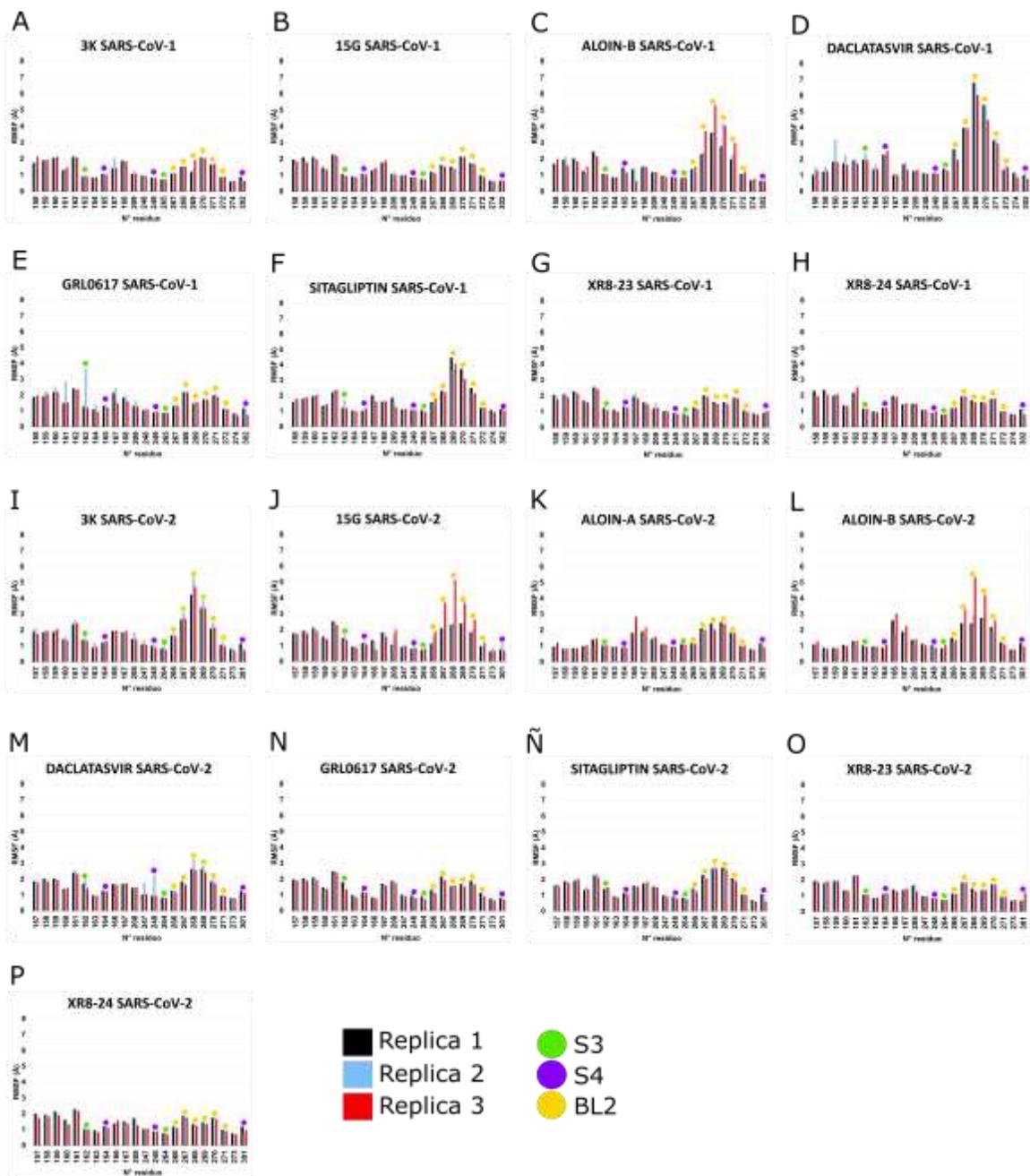
Para comprender la estabilidad conformacional y los cambios en la evolución temporal de los complejos proteína-ligando de SARS-CoV-1-PLpro y SARS-CoV-2-PLpro, se realizó un análisis RMSD considerando el *backbone* de la proteína. Este análisis se llevó a cabo evaluando desde el estado inicial 0 (ns) hasta el estado final 210 (ns) de los complejos, considerando las tres replicas realizadas. De modo general, se puede observar que todos los sistemas se estabilizaron dentro de los 30 (ns) de simulación. La figura suplementaria S3 indica que los valores de RMSD para los sistemas se sitúan alrededor de 2.0 Å, lo que sugiere estabilidad en el sistema, tanto para SARS-CoV-1 como para SARS-CoV-2.

Sin embargo, al analizar las dinámicas de los sistemas que contenían los inhibidores Aoin A y Aoin B, tanto para SARS-CoV-1 como para SARS-CoV-2 PLpro, se puede observar que los inhibidores no permanecen en el sitio de unión durante toda la trayectoria. Para estos casos, se analizaron los instantes en que el inhibidor se encontraba en el sitio de unión de la proteasa, resultado así que 1) en SARS-CoV-1-Aoin-A, el ligando no logró situarse de manera consistente en el sitio de unión en ninguna de las tres réplicas. Por otra parte 2) SARS-CoV-1-Aoin-B, presenta la réplica 1 con el inhibidor presente en el sitio de unión durante toda la trayectoria, la réplica 2 un total de 50 (ns), y para la réplica 3 se consideraron 27 (ns). 3) En cuanto a SARS-CoV-2-Aoin-A, mostró que tanto la réplica 1 como la réplica 3 contenían al inhibidor dentro al sitio de unión a diferencia de la réplica 2 donde el inhibidor se situó en el sitio solamente durante 110 (ns). 4) Por último, de SARS-CoV-2-Aoin-B se analizaron 83 (ns) para su réplica 1, 26 (ns) para la réplica 2 y 210 (ns) para la réplica 3. Al observar los análisis RMSD de las porciones de las trayectorias de los tres complejos que han sido mencionados anteriormente (2, 3 y 4), se han observado los mismos resultados que para el resto de los complejos.

### 5.2.2 RMSF

EL RMSF es un parámetro esencial para identificar las regiones rígidas y flexibles de la estructura de la proteína (Bhowmik et al., 2021). En base a esto, se evaluó la flexibilidad de los residuos presentes en el sitio de unión proteína-ligando, con la finalidad de conocer aquellos que podrían ser causantes de diversidad conformacional en cada uno de los sistemas estudiados.

Al observar los gráficos presentes en la Figura 16, se puede distinguir que existe una tendencia en la fluctuación de los residuos que forman parte del sitio. De manera generalizada, para todas las réplicas, los residuos pertenecientes a los subsitios de unión S3 (Leu163/162, Tyr265-264) y S4 (Asp165/164, Pro249/248, Thr302/301) presentan valores de RMSF menores a los 2.0 Å, lo que podría significar una estabilidad causada por las interacciones entre los diversos inhibidores y los residuos mencionados con anterioridad. Por otro lado, se puede observar que los residuos que conforman el bucle BL2 (Gln267/266-Gln272/271), son los que presentan valores más altos de RMSF dentro de las gráficas. Se ha reportado que el bucle BL2 es flexible y crucial en la unión entre PLpro e inhibidores, ya que da como resultado una conformación abierta (sin ligando) o cerrada cuando un sustrato se une al sitio (Alabbas & Alamri, 2022).



**Figura 16:** Resultados del análisis RMSF. Graficas de los 18 sistemas con sus tres replicas simuladas, donde se destacan los residuos correspondientes al subsitio S3 en verde, al subsitio S4 en morado y en amarillo los que conforman el bucle BL2.

