

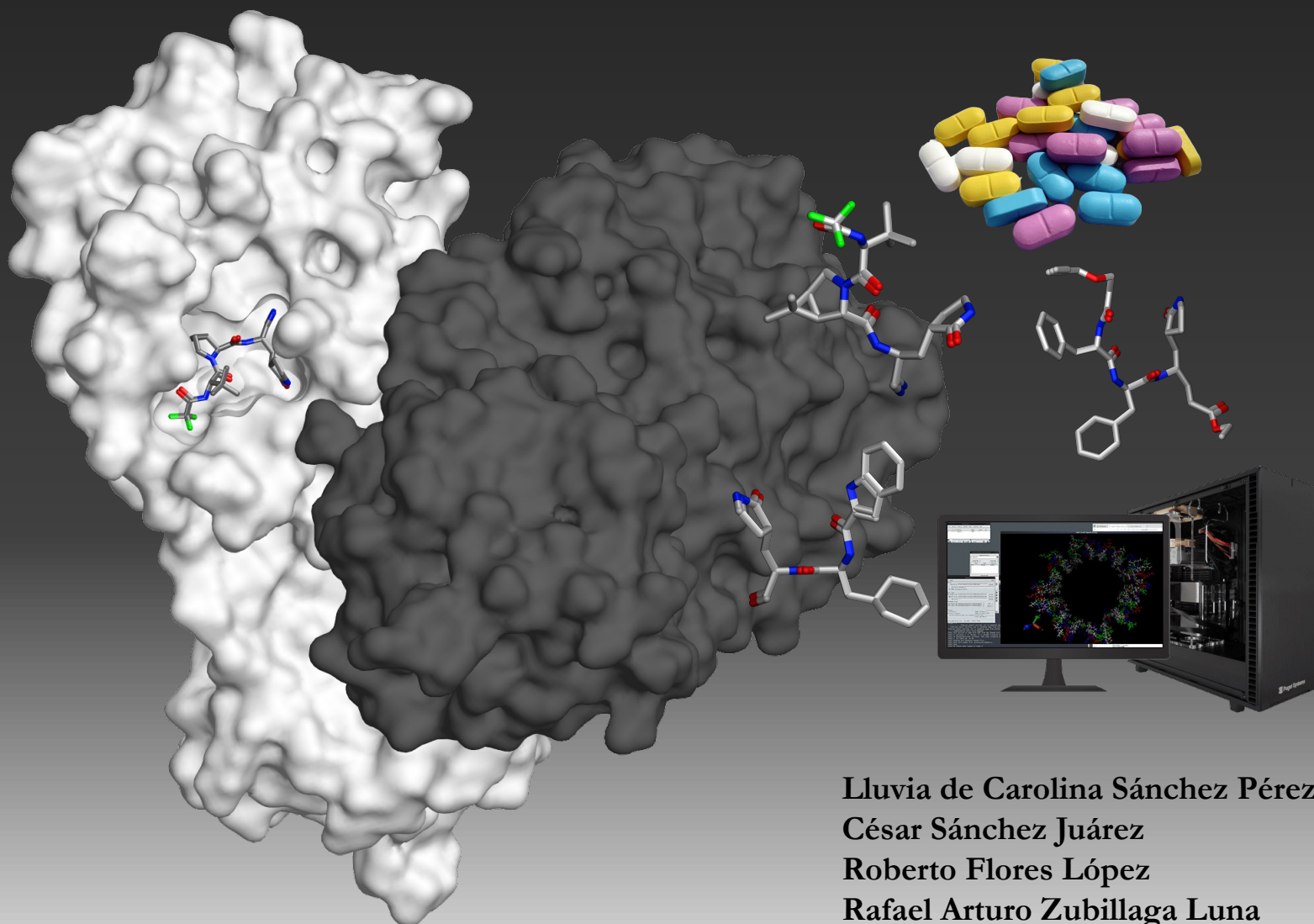


Casa abierta al tiempo

UNIVERSIDAD AUTÓNOMA METROPOLITANA

Colección CBI
Manual de prácticas

Taller de bioquímica computacional para el análisis de proteínas y compuestos bioactivos



Lluvia de Carolina Sánchez Pérez
César Sánchez Juárez
Roberto Flores López
Rafael Arturo Zubillaga Luna
Ponciano García Gutiérrez



Taller de bioquímica computacional

para el análisis de proteínas y compuestos bioactivos

Lluvia de Carolina Sánchez Pérez
César Sánchez Juárez
Roberto Flores López
Rafael Arturo Zubillaga Luna
Ponciano García Gutiérrez

Manual de prácticas



Casa abierta al tiempo

Dr. Gustavo Pacheco López
Rector General

Dra. Esthela Irene Sotelo Núñez
Secretaria General

UNIDAD IZTAPALAPA

Dra. Edith Ponce Alquicira
Rectora de Unidad

Dr. Javier Rodríguez Lagunas
Secretario de Unidad

Dr. Román Linares Romero
Director de la División de Ciencias Básicas e Ingeniería

Taller de Bioquímica Computacional para el Análisis de Proteínas y Compuestos Bioactivos

Primera edición: 2026

© UNIVERSIDAD AUTÓNOMA METROPOLITANA-IZTAPALAPA

Av. Ferrocarril San Rafael Atlixco 186, colonia Leyes de Reforma, Alcaldía Iztapalapa, C. P. 09310, Ciudad de México

ISBN Colección: 978-607-28-2107-1

ISBN Volumen: 978-607-28-3735-5





Índice general

1	Bioinformática: Utilización de las bases de datos para el estudio de proteínas	9
1.1	Ingreso al GenBank para recuperar la secuencia de aminoácidos de la proteasa principal de algunos coronavirus	10
1.2	Ingreso a la base de datos UniProt	13
1.3	Estimación de propiedades fisicoquímicas con ProtParam	15
1.4	Alineamiento de secuencias	18
2	Visualización de proteínas	21
3	Examinando interacciones covalentes y no covalentes en un complejo proteína–ligando relacionado con el COVID-19	33
4	Modelado de la estructura tridimensional de la Mpro dimérica del SARS-CoV-2	42
4.1	Identificación de la plantilla	44
4.2	Construcción del modelo de la SCV2–Mpro dimérica	49
4.3	Validación y comparación del modelo	51
5	Construcción y análisis de la flexibilidad conformacional del 1,3-dimetilciclohexano	55
5.1	Construyendo la molécula en UCSF Chimera	56
5.2	Flexibilidad estructural del 1,3-dimetilciclohexano	56
5.3	Obtención del confórmero de menor energía	59
6	Generación de conformaciones de una proteína mediante simulación de dinámica molecular	63
6.1	Ajuste de los estados de protonación	65
6.2	Preparación del sistema para la simulación dinámica molecular	68
6.2.1	Conversión del formato .pdb a .gmx y generación de la topología	68
6.2.2	Construcción de la caja de simulación y solvatación	71
6.2.3	Solvatación del proteína y llenado de la caja con moléculas de agua	72
6.2.4	Adición de iones	74
6.2.5	Minimización de la energía	75
6.2.6	Estabilización del sistema	77
6.2.7	Ejecución de la simulación de dinámica molecular	83



6.3	Análisis posterior a la simulación	85
7	Análisis estructural de la trayectoria de simulación de dinámica molecular de una proteína	88
7.1	Generación de clústeres	93
8	Acoplamiento molecular y cribado virtual	96
8.1	Preparación del receptor	98
8.2	Cribado virtual: obtención y edición del archivo batch	103
8.3	Ejecutando el cribado virtual	105
9	Diseño de inhibidores para la Mpro del SARS-CoV-2 utilizando un enfoque basado en fragmentos	110
10	Predicción de biodisponibilidad y propiedades farmacológicas de ligandos	123
10.1	Predicción de características con SwissADME	124
10.2	Predicción de propiedades fisicoquímicas y farmacocinéticas	125
10.3	Interpretación de resultados	126
10.4	Predicción de características con ADMETlab 3.0	128
10.5	Predicción de toxicidad con ProTox-3	130
A	Apéndice A. Guía para la instalación de software para la realización del Taller de Bioquímica Computacional para el Análisis de Proteínas y Compuestos Bioactivos	136
B	Apéndice B. Metodología para la selección del parámetro -ng en TTClust	147



Prefacio

El estudio de la estructura y función de las proteínas se ha convertido en una práctica indispensable en diversos campos multidisciplinarios y sectores industriales, particularmente aquellos relacionados con la salud, la biotecnología y los alimentos. Comprender cómo interactúan y se organizan las proteínas es fundamental para modificar enzimas y adaptarlas a nuevos procesos, diseñar anticuerpos terapéuticos y vacunas, así como para optimizar la alergenicidad o el valor nutricional de proteínas alimentarias.