### **5.2.3 Interaction fingerprint (Huellas de interacción molecular)**

El análisis de IFP se utilizó para evidenciar las interacciones generadas entre el receptor (PLpro) y los ligandos acoplados. De esta forma, se obtuvo información más detallada del sitio de unión de PLpro para cada uno de los 17 sistemas analizados. Los resultados de las huellas de interacción molecular se presentan en la Tabla 4, la cual se complementa con las figuras suplementarias S4 y S5 para un mejor entendimiento.

#### **5.2.3.1 Puentes salinos**

En primera instancia se presentan las interacciones de tipo puente salino, donde el residuo más recurrente para la formación de dichas interacciones es Asp165/164. Esta interacción se presenta en un total del 70% de los complejos. Este porcentaje se compone de 1) 5 de 17 sistemas con un grado de ocurrencia superior al 80% de la dinámica y 2) 7/17 sistemas con un grado de ocurrencia superior a 50% pero inferior a 80% a lo largo de la simulación. Por otro lado, los complejos que poseen como inhibidor tanto a XR8-23 y XR8-24, también generan una interacción relevante con el residuo Glu168/167, tanto para SARS-CoV-1, como para SARS-CoV-2. Cabe destacar que se presentan sistemas que no poseen interacción alguna de este tipo, los cuales tienen como inhibidor a GRL0617, Aoin A y Aoin B, tanto para SARS-CoV-1 como SARS-CoV-2.

#### **5.2.3.2 Enlaces de hidrógeno**

Para evaluar este tipo de interacciones, se consideraron aquellos residuos que actúan como aceptor y donador en la formación de enlaces de hidrógeno. Dentro de los aceptores de hidrógenos se destaca el residuo Gln270/269, el cual participa en esta interacción de manera consistente (superior a un 80% de ocurrencia) para ambos SARS, en los sistemas que contienen a los inhibidores GRL0617, XR8-23 y XR8-24, sumado a Sitagliptin en SARS-CoV-2. Esta interacción también ocurre en el resto de los complejos, pero con un grado de ocurrencia menor al 40%.

Por otro lado, dentro de los residuos que actúan como donador en la formación de enlaces de hidrógeno se encuentra Asp165/164, el cual posee una mayor ocurrencia dentro del total de complejos. Sin embargo, este residuo no participa en los sistemas que poseen como inhibidor Aoin A, Daclatasvir y Aoin B tanto en SARS-CoV-1, como en SARS-CoV-2. Cabe mencionar que además de Asp165/164, el residuo Glu168/167 interactúa de manera estable a lo largo de las trayectorias para los sistemas que contienen los inhibidores XR8-23 y XR8-24.

### **5.2.3.3 Interacciones Hidrofóbicas**

En cuanto a las interacciones hidrofóbicas, éstas presentan una gran cantidad de residuos que participan con más de un 80% de ocurrencia a lo largo de las dinámicas para gran parte de los sistemas en estudio. Dentro de los residuos más relevantes se encuentran: Leu163/162, Asp165/164, Pro248/247, Pro249/248, Tyr265/264, Tyr269/268, Gln270/269, Tyr274/273 y Thr302/301.

#### **5.2.3.3.1 Interacciones Pi-stacking**

El análisis de las interacciones del tipo pi-stacking, revela que las porciones aromáticas pertenecientes a dos residuos generan interacciones de este tipo en todos los sistemas. Estos residuos corresponden a Tyr265/264 y Tyr269/268.

#### **5.2.3.3.2 Interacciones de Van der Waals (VdW)**

En cuanto a las interacciones de van der Waals, se han encontrado una gran cantidad de residuos que prevalecen generando este tipo de interacciones en gran parte de la trayectoria y que participan en casi todos los sistemas simulados. Los residuos más relevantes son: Asp165/164, Tyr265/264, Tyr269/268, Gln270/269 y por último Thr302/301.

SARS-CoV-1 Interacciones								
LIGANDO	PDB-ID	Catiónicas	HB.aceptor	HB.donor	Hidrofóbicas	Pi-stacking	VdW	
15G	3MJ5	<b>ASP165</b>	-	<b>ASP165</b> <b>TYR269</b>	<b>LEU163, APS165, MET209, PRO248, PRO249, TYR265, TYR269, GLN270, TYR274, THR302</b>	<b>TYR269</b>	<b>ASP165, PRO249, TYR265, TYR269, GLN270, TYR274</b>	
3K	4OW0	<b>ASP165</b>	-	<b>ASP165</b> <b>TYR269</b>	<b>LEU163, APS165, MET209, PRO248, PRO249, TYR265, TYR269, GLN270, TYR274, THR302</b>	TYR269	<b>ASP165, PRO249, TYR265, TYR269, GLN270, TYR274</b>	
Aloin B	3E9S	-	-	-	TYR265, TYR269	-	-	
Daclatasvir	7LFU	ASP165	-	-	LEU163, PRO249, <b>TYR265</b> , THR302	-	TYR265	
GRL0617	3E9S	-	<b>GLN270</b>	-	<b>LEU163, APS165, PRO248, PRO249, TYR265, TYR269, GLN270, TYR274, THR302</b>	TYR269	<b>ASP165, PRO259, TYR265, TYR269, GLN270</b>	
Sitagliptin	4OW0	<b>ASP165</b>	-	<b>ASP165</b>	APS165, MET209, PRO248, TYR265, TYR269, TYR274, THR302	-	<b>ASP165, TYR265, TYR269, TYR274, THR302</b>	
XR8-23	3E9S	ASP165 <b>GLU168</b>	<b>GLN270</b>	ASP165 <b>GLU168</b> GLY267	<b>LEU163, APS165, PRO248, PRO249, TYR265, TYR269, GLN270, TYR274, PRO300, THR302</b>	TYR269	<b>ASP165, GLU168, PRO249, TYR265, GLY267, TYR269, GLN270, TYR274</b>	
XR8-24	3E9S	ASP165 <b>GLU168</b>	<b>GLN270</b>	ASP165 <b>GLU168</b>	<b>LEU163, APS165, PRO248, PRO249, TYR265, TYR269, GLN270, TYR274, THR302</b>	TYR269	<b>ASP165, GLU168, PRO249, TYR265, GLY267, ASN268, TYR269, GLN270</b>	
SARS-CoV-2 Interacciones								
LIGANDO	PDB-ID	Catiónicas	HB.aceptor	HB.donor	hidrofóbicas	Pi-stacking	VdW	
15G	7CJM	<b>ASP164</b>	-	<b>ASP164</b> TYR268	<b>LEU162, ASP164, PRO247, PRO248, TYR264, TYR268, GLN269, TYR273, THR301</b>	-	<b>ASP164, PRO248, TYR264, TYR268, GLN269, TYR273</b>	
3K	7CJM	ASP164	-	ASP164 GLN269	LEU162, <b>ASP164, PRO248, TYR264, TYR268, GLN269, TYR273</b>	-	LEU162, <b>ASP164, TYR264, TYR268, GLN269, TYR273</b>	
Aloin A	7CMD	-	-	-	<b>PRO247, PRO248, TYR264, TYR268</b>	-	PRO247, PRO248, <b>TYR268</b>	
Aloin B	7CMD	-	-	-	PRO247, PRO248, TYR264, <b>TYR268</b>	-	<b>TYR268</b>	
Daclatasvir	6WX4	ASP164	-	<b>TYR268</b>	<b>LEU162, PRO247, PRO248, TYR264, TYR268, TYR273</b>	<b>TYR264</b>	ASP164, PRO248, TYR264, <b>TYR268, GLN269, TYR273</b>	
GRL0617	7CJM	-	<b>GLN269</b>	-	<b>LEU162, ASP164, PRO247, PRO248, TYR264, TYR268, GLN269, TYR273, THR301</b>	TYR268	<b>ASP164, PRO248, TYR264, TYR268, GLN269</b>	
Sitagliptin	7CJM	<b>ASP164</b>	<b>GLN269</b>	<b>ASP164</b>	ASP164, MET20, <b>TYR264</b>	-	<b>ASP164, TYR264, TYR268, GLN269, TYR273</b>	
XR8-23	7LBS	ASP164 <b>GLU167</b>	<b>GLN269</b>	ASP164 <b>GLU167</b>	<b>LEU162, ASP164, PRO248, TYR264, TYR268, GLN269, TYR273, THR301</b>	<b>TYR268</b>	<b>ASP164, GLU167, PRO248, TYR264, GLY266, ASN267, TYR268, GLN269, TYR273</b>	
XR8-24	7LBS	ASP164 <b>GLU167</b>	<b>GLN269</b>	ASP164 <b>GLU167</b>	<b>LEU162, ASP164, PRO248, TYR264, TYR268, GLN269</b>	TYR268	<b>ASP164, GLU167, PRO248, TYR264, GLY266, ASN267, TYR268, GLN269</b>	

HB:Hydrogen bond, VdW: Van der Waals

**Tabla 4:** Resultados del análisis Interaction Fingerprint, se muestran los residuos que participan en las distintas interacciones con los inhibidores expuestos, los residuos en letra normal poseen desde el 50% hasta el 80% ocurrencia y en negrita desde el 80% en adelante.

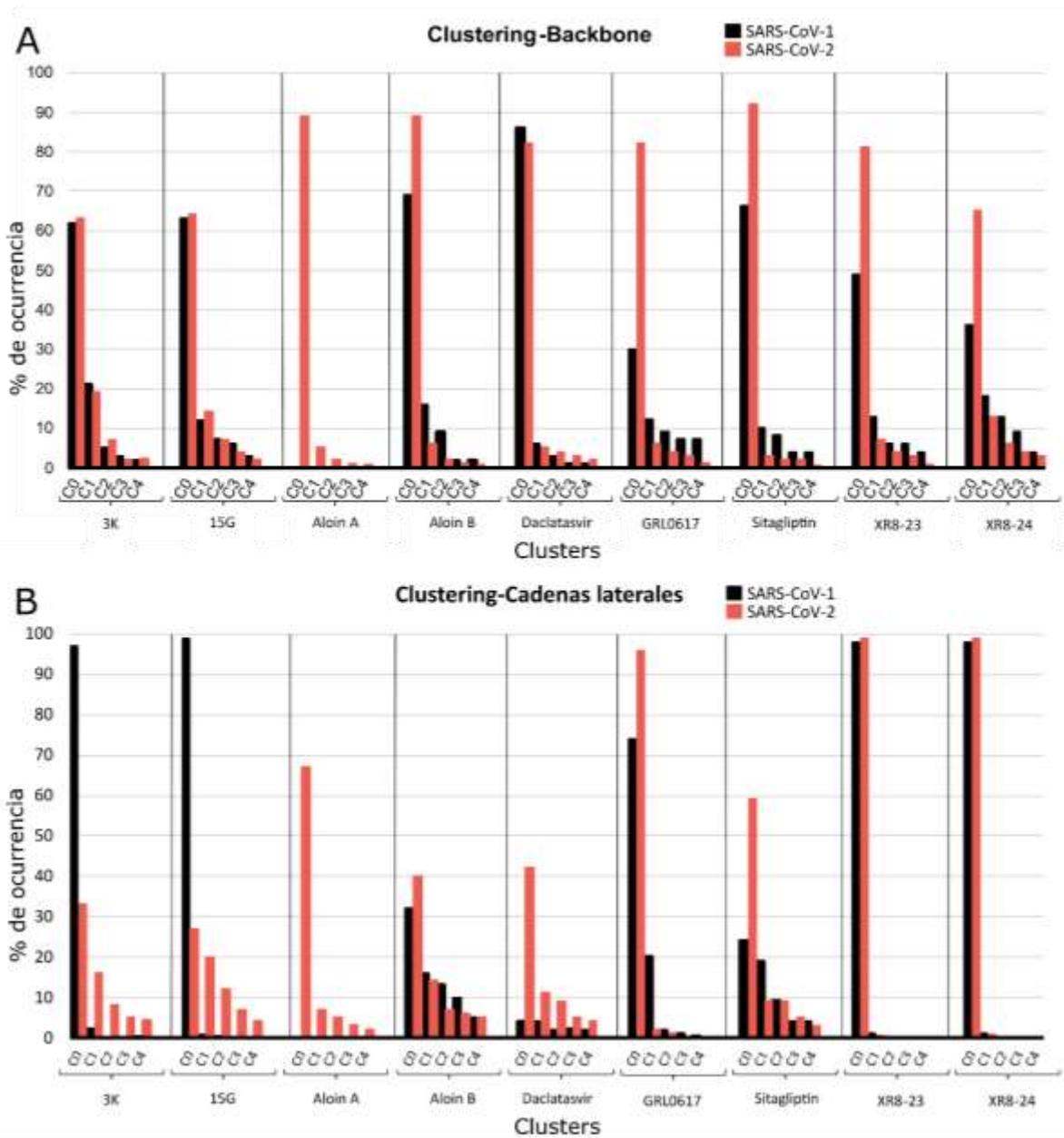
## 5.2.4 Fase 1: Clustering de las dinámicas

Se clusterizaron las dinámicas de los distintos sistemas simulados, utilizando como parámetro el RMSD tanto del *backbone* de las proteínas, como de las cadenas laterales de los residuos que conforman los subsitios de unión S<sub>3</sub> y S<sub>4</sub> en conjunto con el bucle BL2.

El clustering del *backbone* (Figura 17.A) muestra que todos los sistemas presentan un cluster con un alto porcentaje de ocurrencia (cluster 0). Esto indica que las estructuras de las proteínas presentes en cada sistema, poseen un alto grado de similitud a lo largo de las trayectorias.

Por otro lado, el clustering de las cadenas laterales de los subsitios de unión S<sub>3</sub> y S<sub>4</sub> en conjunto con el bucle BL2 (figura 17.B), muestra que para los complejos de SARS-CoV-1-PLpro con los inhibidores 3K, 15G, GRL0617, XR8-23 y XR8-24 presentan un cluster altamente poblado (cluster 0). Esto indica que existe una disposición estructural altamente similar de las cadenas laterales de los residuos estudiados. Por el contrario, para los complejos que presentan los inhibidores Aloin B, Daclatasvir y Sitagliptin, no existe un cluster altamente poblado, lo que es causado por una mayor variación en la disposición de las cadenas laterales.

Al igual que en el caso anterior, pero esta vez para su análogo en SARS-CoV-2, los complejos con los compuestos GRL0617, XR8-23 y XR8-24 presentan un cluster altamente poblado (cluster 0). A este conjunto de sistemas se suman los complejos con los inhibidores Aloin A y Sitagliptin, los cuales aparecen en porcentajes altos de 67% y 59%, respectivamente. De la misma forma que en SARS-CoV-1-PLpro, para los sistemas con inhibidores Aloin B y Daclatasvir no se presenta un cluster altamente predominante, como tampoco para los complejos con los compuestos 3K y 15G lo que es causado por una mayor variación en la disposición de las cadenas laterales presentes en estos sistemas.