Este taller tiene como propósito guiar a los participantes en la exploración de herramientas computacionales de libre acceso, ampliamente utilizadas en la investigación científica para el análisis, estudio y modelado de estructuras proteicas. Asimismo, permite trabajar con moléculas orgánicas pequeñas, analizando sus conformaciones, su interacción con proteínas mediante acoplamiento molecular (*docking*) y la predicción de propiedades farmacocinéticas como absorción, distribución, metabolismo, excreción y toxicidad (ADMET).

Las actividades propuestas han sido concebidas como un apoyo didáctico para las Unidades de Enseñanza-Aprendizaje (UEA) de Diseño de Fármacos, Bioquímica y Biología Molecular I, y Biofísicoquímica, impartidas en la Licenciatura en Química de la Universidad Autónoma Metropolitana, Unidad Iztapalapa. No obstante, este manual también puede ser de utilidad para estudiantes o investigadores principiantes interesados en introducirse en el fascinante mundo del análisis estructural de proteínas y su interacción con ligandos.

Durante el desarrollo del taller, se recomienda que cada participante lleve un registro detallado de las actividades realizadas. En cada práctica se han incluido –resaltadas en fondo gris– preguntas clave y ejercicios que deberán ser respondidos y documentados para su evaluación posterior.

Confiamos en que este manual sea no solo una herramienta útil, sino también una invitación a adentrarse en el apasionante campo del diseño de fármacos asistido por computadora.

Atentamente

Los Autores





Práctica 1. Bioinformática: Utilización de las bases de datos para el estudio de proteínas

Objetivos

- Identificar y recuperar secuencias de proteínas de interés a partir de bases de datos públicas como *GenBank* y *UniProt*.
- Estimar e interpretar propiedades fisicoquímicas básicas a partir de la secuencia utilizando *ProtParam*.
- Identificar regiones conservadas y variaciones evolutivas mediante el alineamiento múltiple de secuencias homólogas utilizando *Clustal Omega*.

En las últimas décadas, los avances en biología molecular y el desarrollo de proyectos genómicos y proteómicos a gran escala han generado un crecimiento exponencial de la información biológica. Este incremento exige almacenar, organizar y analizar grandes volúmenes de datos mediante herramientas especializadas. La disciplina encargada de integrar y gestionar estos recursos, incluidas las secuencias de ácidos nucleicos y proteínas, es la bioinformática.

Una de las principales fuentes de información en bioinformática son las bases de datos públicas, las cuales almacenan registros altamente especializados y accesibles en línea. Una base de datos es una recopilación organizada de datos estructurados, almacenada electrónicamente y gestionada por programas que permiten su recuperación y actualización constante. En el ámbito de las proteínas, resultan esenciales aquellas que resguardan secuencias de aminoácidos junto con metadatos experimentales y funcionales. Estas plataformas no solo almacenan las secuencias, sino que también incorporan algoritmos avanzados para búsquedas rápidas y proporcionar información adicional, por ejemplo, dominios estructurales, propiedades fisicoquímicas o alineamientos, a partir de las secuencias depositadas. Entre las bases de datos más importantes se encuentran:

- *GenBank*: base de datos de secuencias nucleotídicas (ADN y ARN) anotadas, que reúne prácticamente toda la información de secuencias disponibles públicamente. Es mantenida y actualizada por los Institutos Nacionales de Salud de los Estados Unidos de América (NIH), a través del NCBI, y forma parte de la International Nucleotide Sequence Database Collaboration junto con DDBJ (Japón) y ENA (Europa). Su acceso está disponible en: <https://www.ncbi.nlm.nih.gov/genbank/>
- *UniProt* (Universal Protein Resource): un repositorio central de datos sobre proteínas, resultado de la fusión de Swiss-Prot, TrEMBL y PIR. Esta base de datos es no redundante y contiene anotaciones verificadas. Su acceso está disponible a través de <https://www.uniprot.org/>

En esta práctica se utilizarán ambas bases de datos para consultar información relacionada con la proteasa principal (Mpro) de algunos coronavirus, con el propósito de:

- 
- I. Obtener sus secuencias de aminoácidos
 - II. Estimar propiedades fisicoquímicas básicas
 - III. Realizar un alineamiento múltiple de las secuencias –proceso para comparar similitudes y diferencias entre secuencias de aminoácidos– para profundizar en su conservación evolutiva

1.1 Ingreso al *GenBank* para recuperar la secuencia de aminoácidos de la proteasa principal de algunos coronavirus

En primer lugar, se debe acceder a la base de datos *GenBank* para recuperar la secuencia de aminoácidos de la Mpro, (también conocida como 3CL por *3-chymotrypsin-like*, debido a su similitud con la quimotripsina), del SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2), abreviada como SCV2–Mpro. Acceda al portal del NCBI (National Center for Biotechnology Information) en <https://www.ncbi.nlm.nih.gov/genbank/>. En el menú junto a la barra de búsqueda, seleccione la opción **Protein**, para asegurarse de que los resultados devuelvan la secuencia de aminoácidos y no la de nucleótidos. En el campo de búsqueda, escribir *SARS-CoV-2 Main protease* y presionar **Search** para iniciar la búsqueda (ver Figura 1.1).

Es importante tener en cuenta que, como en cualquier motor de búsqueda en internet, pueden aparecer múltiples resultados no relevantes que coincidan parcialmente con las palabras clave utilizadas.

Actividad 1.1. Responder las siguientes preguntas:

- ¿Cuántas secuencias de proteínas obtuvieron como resultado?
- ¿Por qué aparece un número tal elevado de resultados?
- ¿Son estas secuencias iguales o diferentes?

En la lista de resultados, localizar la entrada titulada **Chain A, 3C-like proteinase**, que es una de las múltiples entradas que contienen la información de interés, y hacer clic sobre ella. El servidor desplegará una página similar a la de la Figura 1.2. Es recomendable explorarla con detalle para relacionarse con su contenido.

Actividad 1.2. Comentar la información que se despliega en la pantalla y responder las siguientes preguntas:

- ¿Cuál es el título del artículo científico del que proviene esta secuencia?
- ¿Cuál es el organismo de origen?
- ¿Es posible identificar la secuencia de aminoácidos? En caso afirmativo, ¿cuál es el primer residuo y cuál es el último?

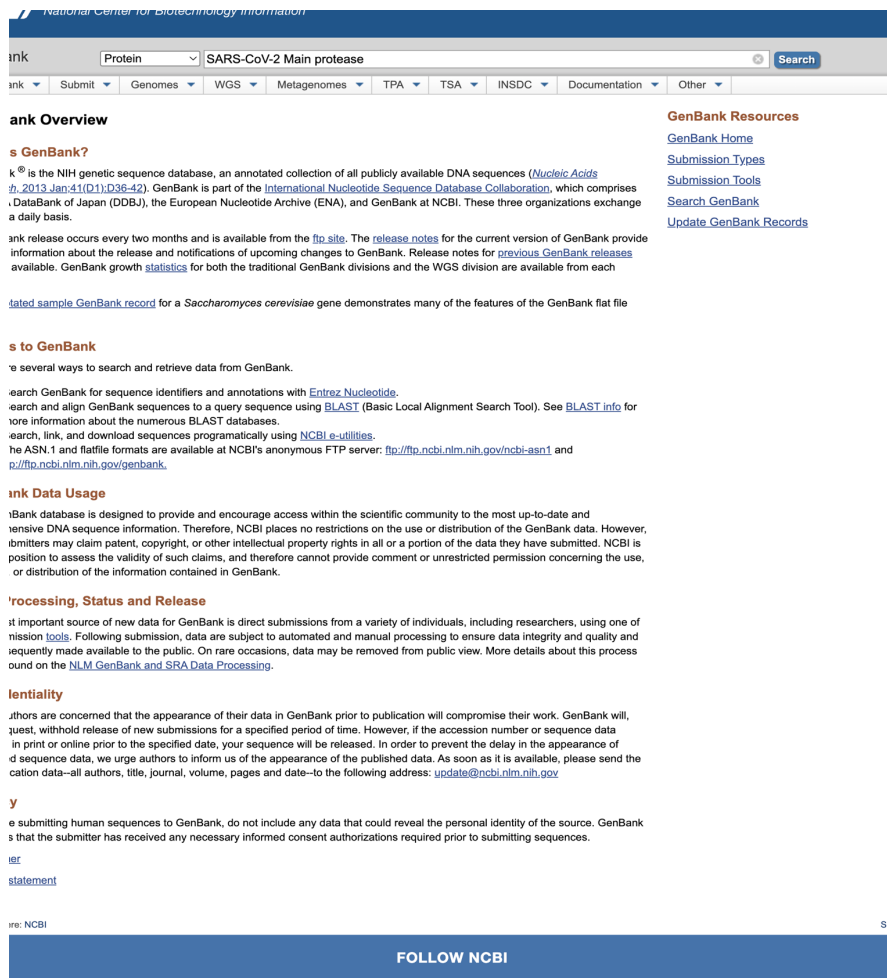


Figura 1.1. Captura de pantalla de la interfaz de la base de datos *GenBank*, utilizada para la búsqueda y recuperación de secuencias genéticas públicas. Se muestra un ejemplo del entorno de trabajo utilizado para localizar la secuencia de interés.