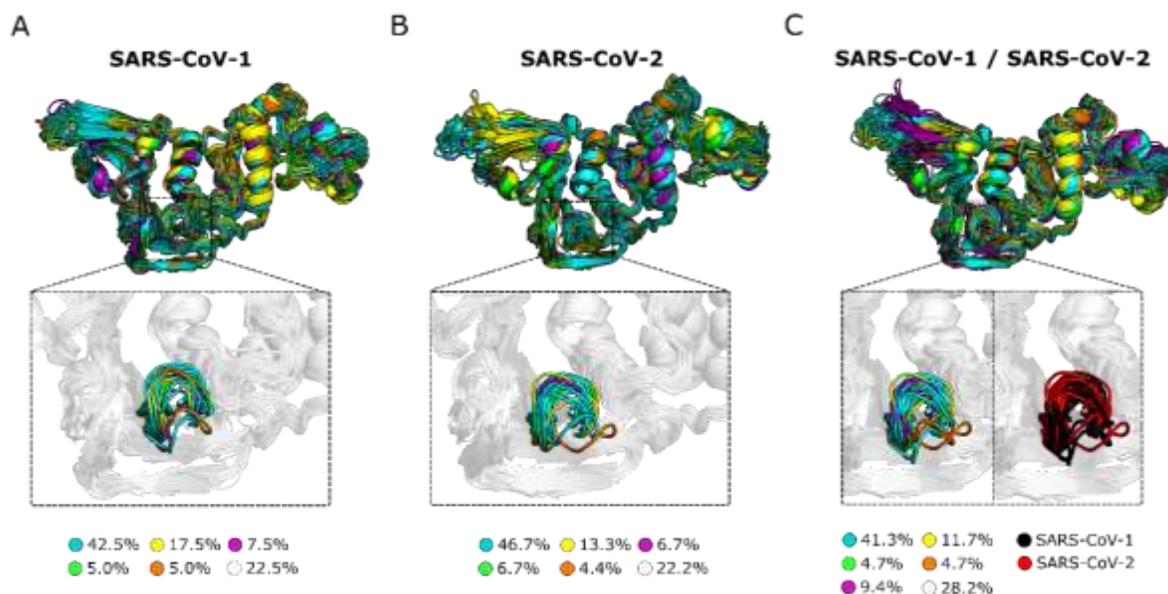


**Figura 17:** Resultados del clustering realizado a las dinámicas. Se grafican los 17 sistemas seleccionados para SARS-CoV-1 y SARS-CoV-2, donde se muestran los cinco clusters obtenidos para cada uno de los complejos con su inhibidor correspondiente.

### 5.2.5 Fase 2: Clustering de los centroides del *backbone*

Una vez que se analizaron los 5 clusters pertenecientes a cada uno de los 17 sistemas considerados, se obtuvieron los centroides correspondientes de cada uno de dichos clusters, resultando un total de 40 centroides para SARS-CoV-1 y 45 para SARS-CoV-2. Dichos centroides se clusterizaron en tres etapas, 1) únicamente los centroides pertenecientes a SARS-CoV-1; 2) centroides correspondientes a SARS-CoV-2 y 3) los 85 centroides de ambos SARS.

Los resultados que se obtuvieron una vez realizadas las tres etapas, muestran que gran parte de los centroides se agrupan en el cluster 0 (Figura 18.A.B.C), destacando que los centroides más representativos de cada sistema (centroide perteneciente al cluster 0 del clustering realizado a las dinámicas) se clasifican dentro de un mismo conjunto (Tabla S1, S2 y S3). Por otro lado, se analizaron las principales diferencias entre los centroides de los clusters obtenidos. En primer lugar, se realizó una inspección visual y posteriormente un análisis de RMSD para las regiones que mostraron mayor variación estructural. De estos análisis se desprende que gran parte de la variabilidad estructural que diferencian los cinco clusters, la aportan los loops presentes en los dominios “UBL” y “Fingers” y el bucle BL2, quien a pesar de contribuir a la diferencia no lo hace como las dos regiones mencionadas anteriormente (Figura S6, Figura S7). Dicho esto, al observar el sitio de unión de las distintas estructuras clusterizadas, no se consigue encontrar grandes diferencias conformacionales, lo que indica el *backbone* de PLpro en esta región posee una conformación estructural similar tanto en SARS-CoV-1 como en SARS-CoV-2 en presencia de inhibidores.



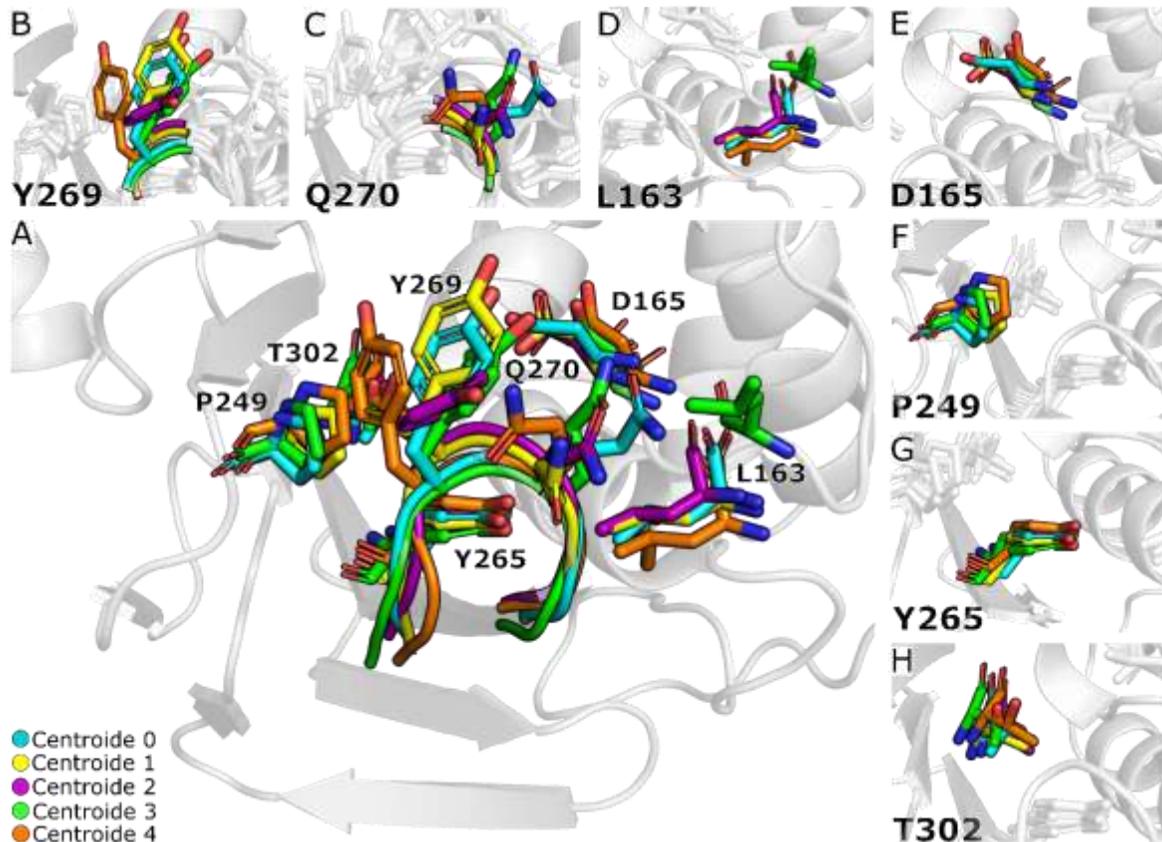
**Figura 18:** Clustering de los centroides del *backbone* **A)** Clustering de los centroides de las dinámicas del *backbone* para SARS-CoV-1 **B)** Clustering de los centroides de las dinámicas para SARS-CoV-2 **C)** Clustering de los centroides de las dinámicas para SARS-CoV-1 y SARS-CoV-2. En celeste se muestra el cluster 0, amarillo el cluster 1, morado cluster 2, verde cluster 3, naranja cluster 4 y en blanco los centroides sin clusterizar.

### **5.2.6 Fase 3: Clustering de los centroides de las cadenas laterales**

Al igual que en la fase 2 de clustering, se analizaron los 5 clústers pertenecientes a cada uno de los 17 sistemas considerados en la fase 1 de clustering. Se obtuvieron los centroides correspondientes a cada uno de dichos clústers, resultando un total de 40 centroides para SARS-CoV-1 y 45 para SARS-CoV-2. Con la finalidad de clusterizar las cadenas laterales de los residuos que se encuentran en los subsitios de unión  $S_3$ ,  $S_4$  y el bucle BL2, se reiteraron las tres etapas de clustering: 1) se agruparon, únicamente los centroides pertenecientes a la PLpro de SARS-CoV-1, 2) posteriormente los centroides correspondientes a la PLpro de SARS-CoV-2 y 3) por último los 85 centroides de ambas PLpro.

#### **5.2.6.1 Centroides de las cadenas laterales de la PLpro de SARS-CoV-1**

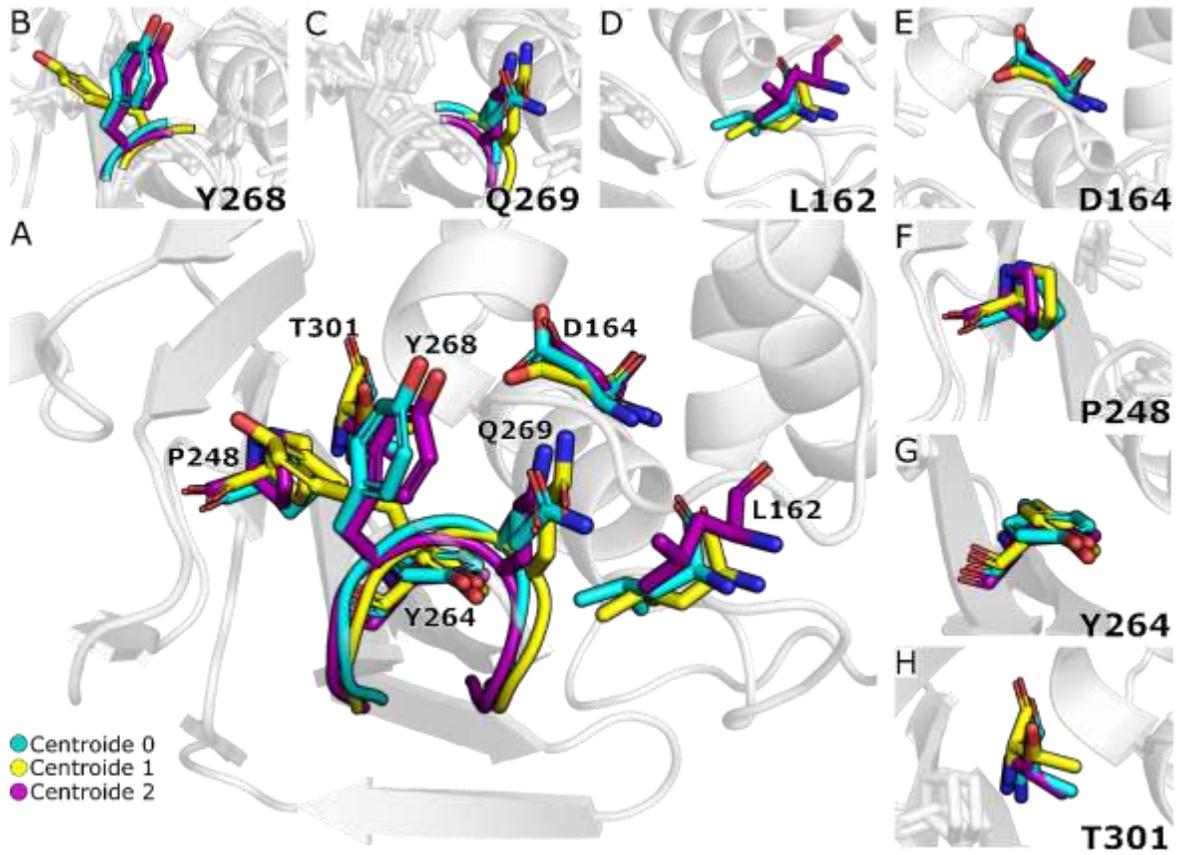
En primera instancia se analizó el clustering de los centroides de las cadenas laterales para SARS-CoV-1. A partir de esto se obtuvieron dos clusters medianamente poblados, cluster 0 y cluster 1, tres clusters con pocas estructuras (cluster 2, 3 y 4) y algunas estructuras no clusterizadas (Figura S8.A). Dentro de las principales características que se observan en este agrupamiento, es que los centroides 0 (centroides más representativos de las dinámicas), se han clasificado dentro de diferentes clusters (Tabla S4). Para poder observar de mejor manera las diferencias que presentan los distintos clusters, se han analizado los centroides correspondientes a cada uno de ellos (Figura 19.A). Dentro de las principales diferencias que se logran dilucidar entre los clusters se encuentra la disposición conformacional de los residuos Tyr269 (Figura 19.B) y Gln270 (Figura 19.C), sumado a un desplazamiento presente en Leu163 (Figura 19.D), perteneciente al cluster 3, lo cual puede originarse debido al gran tamaño del inhibidor Daclatasvir quien se corresponde con este cluster. Por otro lado, los residuos Asp165 (Figura 19.E), Pro249 (Figura 19.F), Tyr265 (Figura 19.G) y Thr302 (Figura 19.H), poseen una orientación similar.



**Figura 19:** Centroides pertenecientes al clustering de la fase 3 para SARS-CoV-1. Se muestran los residuos que conforman el subsitio S3 (Leu163, Tyr265), S4 (Asp165, Pro249, Thr302) y el bucle BL2 (Tyr269, Gln270) se presentan en sticks.

### 5.2.6.2 Centroides de las cadenas laterales de la PLpro de SARS-CoV-2

Se logran observar únicamente tres clusters, donde el cluster 0 se encuentra altamente poblado, a diferencia del cluster 1 y el cluster 2 (Figura S8.B). Además, se puede observar que en este clustering se han agrupado todos los centroides más representativos de las dinámicas en un mismo grupo (cluster 0) (Tabla S5). En adición a esto se han analizado los centroides pertenecientes a los tres clusters, lo cual evidenció que los residuos que aportan la variabilidad estructural a este clustering son aquellos pertenecientes al bucle BL2 (Tyr268 (Figura 20.B) y Gln269 (Figura 20.C)), sumados a Leu162 (Figura 20.D) perteneciente al subsitio S3, el cual presenta un leve desplazamiento. Por otro lado, los residuos Asp164 (Figura 20.E), Pro248 (Figura 20.F), Tyr264 (Figura 20.G) y Thr301 (Figura 20.H), poseen una orientación similar.