En la parte superior derecha, seleccione el enlace **FASTA** para visualizar la secuencia de aminoácidos en un formato estándar ampliamente utilizado para describir secuencias de proteínas y ácidos nucleicos, conocido como formato *fasta* (ver Figura 1.3).

Actividad 1.3. Copiar y pegar la secuencia en formato *fasta* de la SCV2-Mpro en un archivo de texto (WordPad, TextEdit, o Bloc de notas). Es importante señalar que la primera línea, conocida como *encabezado* (por ejemplo: >pdb|7JU7|A Chain A, 3C-like proteinase), forma parte del formato *fasta* y debe conservarse sin modificar.

¿Cómo se define el formato *fasta*?



NIH National Library of Medicine
National Center for Biotechnology Information

Protein

GenPept

Chain A, 3C-like proteinase

PDB: 7JUT_A
[Identical Proteins](#) [FASTA](#) [Graphics](#)

Go to:

LOCUS 7JUT_A 306 aa linear VRL 18-OCT-2023
DEFINITION Chain A, 3C-like proteinase.
ACCESSION 7JUT_A
VERSION 7JUT_A
DBSOURCE pdb: molecule 7JUT, chain A, release Oct 18, 2023;
deposition: Aug 19, 2020;
class: HYDROLASE;
source: Mmdb_id: 192276, Pdb_id 1: 7JUT;
Exp. method: X-ray Diffraction.

KEYWORDS Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)
SOURCE *Severe acute respiratory syndrome coronavirus 2*
ORGANISM Viruses; Riboviria; Orthornavirae; Pisuviricota; Pisoniviricetes;
Nidovirales; Coronavirineae; Coronaviridae; Orthocoronavirinae;
Betacoronavirus; Sarbecovirus; Betacoronavirus pandemicum.
1 (residues 1 to 306)

REFERENCE Drayman, N., Jones, K.A., Azizi, S.A., Froggatt, H.M., Tan, K.,
Maltseva, N.I., Chen, S., Nicolaescu, V., Dvorkin, S., Furlong, K.,
Kathayat, R.S., Firpo, M.R., Mastrodomenico, V., Bruce, E.A.,
Schmidt, M.M., Jedrzejczak, R., Muñoz-Alia, M.Á.;, Schuster, B.,
Nair, V., Botten, J.W., Brooke, C.B., Baker, S.C., Mounce, B.C.,
Heaton, N.S., Dickinson, B.C., Joachimiak, A., Randall, G. and Tay, S.
Drug repurposing screen identifies masitinib as a 3CLpro inhibitor
that blocks replication of SARS-CoV-2 <i>in vitro</i>
JOURNAL bioRxiv (2020)
PUBMED 32988976
REFERENCE 2 (residues 1 to 306)
AUTHORS Drayman, N., DeMarco, J.K., Jones, K.A., Azizi, S.A., Froggatt, H.M.,
Tan, K., Maltseva, N.I., Chen, S., Nicolaescu, V., Dvorkin, S.,
Furlong, K., Kathayat, R.S., Firpo, M.R., Mastrodomenico, V.,
Bruce, E.A., Schmidt, M.M., Jedrzejczak, R., Muñoz-Alia, M.Á.;,
Schuster, B., Nair, V., Han, K.Y., O'Brien, A., Tomatsidou, A.,
Meyer, B., Vignuzzi, M., Missiakas, D., Botten, J.W., Brooke, C.B.,
Lee, H., Baker, S.C., Mounce, B.C., Heaton, N.S., Severson, W.E.,
Palmer, K.E., Dickinson, B.C., Joachimiak, A., Randall, G. and Tay, S.
Masitinib is a broad coronavirus 3CL inhibitor that blocks
replication of SARS-CoV-2
JOURNAL Science 373 (6557), 931-936 (2021)
PUBMED 34285113
REFERENCE 3 (residues 1 to 306)
AUTHORS Tan, K., Maltseva, N.I., Welk, L.F., Jedrzejczak, R.P., Joachimiak, A.
and Center for Structural Genomics of Infectious Diseases (CSGID).
Direct Submission
JOURNAL Submitted (19-AUG-2020)
COMMENT The crystal structure of SARS-CoV-2 Main Protease in complex with
masitinib.

FEATURES Location/Qualifiers
source 1..306
organism="Severe acute respiratory syndrome coronavirus
2"

Analyze this sequence

Protein 3D Structure

Room-temperature X-ray
structure of SARS-CoV-2
main protease in complex
with masitinib
PDB: 9NSL
Source: Severe acute
respiratory syndrome
coronavirus 2
Method: X-ray Diffraction
Resolution: 2 Å

NCBI Virus

Related information

Recent activity
 Protein

Figura 1.2. Captura de pantalla de una entrada de *GenBank* correspondiente a la Mpro del SARS-CoV-2. Se observa información asociada a la proteína, incluyendo datos del gen, la especie de origen y las secuencias correspondientes de ADN, ARN mensajero y aminoácidos, entre otros elementos.

Repetir el procedimiento para obtener la secuencia de la Mpro de los coronavirus SARS-CoV-1 y MERS (este último con una letalidad mayor que la del SARS-CoV-2). Si la búsqueda resulta complicada, consultar las sugerencias de palabras clave al final de este protocolo. Al igual que en el caso de la SCV2-Mpro, se debe tener en cuenta que puede haber varias entradas que contengan la misma información correcta. Se deben pegar las tres secuencias obtenidas, una tras otra, en el mismo archivo de texto previamente generado. El documento final debería tener un formato similar al mostrado en la Figura 1.4.

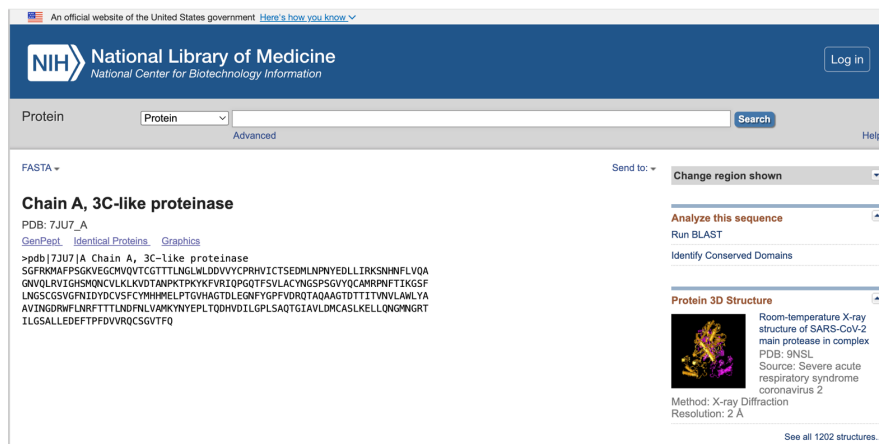


Figura 1.3. Captura de pantalla de la ventana de visualización de la secuencia de la proteína seleccionada en formato *fasta*. Este formato incluye un *encabezado* descriptivo seguido de la secuencia de aminoácidos en código de una sola letra, ampliamente utilizado en bioinformática para el análisis y procesamiento de secuencias.

```
>pdb|7DAV|A Chain A, COVID-19 MAIN PROTEASE
SGFRKMAFSPGKVEGCMVQVTCGTTTLNGLWLDVVYCPRHVICTSEDMLNPNYEDLLIRKSNHNLVQA
GNVQLRVIGHSMQNCVLKLVDTANPKTPKYKFVRIQPGQTFSLACYNNGSPSGVYQCAMPNFTIKGSF
LNGSCGSGVFNIDYDCVSFCYMHMELPTGVHAGTDLEGNFYGPVDRQTAQAAGTDTTITVNLAWLYA
AVINGDRWFLNRFTTTLNDFNLVAMKYNIEPLTQDHVDILGPLSAQTGIAVLDMCASLKELLQNGMNGRT
ILGSALLEDEFTPFDDVVRQCSGVTFQ

>pdb|4WMF|C Chain C, MERS-CoV 3CL protease
SGLVKMSHPDVEACMVQVTCGSMTLNGLWLDNTVWCPRHVMCPADQLSDPNYDALLISMTNHSFVQK
HIGAPANLRVVGHAMQGTLLKLTVDVANPSTPAYTFTTVKPGAASFVLACYNRPTGTFTVVMRPNYTIK
GSFLCGSAGSVGYTKESGVINFCYMHQMELANGHTGSAFDGTMYGAFMDKQVHQVQLTDKYCSVNVAW
LYAAILNGCAWFVKPNRTSVVSFNWALANQFTEFVGTQSDMLAVKTGVAIEQLLYAIQQLYTGFGQKQ
ILGSTMLEDEFTPDVNMQIMGVVMQ

>pdb|3TIU|A Chain A, SARS coronavirus main protease
SGFRKMAFSPGKVEGCMVQVTCGTTTLNGLWLDVTVCPRHVICTAEDMLNPNYEDLLIRKSNHNLVQA
GNVQLRVIGHSMQNCVLRKLVDTSNPKTPKYKFVRIQPGQTFSLACYNNGSPSGVYQCAMPNHTIKGSF
LNGSCGSGVFNIDYDCVSFCYMHMELPTGVHAGTDLEGFYGPVDRQTAQAAGTDTTITLNLAWLYA
AVINGDRWFLNRFTTTLNDFNLVAMKYNIEPLTQDHVDILGPLSAQTGIAVLDMCAALKELLQNGMNGRT
ILGSTILEDEFTPFDDVVRQCSGVTFQ
```

Figura 1.4. Ejemplo de un archivo de texto plano que contiene las secuencias en formato *fasta* de Mpro de los tres coronavirus identificados. Cada secuencia debe conservar su *encabezado* para facilitar su reconocimiento por programas de alineamiento múltiple de secuencias.

1.2 Ingreso a la base de datos *UniProt*

A continuación, se utilizará la segunda base de secuencias de proteínas mencionada: *UniProt*. Acceder al sitio <https://www.uniprot.org> e intentar localizar la secuencia de aminoácidos en formato *fasta* correspondiente a la SCV2-Mpro. Las mismas palabras clave utilizadas en *GenBank* pueden ser útiles; deben ser escritas en el campo de búsqueda de la página principal. Es importante notar que, nuevamente, se obtendrán múltiples entradas, aunque en menor cantidad que en *GenBank*, ya que *UniProt* es una

base de datos no redundante y curada, lo cual implica que sus secuencias han sido verificadas y anotadas por expertos antes de ser publicadas.

Actividad 1.4. Responder las siguientes preguntas:

¿Cuántas secuencias de proteínas se obtuvieron como resultado de la búsqueda? Registra el número total e indica, si aplica, si corresponden a variantes, isoformas o entradas redundantes.

¿Por qué no coinciden con las obtenidas en *GenBank*? Considera diferencias entre las bases de datos (tipo de entrada, criterios de anotación, presencia de isoformas, secuencias revisadas vs. no revisadas, filtros aplicados en la búsqueda, etc.). Comenta brevemente qué implicaciones tiene esto al momento de seleccionar secuencias para análisis posteriores.

Prestar especial atención a las cinco primeras entradas y observar el ícono con forma de carpeta dorada que aparece en la segunda columna, **Entry Name** (ver 1.5). Este símbolo indica que la entrada ha sido revisada (*reviewed*), es decir, verificada por expertos. La columna **Length** indica el número de aminoácidos que conforman a cada una de las proteínas. Se puede observar que la primera entrada corresponde a la poliproteína 1 y la segunda a la poliproteína 2 del SARS-CoV-2, por lo que ninguna de ellas representa directamente la secuencia específica de la SCV2–Mpro. Para continuar, se debe hacer clic en la primera entrada con código **P0DTC1** y navegar a través de la información desplegada.

Entry	Entry Name	Protein Names	Gene Names	Organism	Length
P0DTC1	R1A_SARS2	Replicase polyprotein 1a[...]		Severe acute respiratory syndrome coronavirus 2 (2019-nCoV) (SARS-CoV-2)	4,405 AA
P0DTC1	R1A_SARS2	Replicase polyprotein 1a[...]	rep. 1a-1b	Severe acute respiratory syndrome coronavirus 2 (2019-nCoV) (SARS-CoV-2)	7,096 AA
P0C6U8	R1A_SARS	Replicase polyprotein 1a[...]	1a	Severe acute respiratory syndrome coronavirus (SARS-CoV)	4,382 AA
P0C6X7	R1A_SARS	Replicase polyprotein 1a[...]	rep. 1a-1b	Severe acute respiratory syndrome coronavirus (SARS-CoV)	7,073 AA
P0DTC2	SPIKE_SARS2	Spike glycoprotein[...]	S, 2	Severe acute respiratory syndrome coronavirus 2 (2019-nCoV) (SARS-CoV-2)	1,273 AA
A0A6H0JLS9	A0A6H0JLS9_SARS2	ORF1ab polyprotein	ORF1ab	Severe acute respiratory syndrome coronavirus 2 (2019-nCoV) (SARS-CoV-2)	7,096 AA
A0A8B1KEG5	A0A8B1KEG5_SARS2	ORF1ab polyprotein	ORF1ab	Severe acute respiratory syndrome coronavirus 2 (2019-nCoV) (SARS-CoV-2)	6,961 AA
A0A6G7KRG6	A0A6G7KRG6_SARS2	Replicase polyprotein	ORF1ab	Severe acute respiratory syndrome	4,405 AA

Figura 1.5. Captura de pantalla de la interfaz de la base de datos *UniProt* utilizada para la búsqueda de secuencias e información de proteínas. La información se organiza en columnas que incluyen el identificador de entrada, el nombre de la proteína, el gen correspondiente y otros datos relevantes.

La información se organiza en distintas secciones, accesibles desde el panel izquierdo, de acuerdo con su naturaleza. Hacer clic en **Function, Names & Taxonomy, Subcellular Location** para familiarizarse con el contenido que ofrece la base de datos.

A continuación, dirigirse a la sección **PTM/Processing**, que contiene información sobre modificaciones pos-traduccionales y procesamiento proteico. Es posible acceder a ella directamente mediante el enlace correspondiente. Se debe analizar el contenido de esta sección, el cual deberá de ser similar al mostrado en la Figura 1.6.

TYPE	ID	POSITION(S)	DESCRIPTION
Chain	PRO_0000449635	1-180	Host translation inhibitor nsp1
Chain	PRO_0000449634	1-4405	Replicase polypeptide 1a
Chain	PRO_0000449636	181-818	Non-structural protein 2
Chain	PRO_0000449637	819-2763	Papain-like protease nsp3
Disulfide bond		2263--2291	PROSITE-ProRule Annotation
Disulfide bond		2282--2288	PROSITE-ProRule Annotation
Chain	PRO_0000449638	2764-3263	Non-structural protein 4
Chain	PRO_0000449639	3264-3569	3C-like proteinase nsp5
Chain	PRO_0000449640	3570-3859	Non-structural protein 6
Chain	PRO_0000449641	3860-3942	Non-structural protein 7
Chain	PRO_0000449642	3943-4140	Non-structural protein 8
Chain	PRO_0000449643	4141-4253	RNA-capping enzyme subunit nsp9
Chain	PRO_0000449644	4254-4392	Non-structural protein 10
Chain	PRO_0000449645	4393-4405	Non-structural protein 11

Figura 1.6. Captura de pantalla de la sección **PTM/Processing** de la entrada P0DTC1 en *UniProt*, donde se muestran las modificaciones pos-traduccionales y el procesamiento proteico asociados a la proteína seleccionada.

Actividad 1.5. Responder las siguientes preguntas:

¿La información contenida en esta entrada es la misma que la reportada en cualquiera de las entradas localizadas en *GenBank*?

Analizar el contenido de la sección **PTM/Processing** y anotar comentarios al respecto. ¿Qué tipo de modificaciones postraduccionales (PTM) o eventos de procesamiento se describen? ¿Podrían tener implicaciones funcionales o estructurales para la SCV2-Mpro?

Tratar de recuperar la secuencia en formato *fasta* para la SCV2-Mpro y compararla con la obtenida en *GenBank*. Determinar si ambas secuencias son iguales o presentan diferencias, e indicar brevemente a qué podrían deberse.

1.3 Estimación de propiedades fisicoquímicas con *ProtParam*

ExPASy (<https://www.expasy.org/>) es un meta-servidor que proporciona acceso a múltiples recursos bioinformáticos para análisis de secuencias y estructuras proteicas. Fue creado en 1993 y es mantenido por el Instituto Suizo de Bioinformática (SIB). En esta actividad se utilizará la herramienta *ProtParam* para estimar propiedades

fisicoquímicas básicas a partir de la secuencia de la SCV2–Mpro. Dirigirse a la página de *ExPASy*, escribir la palabra *ProtParam* en el campo de búsqueda y luego hacer clic en el ícono de lupa para iniciar la búsqueda. Luego, seleccionar los enlaces correspondientes hasta llegar a la interfaz que se muestra en la Figura 1.7.

Figura 1.7. Captura de pantalla de la interfaz de la herramienta *ProtParam* en el servidor *ExPASy*, utilizada para el análisis teórico de secuencias proteicas. La secuencia puede ingresarse mediante un número de acceso o pegándola directamente en el recuadro para realizar el análisis.