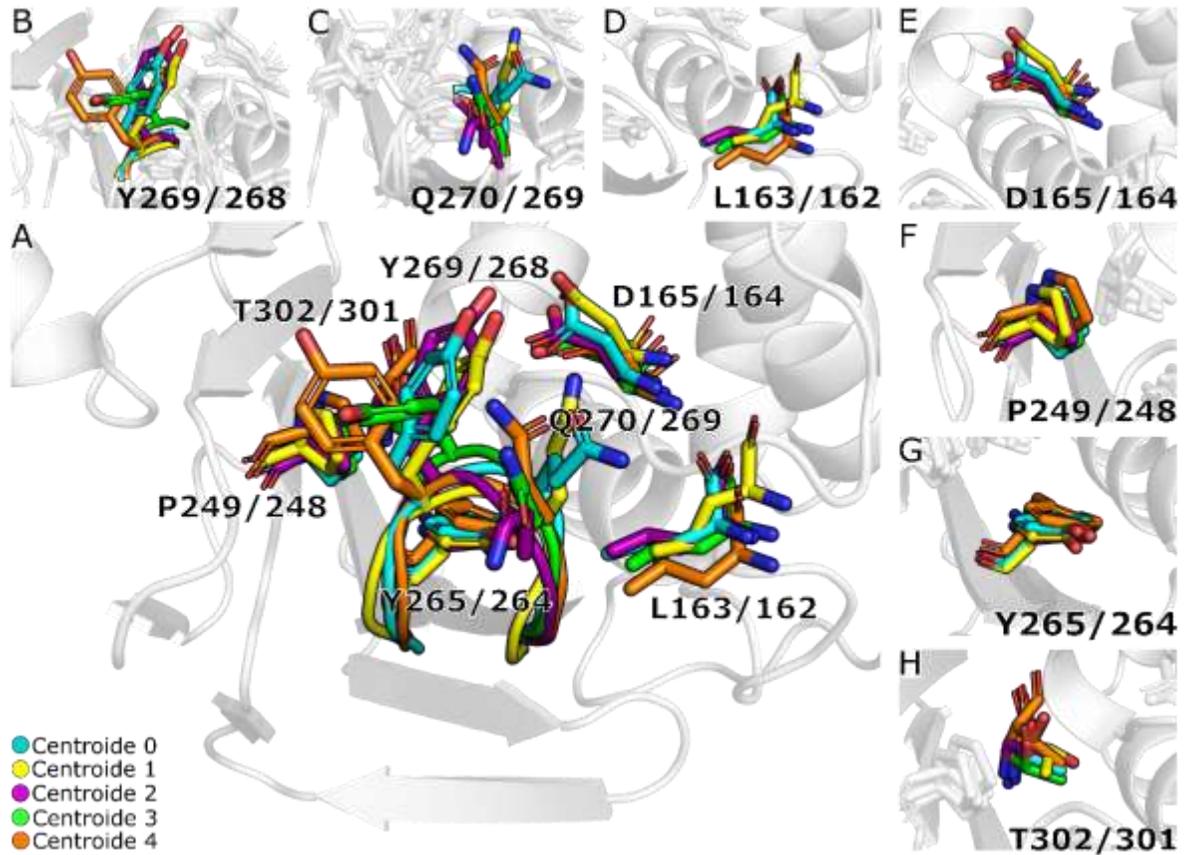


**Figura 20:** Centroides pertenecientes al proceso de clustering de la fase 3 para SARS-CoV-2. Los residuos que conforman los subsitios S3 (Leu162, Tyr264), S4 (Asp164, Pro248, Thr301) y el bucle BL2 (Tyr268, Gln269) se presentan en sticks.

### 5.2.6.3 Centroides de las cadenas laterales de las PLpro de SARS-CoV-1/SARS-CoV-2

Por último, se realizó la etapa 3 asociada a las cadenas laterales de los residuos (Figura S8.C). Los resultados del proceso revelan un agrupamiento de las estructuras en el cluster 0. En este se puede observar que, al igual que en el clustering de los centroides de SARS-CoV-2, todas las estructuras representativas de las dinámicas (centroides correspondientes al cluster 0 de las dinámicas) se clasifican en el mismo grupo. Por otro lado, las estructuras más representativas de la PLpro de SARS-CoV-1, a pesar de que su gran mayoría se ubica dentro del cluster 0, también se distribuyen en una menor proporción en el cluster 1. La conformación de la PLpro que contiene daclatasvir no logró ser clusterizada (Tabla S6). Sumado a esto, al analizar las estructuras clusterizadas, diferenciándolas por el tipo de SARS, no se logra establecer una clara diferencia estructural entre ambos tipos de virus, lo que explicaría el hecho de que las estructuras se clasifiquen dentro de un cluster común (Figura S9).

Con el fin de analizar con mayor detalle las diferencias que presentan los distintos clusters, se estudiaron los centroides correspondientes a cada uno de ellos (Figura 21.A). Dentro de las principales diferencias que se logran dilucidar entre los clusters, se encuentra la disposición conformacional de los residuos Tyr269/268 (Figura 21.B) y Gln270/269 (Figura 21.C), sumado a un desplazamiento presente en Leu163/162 (Figura 21.D). Por otro lado, los residuos Asp165/164 (Figura 21.E), Pro249/248 (Figura 21.F), Tyr265/264 (Figura 21.G) y Thr302/301 (Figura 21.H) poseen una orientación similar.



**Figura 21:** Centroides pertenecientes al clustering de la fase 3 para SARS-CoV-1 y SARS-CoV-2. Los residuos que conforman el subsitio S3 (Leu163/162, Tyr265/264), S4 (Asp165/164, Pro249/248, Thr302/301) y el bucle BL2 (Tyr269/268, Gln270/269) se presentan en sticks.

## 6. Conclusiones

De manera general, se realizó el estudio de 18 complejos compuestos de inhibidores diversos estructuralmente de las proteasas PLpro tanto para SARS-CoV-1 y SARS-CoV-2. Este trabajo nos permitió conocer con mayor detalle las características y el comportamiento de los inhibidores seleccionados (Tabla 1) en el sitio de unión de los diferentes estados conformacionales de SARS-CoV-1-PLpro y SARS-CoV-2-PLpro.

Dentro de los sistemas simulados, se pudo observar que en 14 de los 18 complejos, los inhibidores presentes en los complejos lograron interactuar de manera constante a lo largo de las dinámicas. Caso opuesto ocurre en los sistemas que presentan Aloin A y Aloin B, pues los inhibidores no se mantenían estables dentro del sitio de unión. Esto se corresponde con el bajo grado de afinidad con SARS-CoV-2-PLpro reportado para estos ligandos (Lewis et al., 2022).

En cuanto a las huellas de interacción molecular (Interaction Fingerprint), se demostró que las interacciones relevantes para mantener los inhibidores enlazados al sitio de unión para las PLpro tanto de SARS-CoV-1 como de SARS-CoV-2, involucran de manera más recurrente a los residuos Leu163/162, Asp165/164, Pro249/248, Tyr265/264, Tyr269/268, Gln270/269, y Thr302/301, los cuales pertenecen a los subsitios S3, S4 y al bucle BL2 de la PLpro.

Sumado a lo anterior, los resultados del clustering muestran la presencia de patrones estructurales que se han encontrado en el sitio de unión de PLpro tanto para SARS-CoV-1 y SARS-CoV-2. A partir de esto, se obtuvo una estructura común la cual es independiente del SARS-CoV y del inhibidor que se encuentre alojado en este sitio en la proteasa, la cual corresponde al centroide 0 (Figura 21). Esta estructura, en comparación con el resto de los centroides obtenidos, mantiene una similitud conformacional causada por los residuos Asp165/164, Pro249/248, Tyr265/264 y Thr302/30. Por otro lado, los residuos Tyr269/268, Gln270/269 y Leu163/162 son quienes confieren la variabilidad estructural, resultando esenciales para el ajuste del inhibidor al sitio de unión.

Para finalizar, el proveer nueva información a los antecedentes estructurales que se encuentran disponibles en literatura utilizando nuevos protocolos computacionales, permite contribuir en el campo investigativo dirigido a los virus SARS. Hasta la fecha, solo se han realizado estudios de clustering asociados a proteínas de SARS-CoV-1 y SARS-CoV-2 de manera independiente. Por esta razón, encontrar patrones estructurales comunes entre ambos virus, permite obtener información relevante y necesaria para la síntesis de nuevas moléculas capaces de inhibir la proteasa PLpro presente en los virus SARS-CoV.