Actividad 1.6. Explorar al menos tres herramientas bioinformáticas disponibles en *ExPASy* distintas de *ProtParam*, y describir brevemente su utilidad. Se sugiere seleccionar herramientas enfocadas en análisis de secuencias, identificación de dominios, predicción estructural o caracterización funcional.

Copiar la secuencia de la SCV2–Mpro desde el archivo de texto previamente creado (excluir el *encabezado*, por ejemplo, `>pdb|7DAV| A...`) y pegarla en el recuadro de entrada de la interfaz de *ProtParam*. A continuación, hacer clic en **Compute parameters** para iniciar el análisis. Los resultados serán similares a los mostrados en la Figura 1.8.

Repetir el procedimiento para la secuencias de la Mpro correspondiente a los otros dos coronavirus. Los resultados obtenidos deben copiarse en el mismo archivo de texto previamente generado.

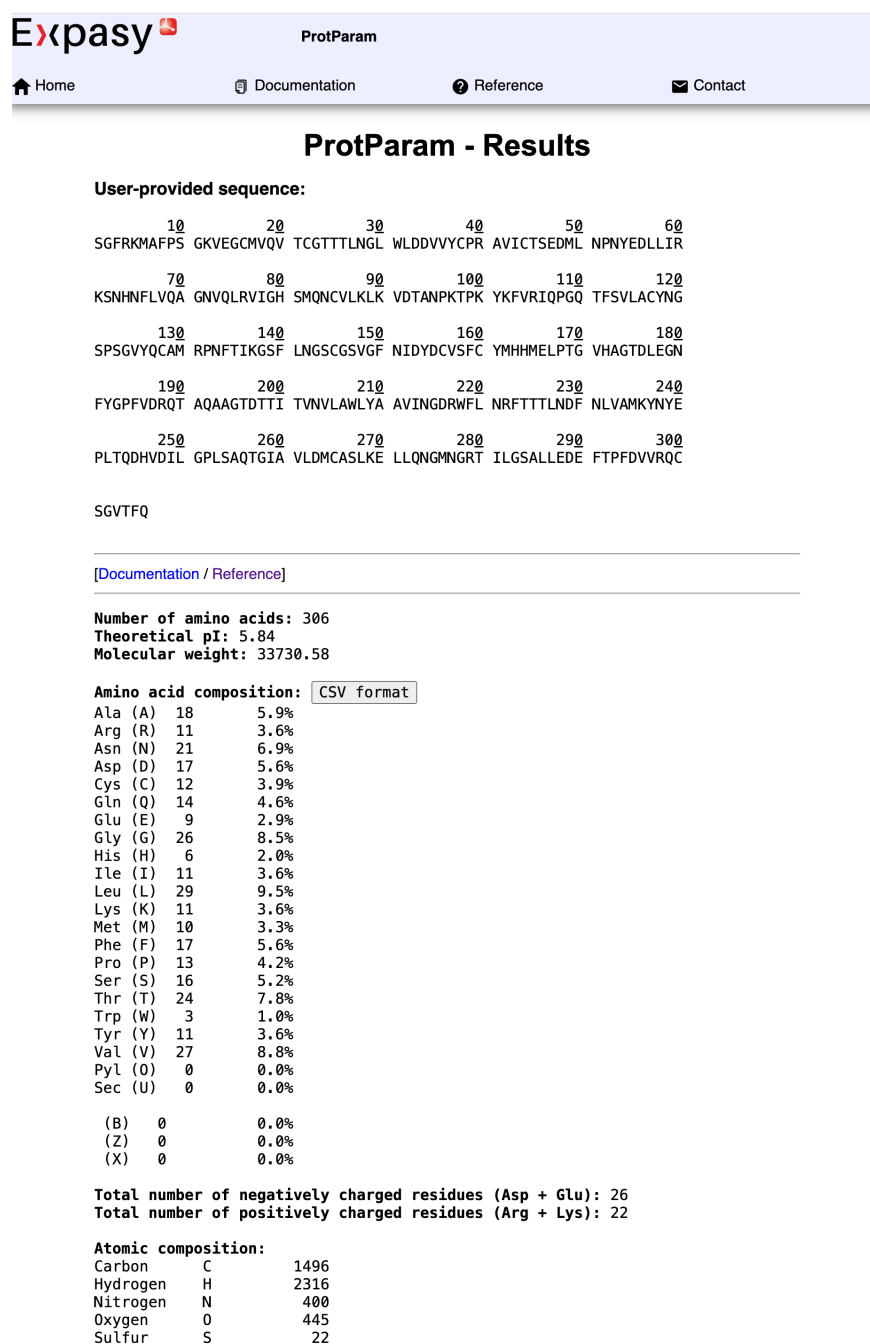


Figura 1.8. Captura de pantalla de los resultados del análisis bioinformático realizado con la herramienta *ProtParam* del servidor *ExPASy*. Se muestran propiedades fisicoquímicas estimadas para la SCV2-Mpro, como el peso molecular, el índice de inestabilidad, el punto isoelectrico teórico (pI), la composición de aminoácidos, entre otros parámetros relevantes.



Actividad 1.7. Analizar las semejanzas y diferencias en los parámetros obtenidos para las tres secuencias.

¿Por qué resulta importante estimar estos parámetros fisicoquímicos a partir de una secuencia proteica?

1.4 Alineamiento de secuencias

A continuación, se compararán las secuencias de la proteasa Mpro de los tres virus relacionados mediante un alineamiento múltiple. Este procedimiento es fundamental para identificar similitudes y diferencias entre proteínas que realizan la misma función, así como para localizar regiones conservadas útiles en el modelado estructural.

Existen diversos servidores que permiten realizar alineamientos de secuencias, accesibles desde *ExPASy* o directamente desde sus sitios web. Uno de los más utilizados es *Clustal Omega-Multiple Sequence Alignment* (o simplemente *Clustal Omega*), un algoritmo ampliamente empleado para efectuar alineamientos múltiples, incluso con miles de secuencias. Acceder a la versión alojada en <https://www.ebi.ac.uk/jdispatcher/msa/clustalo>.

Es importante familiarizarse con la interfaz del servidor. Pegar las secuencias de las tres Mpro (incluidos sus *encabezados*) en el recuadro **Paste your sequences here - or use the example sequence**. Luego, hacer clic en **Submit** para iniciar el alineamiento (ver Figura 1.9).

Copiar el alineamiento obtenido en el documento que contiene las secuencias en formato *fasta* de las Mpro. Posteriormente, acceder a la pestaña **Result Files** y localizar la sección **Percent Identity Matrix**. Al hacer clic en el enlace correspondiente, se mostrará la matriz de identidad (ver Figura 1.10), la cual debe copiarse en el mismo documento.

Actividad 1.8. Interpretar la mayor cantidad de detalles posibles en los resultados del alineamiento:

Explicar el significado del asterisco (*) que aparece debajo de cada posición alineada.

¿Qué indican los dos puntos (:) y el punto (.) en el alineamiento?

¿Qué implica que solo haya un espacio en blanco debajo de las secuencias?

¿Qué proteasas son más similares entre sí?

¿Cuáles son las más diferentes?

A continuación, repetir el alineamiento utilizando la versión de *Clustal Omega* alojada en *ExPASy* (*UniProt ClustalO*) y comparar los resultados con el alineamiento previo:

¿Existen diferencias entre ambos alineamientos?

¿Cuál es la función de la herramienta **Amino acid properties**?

Finalmente, entregar el archivo de texto generado con los resultados, análisis y comentarios a lo largo de la práctica.



II. SARS-CoV-1: ingresar *SARS coronavirus main protease*. Buscar la entrada **Chain A, SARS coronavirus main protease**

Bibliografía

- Akalin P. K.** (2006). Introduction to bioinformatics. *Molecular nutrition & food research*, 50(7), 610–619.
- Brandt G. S.** & Novak W. R. P. (2021). SARS-CoV-2 virtual biochemistry labs on bioinformatics and drug design. *Biochem Mol Biol Educ.* 49:26-28.
- Clark, K.,** Karsch-Mizrachi., D. J. Lipman., J. Ostell., E. W. Sayers. (2016). GenBank. *Nucleic Acids Res.* 2016 Jan 4;44. 2023.
- Gasteiger, E.,** Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., & Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic acids research* 31(13): 3784–3788.
- Pundir, S.,** Martin, M. J., & O'Donovan, C. (2017). UniProt Protein Knowledgebase. *Methods in molecular biology* (Clifton, N.J.) 1558: 41–55.
- Sayers, E. W.,** Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K. D., & Karsch-Mizrachi, I. (2019). GenBank. *Nucleic acids research* 47(D1), D94–D99.
- Wooller, S. K.,** Benstead-Hume, G., Chen, X., Ali, Y., & Pearl, F. M. G. (2017). Bioinformatics in translational drug discovery. *Bioscience reports* 37(4), BSR20160180.