## 7. Referencias

- Alabbas, A. B., & Alamri, M. A. (2022). Analyzing the effect of mutations in SARS-CoV2 papain-like protease from Saudi isolates on protein structure and drug-protein binding: Molecular modelling and dynamics studies. *Saudi Journal of Biological Sciences*, 29(1), 526-533. <https://doi.org/10.1016/j.sjbs.2021.09.028>
- Amin, Sk. A., Banerjee, S., Ghosh, K., Gayen, S., & Jha, T. (2021). Protease targeted COVID-19 drug discovery and its challenges: Insight into viral main protease (Mpro) and papain-like protease (PLpro) inhibitors. *Bioorganic & Medicinal Chemistry*, 29, 115860. <https://doi.org/10.1016/j.bmc.2020.115860>
- Asale, R.-, & RAE. (s. f.). *Pandemia | Diccionario de la lengua española. «Diccionario de la lengua española» - Edición del Tricentenario*. Recuperado 9 de abril de 2022, de <https://dle.rae.es/pandemia>
- 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. <https://doi.org/10.1016/j.antiviral.2014.12.015>
- Bagherzadeh, K., Daneshvarnejad, K., Abbasinazari, M., & azizian, homa. (2020). In silico Repositioning for Dual Inhibitor Discovery of SARS-CoV-2 (COVID-19) 3C-like Protease and Papain-like Peptidase [Preprint]. *CHEMISTRY*. <https://doi.org/10.20944/preprints202004.0084.v1>
- Benedictow, O. J. (2011). *La peste negra, 1346-1353: la historia completa (Vol. 316)*. Ediciones Akal.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research*, 28(1), 235-242. <https://doi.org/10.1093/nar/28.1.235>
- Bhowmik, D., Sharma, R. D., Prakash, A., & Kumar, D. (2021). "Identification of Nafamostat and VR23 as COVID-19 drug candidates by targeting 3CLpro and PLpro". *Journal of Molecular Structure*, 1233, 130094. <https://doi.org/10.1016/j.molstruc.2021.130094>
- Bouysset, C., & Fiorucci, S. (2021). ProLIF: A library to encode molecular interactions as fingerprints. *Journal of Cheminformatics*, 13(1), 72. <https://doi.org/10.1186/s13321-021-00548-6>
- Calleja, D. J., Lessene, G., & Komander, D. (2022). Inhibitors of SARS-CoV-2 PLpro. *Frontiers in Chemistry*, 10, 876212. <https://doi.org/10.3389/fchem.2022.876212>
- Cardona-Arias, J. A., & Higuera-Gutiérrez, L. F. (2014). Impacto del VIH/SIDA sobre la calidad de vida: Metaanálisis 2002-2012. *Revista Española de Salud Pública*, 88(1), 87-101. <https://doi.org/10.4321/S1135-57272014000100006>
- Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., & Woods, R. J. (2005). The Amber biomolecular simulation programs. *Journal of Computational Chemistry*, 26(16), 1668-1688. <https://doi.org/10.1002/jcc.20290>
- Cheng, V. C., Lau, S. K., Woo, P. C., & Yuen, K. Y. (2007). Severe Acute Respiratory Syndrome Coronavirus as an Agent of Emerging and Reemerging Infection. *Clinical Microbiology Reviews*, 20(4), 660-694. <https://doi.org/10.1128/CMR.00023-07>

- 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>
- Groot, R. J., Baker, S. C., Baric, R. S., Brown, C. S., Drosten, C., Enjuanes, L., Fouchier, R. A., Galiano, M., Gorbalenya, A. E., Memish, Z. A., Perlmann, S., Poon, L. L., Snijder, E. J., Stephens, G. M., Woo, P. C., Zaki, A. M., Zambon, M., & Ziebuhr, J. (2013). Commentary: Middle East Respiratory Syndrome Coronavirus (MERS-CoV): Announcement of the Coronavirus Study Group. *Journal of Virology*, 87(14), 7790-7792. <https://doi.org/10.1128/JVI.01244-13>
- Drosten, C., Günther, S., Preiser, W., van der Werf, S., Brodt, H.-R., Becker, S., Rabenau, H., Panning, M., Kolesnikova, L., Fouchier, R. A. M., Berger, A., Burguière, A.-M., Cinatl, J., Eickmann, M., Escriou, N., Grywna, K., Kramme, S., Manuguerra, J.-C., Müller, 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>
- Fu, Z., Huang, B., Tang, J., Liu, S., Liu, M., Ye, Y., Liu, Z., Xiong, Y., Zhu, W., Cao, D., Li, J., Niu, X., Zhou, H., Zhao, Y. J., Zhang, G., & Huang, H. (2021). The complex structure of GRL0617 and SARS-CoV-2 PLpro reveals a hot spot for antiviral drug discovery. *Nature Communications*, 12(1), 488. <https://doi.org/10.1038/s41467-020-20718-8>
- Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., & Shenkin, P. S. (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>
- Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., Sanschagrin, P. C., & Mainz, D. T. (2006). Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *Journal of Medicinal Chemistry*, 49(21), 6177-6196. <https://doi.org/10.1021/jm051256o>
- Gallo, M. I. P. (2020). La gripe española: 1918-1919. Los libros de la catarata.
- 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>
- González-Alemán, R., Hernández-Castillo, D., Rodríguez-Serradet, A., Caballero, J., Hernández-Rodríguez, E. W., & Montero-Cabrera, L. (2020). BitClust: Fast Geometrical Clustering of Long Molecular Dynamics Simulations. *Journal of Chemical Information and Modeling*, 60(2), 444-448. <https://doi.org/10.1021/acs.jcim.9b00828>
- Gorbalenya, A. E., Baker, S. C., Baric, R. S., de Groot, R. J., Drosten, C., Gulyaeva, A. A., Haagmans, B. L., Lauber, C., Leontovich, A. M., Neuman, B. W., Penzar, D., Perlman, S., Poon, L. L. M., Samborskiy, D. V., Sidorov, I. A., Sola, I., Ziebuhr, J., & Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. (2020). The species severe acute respiratory syndrome-related coronavirus:

Classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology*, 5(4), 536-544. <https://doi.org/10.1038/s41564-020-0695-z>