Práctica 2. Visualización de proteínas

Objetivos

- Buscar y descargar archivos estructurales (*.pdb*) desde el repositorio *Protein Data Bank*.
- Visualizar y manipular modelos tridimensionales de proteínas usando UCSF Chimera para identificar elementos estructurales.
- Identificar interacciones no covalentes simples (puentes salinos) en estructuras proteicas mediante criterios geométricos.
- Alinear y comparar estructuras proteicas relacionadas, interpretando métricas cuantitativas como el RMSD.

Software especializado requerido

- UCSF Chimera versión 1.18 o superior

La Biología Estructural es una disciplina que se sitúa en la intersección de la Biología Molecular, la Bioquímica y la Biofísica, enfocada en el estudio de la estructura de macromoléculas biológicas —como las proteínas y los ácidos nucleicos—, así como de su origen y su relación con la función biológica. Entre las técnicas más utilizadas para determinar la estructura tridimensional de estas biomacromoléculas destacan la resonancia magnética nuclear (NMR), la microscopía electrónica (EM), y, de manera destacada, la cristalografía de rayos X (XRC).

El *Protein Data Bank* (PDB) (<https://www.rcsb.org/>) es un repositorio público que alberga, principalmente, estructuras tridimensionales de proteínas y ácidos nucleicos. Este recurso bioinformático, de acceso libre, es ampliamente utilizado por miles de investigadores en todo el mundo.

Hasta el 10 de noviembre de 2025, el PDB había registrado más de 244 mil estructuras depositadas, de las cuales aproximadamente 210 mil corresponden exclusivamente a proteínas. El resto incluye complejos proteína–oligosacárido, proteína–ácido nucleico, y estructuras de ácidos nucleicos. Esta cifra se actualiza diariamente y puede consultarse en el sitio web oficial del PDB.

Cada estructura es depositada por investigadores de todo el mundo y recibe un identificador único de cuatro caracteres, conocido como PDB ID, que generalmente está compuesto por un número seguido de tres letras. Este identificador permite descargar un archivo estándar que contiene las coordenadas tridimensionales de todos los átomos que conforman la biomacromolécula. Cabe señalar que, en la mayoría de los casos, las coordenadas de los átomos de hidrógeno no se incluyen, excepto en muchas estructuras determinadas por NMR. Esto se debe a que, en XRC, los hidrógenos dispersan muy débilmente la radiación y suelen ser difíciles de resolver experimentalmente, por lo que a menudo se omiten o se añaden posteriormente mediante métodos computacionales.

Existen diversos programas para visualizar estructuras biomoleculares, siendo VMD (Visual Molecular Dynamics, <https://www.ks.uiuc.edu/Research/vmd/>), PyMOL

(<https://www.pymol.org/>) y UCSF Chimera los más utilizados. En esta práctica se proporcionarán instrucciones detalladas para el uso de UCSF Chimera version 1.18, el cual puede descargarse gratuitamente desde <https://www.cgl.ucsf.edu/chimera/>. Estos programas permiten manipular las estructuras mediante la línea de comandos, utilizando un lenguaje de programación simple, o mediante la interfaz gráfica de usuario (GUI), utilizando el ratón, el teclado y los menús desplegables.

Para comenzar, se debe ingresar al sitio web del PDB. La página principal ofrece herramientas de búsqueda básica y avanzada, similares a las del *GenBank* o *UniProt*, que permiten localizar estructuras mediante distintos criterios, como el PDB ID, el nombre de la biomacromolécula, los autores, la secuencia de aminoácidos, o características estructurales específicas.

En años recientes, el PDB ha incorporado numerosas estructuras relacionadas con el SARS-CoV-2. Es importante recordar que previamente ya se habían registrado dos problemas de salud pública mundial asociados con SARS-CoV-1 (Severe acute respiratory syndrome coronavirus 1) y MERS (Middle East respiratory syndrome), y que muchas estructuras tridimensionales de biomacromoléculas relacionadas con estos virus ya habían sido depositadas. Esta acumulación de estructuras puede dificultar la búsqueda específica de aquellas correspondientes a cada virus en particular.

Para realizar una búsqueda dirigida, se debe localizar el campo de búsqueda básica en la página principal del PDB, escribir una palabra clave relacionada con la SCV2–Mpro –por ejemplo, *3CL protease*– y hacer clic en el ícono de lupa situado a la derecha de la barra de búsqueda (ver Figura 2.1).

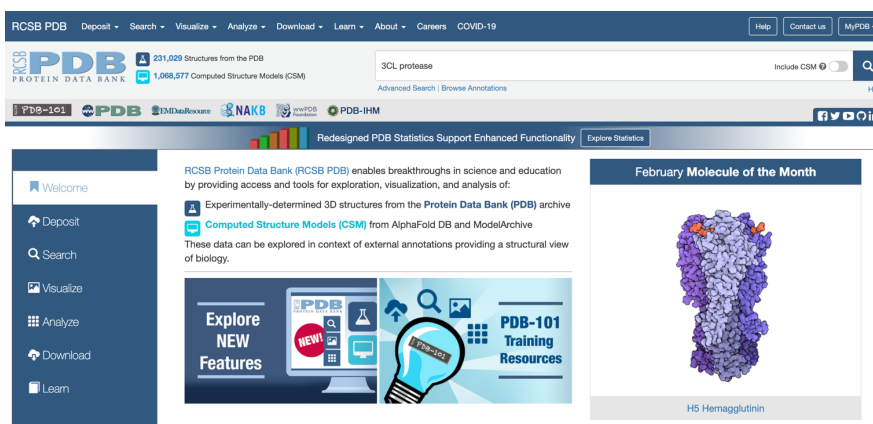


Figura 2.1. Captura de pantalla de la interfaz principal de la base de datos PDB, utilizada principalmente para la búsqueda y descarga de estructuras tridimensionales de proteínas y ácidos nucleicos. La página se actualiza de manera constante, por lo que su apariencia puede variar con el tiempo.

Actividad 2.1. Describir el resultado obtenido tras la búsqueda de una estructura relacionada con la SCV2-Mpro en el PDB. Además, responder las siguientes preguntas:

¿Cuántos resultados arrojó la búsqueda?

¿Todos ellos corresponden a la SCV2-Mpro?

¿Cuándo fue depositada y cuál es su PDB ID de la última estructura relacionada con la SCV2-Mpro?

A continuación, se explorará una de las tantas estructuras de la SCV2-Mpro que han sido depositadas recientemente en el PDB. Es importante recordar que el brote de COVID-19 surgió en Wuhan en noviembre de 2019, lo que dio lugar a un crecimiento exponencial en las investigaciones científicas y en la disponibilidad de datos estructurales sobre el virus.

Para comenzar, se debe regresar a la página principal del PDB. En el campo de búsqueda básica, escribir *6M2Q* y presionar **Enter** para iniciar la búsqueda. Esto conducirá a una página con los detalles de la estructura, similar a la mostrada en la Figura 2.2.

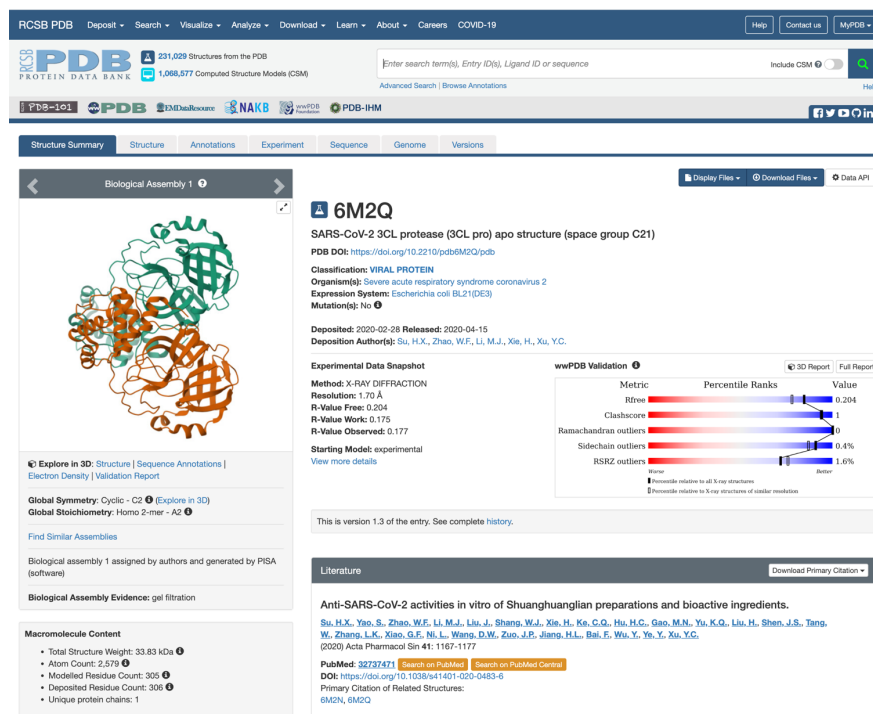


Figura 2.2. Captura de pantalla del resultado de búsqueda correspondiente a la SCV2-Mpro con PDB ID 6M2Q en la base de datos PDB. Se muestra información relevante como el organismo de origen, el sistema de expresión, la presencia de mutaciones, un resumen de los datos experimentales utilizados para resolver la estructura, así como una métrica de validación estructural. En el panel izquierdo de la interfaz se incluye una

visualización preliminar de la estructura en representación de cintas, donde se destacan elementos de estructura secundaria como hélices alfa y hebras beta.