- Hamelberg, D., Mongan, J., & McCammon, J. A. (2004). Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *The Journal of Chemical Physics*, 120(24), 11919-11929. <https://doi.org/10.1063/1.1755656>
- Harcourt, B. H., Jukneliene, D., Kanjanahaluethai, A., Bechill, J., Severson, K. M., Smith, C. M., Rota, P. A., & Baker, S. C. (2004). Identification of Severe Acute Respiratory Syndrome Coronavirus Replicase Products and Characterization of Papain-Like Protease Activity. *Journal of Virology*, 78(24), 13600-13612. <https://doi.org/10.1128/JVI.78.24.13600-13612.2004>
- Henzler-Wildman, K., & Kern, D. (2007). Dynamic personalities of proteins. *Nature*, 450(7172), 964-972. <https://doi.org/10.1038/nature06522>
- Hilgenfeld, R., Anand, K., Mesters, J. R., Rao, Z., Shen, X., Jiang, H., Tan, J., & Verschuere, K. H. G. (2006). Structure and Dynamics of Sars Coronavirus Main Proteinase (MPRO). In S. Perlman & K. V. Holmes (Eds.), *The Nidoviruses* (Vol. 581, pp. 585-591). Springer US. [https://doi.org/10.1007/978-0-387-33012-9\\_106](https://doi.org/10.1007/978-0-387-33012-9_106)
- Huang, C., Wei, P., Fan, K., Liu, Y., & Lai, L. (2004). 3C-like Proteinase from SARS Coronavirus Catalyzes Substrate Hydrolysis by a General Base Mechanism. *Biochemistry*, 43(15), 4568-4574. <https://doi.org/10.1021/bi036022q>
- Kalliokoski, T., Kramer, C., Vulpetti, A., & Gedeck, P. (2013). Comparability of Mixed IC50 Data – A Statistical Analysis. *PLOS ONE*, 8(4), e61007. <https://doi.org/10.1371/journal.pone.0061007>
- Knapp, B., Lederer, N., Omasits, U., & Schreiner, W. (2010). vmdICE: A plug-in for rapid evaluation of molecular dynamics simulations using VMD. *Journal of Computational Chemistry*, 31(16), 2868-2873. <https://doi.org/10.1002/jcc.21581>
- Kokh, D. B., Doser, B., Richter, S., Ormersbach, F., Cheng, X., & Wade, R. C. (2020). A workflow for exploring ligand dissociation from a macromolecule: Efficient random acceleration molecular dynamics simulation and interaction fingerprint analysis of ligand trajectories. *The Journal of Chemical Physics*, 153(12), 125102. <https://doi.org/10.1063/5.0019088>
- Lewis, D. S. M., Ho, J., Wills, S., Kawall, A., Sharma, A., Chavada, K., Ebert, M. C. C. J. C., Evoli, S., Singh, A., Rayalam, S., Mody, V., & Taval, S. (2022). Aoin isoforms (A and B) selectively inhibits proteolytic and deubiquitinating activity of papain like protease (PLpro) of SARS-CoV-2 in vitro. *Scientific Reports*, 12(1), 2145. <https://doi.org/10.1038/s41598-022-06104-y>
- Li, F. (2016). Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annual Review of Virology*, 3(1), 237-261. <https://doi.org/10.1146/annurev-virology-110615-042301>
- 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>
- 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), 93. <https://doi.org/10.1038/s42003-020-01577-x>
- Morens, D. M., Folkers, G. K., & Fauci, A. S. (2009). What Is a Pandemic? *The Journal of Infectious Diseases*, 200(7), 1018-1021. <https://doi.org/10.1086/644537>

- Narayanan, A., Narwal, M., Majowicz, S. A., Varricchio, C., Toner, S. A., Ballatore, C., Brancale, A., Murakami, K. S., & Jose, J. (2022). Identification of SARS-CoV-2 inhibitors targeting Mpro and PLpro using in-cell-protease assay. *Communications Biology*, 5(1), 169. <https://doi.org/10.1038/s42003-022-03090-9>
- Patel, R., Prajapati, J., Rao, P., Rawal, R. M., Saraf, M., & Goswami, D. (2021). Repurposing the antibacterial drugs for inhibition of SARS-CoV2-PLpro using molecular docking, MD simulation and binding energy calculation. *Molecular Diversity*. <https://doi.org/10.1007/s11030-021-10325-0>
- 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. <https://www.frontiersin.org/article/10.3389/fchem.2020.623971>
- Ratia, K., Saikatendu, K. S., Santarsiero, B. D., Barretto, 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*, 103(15), 5717-5722. <https://doi.org/10.1073/pnas.0510851103>
- Ratia, K., Pegan, S., Takayama, J., Sleeman, K., Coughlin, M., Baliji, S., Chaudhuri, R., Fu, W., Prabhakar, B. S., Johnson, M. E., Baker, S. C., Ghosh, A. K., & Mesecar, A. D. (2008). A noncovalent class of papain-like protease/deubiquitinase inhibitors blocks SARS virus replication. *Proceedings of the National Academy of Sciences*, 105(42), 16119-16124. <https://doi.org/10.1073/pnas.0805240105>
- Ratia, K., Kilianski, A., Baez-Santos, Y. M., Baker, S. C., & Mesecar, A. (2014). Structural Basis for the Ubiquitin-Linkage Specificity and deISGylating Activity of SARS-CoV Papain-Like Protease. *PLOS Pathogens*, 10(5), e1004113. <https://doi.org/10.1371/journal.ppat.1004113>
- Roberts, E., Eargle, J., Wright, D., & Luthey-Schulten, Z. (2006). MultiSeq: Unifying sequence and structure data for evolutionary analysis. *BMC Bioinformatics*, 7(1), 382. <https://doi.org/10.1186/1471-2105-7-382>
- Roe, D. R., & Cheatham, T. E. (2013). PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation*, 9(7), 3084-3095. <https://doi.org/10.1021/ct400341p>
- Salmaso, V., Sturlese, M., Cuzzolin, A., & Moro, S. (2017). Exploring Protein-Peptide Recognition Pathways Using a Supervised Molecular Dynamics Approach. *Structure*, 25(4), 655-662.e2. <https://doi.org/10.1016/j.str.2017.02.009>
- Salmaso, V., & Moro, S. (2018). Bridging Molecular Docking to Molecular Dynamics in Exploring Ligand-Protein Recognition Process: An Overview. *Frontiers in Pharmacology*, 9. <https://www.frontiersin.org/article/10.3389/fphar.2018.00923>
- Sheahan, T. P., Sims, A. C., Graham, R. L., Menachery, V. D., Gralinski, L. E., Case, J. B., Leist, S. R., Pirc, K., Feng, J. Y., Trantcheva, I., Bannister, R., Park, Y., Babusis, D., Clarke, M. O., Mackman, R. L., Spahn, J. E., Palmiotti, C. A., Siegel, D., Ray, A. S., ... Baric, R. S. (2017). Broad-spectrum antiviral GS-5734 inhibits both

- epidemic and zoonotic coronaviruses. *Science Translational Medicine*, 9(396), eaal3653. <https://doi.org/10.1126/scitranslmed.aal3653>
- Shen, Z., Ratia, K., Cooper, L., Kong, D., Lee, H., Kwon, Y., Li, Y., Alqarni, S., Huang, F., Dubrovskiy, O., Rong, L., Thatcher, G. R. J., & Xiong, R. (2022). Design of SARS-CoV-2 PLpro Inhibitors for COVID-19 Antiviral Therapy Leveraging Binding Cooperativity. *Journal of Medicinal Chemistry*, 65(4), 2940-2955. <https://doi.org/10.1021/acs.jmedchem.1c01307>
  - Tian, C., Kasavajhala, K., Belfon, K. A. A., Raguette, L., Huang, H., Miguez, A. N., Bickel, J., Wang, Y., Pincay, J., Wu, Q., & Simmerling, C. (2020). ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *Journal of Chemical Theory and Computation*, 16(1), 528-552. <https://doi.org/10.1021/acs.jctc.9b00591>
  - Velázquez-Libera, J. L., Murillo-López, J. A., F. de la Torre, A., & Caballero, J. (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), 6133. <https://doi.org/10.3390/ijms20246133>
  - Velázquez-Libera, J. L., Durán-Verdugo, F., 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>
  - Zhang, Y., & Skolnick, J. (2005). TM-align: A protein structure alignment algorithm based on the TM-score. *Nucleic Acids Research*, 33(7), 2302-2309. <https://doi.org/10.1093/nar/gki524>
  - Zheng, J. (2020). SARS-CoV-2: An Emerging Coronavirus that Causes a Global Threat. *International Journal of Biological Sciences*, 16(10), 1678-1685. <https://doi.org/10.7150/ijbs.45053>
  - Zowalaty, M. E., & Järhult, J. D. (2020). From SARS to COVID-19: A previously unknown SARS- related coronavirus (SARS-CoV-2) of pandemic potential infecting humans – Call for a One Health approach. *One Health*, 9, 100124. <https://doi.org/10.1016/j.onehlt.2020.100124>