Actividad 2.2. Responder las siguientes preguntas con base en la estructura seleccionada.

¿Cuál es el título del archivo de la estructura y cuál es el título del artículo científico asociado?

¿En qué año fue depositada esta estructura en el PDB?

¿Qué técnica experimental se utilizó para resolver la estructura tridimensional?

¿Cuál es el peso molecular de la proteína y cuántos átomos la constituyen?

¿Cuál es la resolución de la estructura? ¿Este valor indica una buena calidad estructural? Justifique brevemente su respuesta.

Para descargar la estructura, se debe hacer clic en el menú **Download files**, ubicado en la esquina superior derecha de la interfaz, y luego elegir **Legacy PDB format** en el menú desplegable.

Una vez descargado el archivo (*6m2q.pdb*), debe abrirse con un editor de texto como WordPad, TextEdit, o Bloc de notas. Se puede observar que el archivo contiene diversa información técnica, escrita en inglés y organizada por secciones relacionada con la metodología experimental utilizada para determinar la estructura, aspectos bioquímicos, y datos cristalográficos (por ejemplo, operadores de simetría), así como detalles estructurales. Una de las más relevantes es la sección marcada con la etiqueta **ATOM**, que contiene las coordenadas tridimensionales de los átomos que conforman la proteína (ver Figura 2.3).

SCALE3		0.000000	0.000000	0.022807	0.000000		
ATOM	1	N	SER	A	1	102.894	-9.765 26.870 1.00 38.53 N
ANISOU	1	N	SER	A	1	5452	4962 4227 -388 -1476 799 N
ATOM	2	CA	SER	A	1	103.698	-8.740 27.515 1.00 38.83 C
ANISOU	2	CA	SER	A	1	5320	5060 4373 -401 -1327 900 C
ATOM	3	C	SER	A	1	103.331	-8.580 28.990 1.00 37.66 C
ANISOU	3	C	SER	A	1	5009	4777 4521 -427 -1280 841 C
ATOM	4	O	SER	A	1	102.719	-9.469 29.585 1.00 35.30 O
ANISOU	4	O	SER	A	1	4726	4384 4301 -419 -1306 720 O
ATOM	5	CB	SER	A	1	103.552	-7.407 26.785 1.00 45.95 C
ANISOU	5	CB	SER	A	1	6160	5993 5306 -485 -1417 1097 C
ATOM	6	OG	SER	A	1	102.193	-7.026 26.690 1.00 51.46 O
ANISOU	6	OG	SER	A	1	6818	6536 6197 -562 -1614 1122 O
ATOM	7	N	GLY	A	2	103.702	-7.446 29.566 1.00 35.21 N

Figura 2.3. Captura de pantalla de la visualización del archivo *6m2q.pdb* abierto con un editor de textos. Se observa un fragmento de la sección **ATOM**, en la cual se especifican datos como el número del átomo, el tipo de átomo, el nombre del residuo al que pertenece, el identificador de cadena, el número de residuo y coordenadas espaciales x , y y z .

Actividad 2.3. Explicar brevemente el significado de cada una de las once secciones que aparecen en una línea típica de la sección **ATOM** dentro de un archivo *.pdb*.



Los programas de visualización estructural como UCSF Chimera, leen las coordenadas contenidas en archivos *.pdb* y las representan de diversas formas, permitiendo realizar análisis estructurales que van desde observaciones básicas hasta estudios muy especializados.

Para comenzar, se debe iniciar una sesión en UCSF Chimera version 1.18 y cargar la estructura correspondiente al archivo con PDB ID 6M2Q previamente descargado. Esto puede realizarse accediendo al menú **File>Fetch by ID.../PDB** y escribiendo *6m2q*; usando la barra de comandos (accesible desde **Favorites>CommandLine**) e ingresando el comando *open 6m2q*; o bien, cargando el archivo previamente descargado (**File>Open...**).

La estructura de la SCV2–Mpro aparecerá en la pantalla principal en una representación conocida como *cintas* o *ribbons*, una forma común de visualizar proteínas en biología estructural. Para facilitar la manipulación del modelo, se recomienda usar un ratón. Una guía útil para aprender a operar UCSF Chimera version 1.18 está disponible en:

<https://www.cgl.ucsf.edu/Outreach/Tutorials/GettingStarted.html>.

Para familiarizarse con el entorno de UCSF Chimera version 1.18, se recomienda realizar las siguientes pruebas básicas de manipulación y selección de la estructura:

- I. *Zoom*: girar la rueda del ratón hacia arriba o hacia abajo para acercar o alejar la vista
- II. *Traslación*: oprimir la rueda del ratón y moverlo en cualquier dirección para trasladar la estructura en el plano de la pantalla
- III. *Rotación*: hacer clic con el botón izquierdo del ratón y arrastrar para rotar la estructura respecto de su centro de masa
- IV. *Selección de la cadena principal (backbone)*: hacer clic derecho sobre cualquier parte de la estructura mientras se presiona la tecla **Control**. El segmento seleccionado se resaltará con un contorno verde (ver Figura 2.4). Para deseleccionar, se debe presionar **Control** + clic en cualquier área vacía de la pantalla

Es importante señalar que la forma biológicamente activa de la SCV2–Mpro, al igual que la de otros coronavirus como MERS y SARS-CoV-1, es un homodímero, aun cuando cada monómero contiene un sitio catalítico. En este contexto, un homodímero funcional se define como un complejo constituido por dos subunidades idénticas cuya asociación es necesaria para mantener la arquitectura del sitio activo, la estabilidad estructural y, en consecuencia, la actividad biológica de la proteína.

Esta situación se debe a que, en los estudios de XRC, la información depositada en el PDB suele corresponder a la unidad asimétrica del cristal, la cual no necesariamente coincide con la forma biológica funcional de la proteína. La unidad asimétrica representa la fracción mínima del cristal a partir de la cual puede reconstruirse toda la red cristalina mediante la aplicación de operaciones de simetría. En muchos casos,

dicha unidad contiene únicamente un monómero, por lo que el dímero funcional completo debe reconstruirse aplicando las operaciones de simetría cristalina apropiadas.



Figura 2.4. Representación en cintas de la estructura tridimensional de la SCV2-Mpro (PDB ID 6M2Q), generada con UCSF Chimera version 1.18. Resaltado mediante un contorno verde, se muestra el *backbone* del residuo Ser123.

Actividad 2.4. Realizar las siguientes acciones y describa sus observaciones:

Resaltar un punto a lo largo de una de las hebras beta de la estructura y presionar una sola vez la tecla flecha hacia arriba. ¿Qué sucede con la molécula al aplicar este comando?

Volver a presionar flecha hacia arriba. ¿Qué ocurre ahora con la molécula?

Repetir el procedimiento anterior, pero ahora utilizando la flecha hacia abajo. Describa lo que sucede en cada caso.

Dirigirse a **Select>Residue** y hacer clic en Cys. ¿Qué ocurre tras realizar la acción? Luego ir a **Actions>Atoms/Bonds>show**. Cambiar el color con **Actions>Color>yellow** y, finalmente, deseleccionar haciendo **Control** + clic en algún lugar del fondo de la ventana gráfica.



Como se mostrará más adelante, varios archivos *.pdb* de la SCV2-Mpro incluyen la forma homodimérica completa (por ejemplo, 5C3N y 6LU7); en cambio, 6M2Q contiene únicamente el monómero. En la **Práctica 9. Diseño de inhibidores para la Mpro del SARS-CoV-2 utilizando un enfoque basado en fragmentos** se explica cómo visualizar el ensamblaje biológicamente activo de la SCV2-Mpro cuando la información está en el archivo, pero UCSF Chimera no la muestra de forma predeterminada.

A continuación, se explorará un tipo de interacción electrostática común en proteínas: los *puentes salinos*. El objetivo es verificar si este tipo de interacciones está presente en la estructura de la SCV2-Mpro descargada. Para facilitar su visualización, primero se debe colorear la representación en cintas de color gris. Los pasos para lograrlo en UCSF Chimera version 1.18 son los siguientes:

- I. Seleccionar toda la estructura: **Select>SelectAll**
- II. Ocultar los átomos y enlaces: **Actions>Atoms/Bonds>hide**
- III. Colorear la estructura en gris: **Actions>Color>Gray**
- IV. Deseleccionar la estructura: hacer clic en **Select>Clear Selection**, o presionar **Control** + clic en un área vacía

El siguiente paso consiste en identificar posibles interacciones tipo *punto salino*. Estas interacciones electrostáticas ocurren entre residuos positivamente cargados, como lisina (Lys) y arginina (Arg), y residuos negativamente cargados como ácido aspártico (Asp) y ácido glutámico (Glu), siempre que se encuentren a una distancia aproximada no mayor a 4 Å.

Para facilitar su identificación visual, se deben colorear en rojo los residuos Asp y Glu, y en azul los residuos Lys y Arg. A continuación, se buscarán pares de residuos cuyos grupos amino y carboxilo estén a una distancia ≤ 3.5 Å, valor aproximado aceptado para considerar un *punto salino*. Los pasos a seguir en UCSF Chimera version 1.18 son los siguientes:

- I. Seleccionar los residuos Asp: **Select>Residue>ASP**
- II. Mostrar los átomos y enlaces: **Actions>Atoms/Bonds>Show**
- III. Colorear en rojo: **Actions>Color>Red**
- IV. Seleccionar los residuos Glu: **Select>Residue>GLU**
- V. Mostrar los átomos y enlaces: **Actions>Atoms/Bonds>Show**
- VI. Colorear en rojo: **Actions>Color>Red**
- VII. Seleccionar los residuos Lys: **Select>Residue>LYS**
- VIII. Mostrar los átomos y enlaces: **Actions>Atoms/Bonds>Show**
- IX. Colorear en azul: **Actions>Color>Blue**
- X. Seleccionar los residuos Arg: **Select>Residue>ARG**
- XI. Mostrar los átomos y enlaces: **Actions>Atoms/Bonds>Show**
- XII. Colorear en azul: **Actions>Color>Blue**

XIII. Deseleccionar la estructura: **Select>Clear Selection**

Actividad 2.5. Se deben consultar los valores de pK_a de los grupos ionizables presentes en las cadenas laterales de los residuos mencionados y discutir cuál sería su carga probable a pH 7.3, considerando además el contexto de su entorno dentro de la estructura de la SCV2-Mpro. Finalmente, responder: ¿por qué no se considera al residuo His para identificar *puentes salinos*?

Una vez coloreados los residuos, se recomienda girar la estructura en todas direcciones utilizando el ratón para facilitar la identificación visual de posibles *puentes salinos*, observando la proximidad espacial de los grupos carboxilato (Asp/Glu) y los grupos amino (Lys/Arg). Cuando se sospeche de una posible interacción, se deben seleccionar dos átomos cercanos, por ejemplo, el átomo de oxígeno del grupo carboxilato y el átomo de nitrógeno del grupo amino más próximos. Para ello, hacer **Control** + clic en el primer átomo y luego **Mayús** + **Control** + clic en el segundo átomo. A continuación, ir a **Tools>Structure Analysis>Distances** y, en el recuadro que aparece, hacer clic en **Create** para medir la distancia entre ambos átomos. Debería obtenerse una imagen similar a la mostrada en la Figura 2.5, donde se visualiza la distancia en Å entre los átomos seleccionados. Una distancia menor o igual a 3.5 Å sugiere la posible existencia de un *punto salino*.

Actividad 2.6. Describir las interacciones *punto salino* identificadas, indicando los residuos involucrados, así como el grupo funcional participante y la distancia entre los átomos con cargas opuestas (oxígeno del grupo carboxilato y nitrógeno del grupo amino).

Hay que recordar que anteriormente se compararon las secuencias de la Mpro de SARS-CoV-1, SARS-CoV-2 y MERS mediante un alineamiento múltiple, cuyo propósito fue identificar regiones conservadas y divergentes entre estas proteínas, aunque también es posible anticipar similitudes estructurales y posibles diferencias funcionales. Dado que en el PDB se dispone de la estructura de la Mpro de los tres virus mencionados, es posible realizar ahora una comparación estrictamente tridimensional de sus respectivas estructuras. Para ello, se debe cerrar la sesión actual UCSF Chimera seleccionando **File>Close Session**. A continuación, se debe abrir nuevamente la estructura 6M2Q mediante cualquiera de las siguientes opciones:

- I. Ir a **File>Open>Fetch by ID...**, escribir 6M2Q y hacer clic en **Fetch**
- II. Hacer clic directamente sobre el recuadro **6m2q** (si aún aparece) en la esquina superior derecha de la ventana principal
- III. Utilizar la línea de comandos activando **Favorites>Command Line** e ingresando *open 6M2Q*

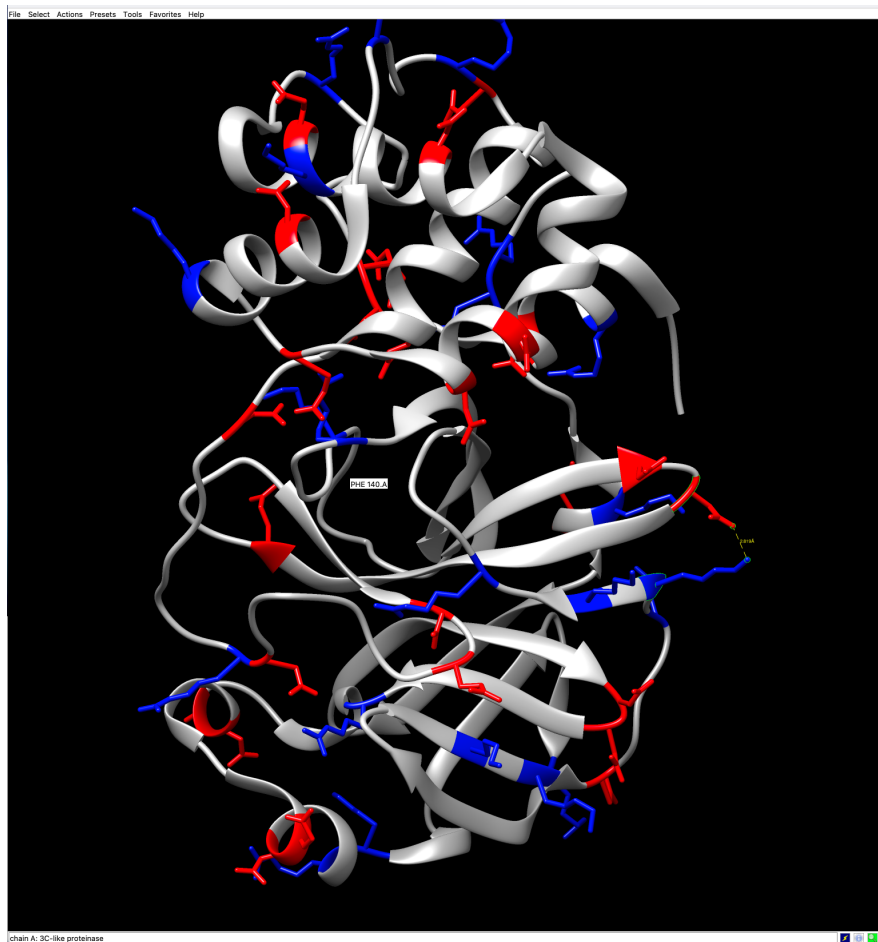


Figura 2.5. Representación en cintas de la estructura de la SCV2-Mpro (PDB ID 6M2Q), en la que se resaltan en rojo los residuos ácidos y en azul los residuos básicos. Se indica, mediante una línea punteada amarilla la distancia (2.99 Å) entre el oxígeno del grupo carboxilato y el nitrógeno del grupo amino en Asp289 y Arg131, respectivamente, implicados en la interacción *punto salino*. Imagen generada con UCSF Chimera version 1.18.

Una vez cargada esta estructura, se debe abrir la estructura de la Mpro del MERS, con PDB ID es 5C3N. Para ello, seleccionar **File>Fetch by ID...**, escribir **5C3N** y hacer clic en **Fetch**. Si esta última estructura no se visualiza, se puede intentar alejar la vista girando la rueda del ratón hacia abajo, para ajustar el campo visual y asegurarse de que ambas estructuras sean visibles.

Al observar ambas estructuras, se notará que la proteasa del MERS (PDB ID 5C3N) se presenta como un homodímero, mientras que la SCV2-Mpro (PDB ID 6M2Q) aparece como monómero. Es importante destacar que muchas proteínas funcionan como oligómeros, es decir, ensamblajes de múltiples subunidades. Sin embargo, en ocasiones estos ensamblajes no representan la forma biológica funcional, sino que resultan del proceso de cristalización.



Para comparar ambas estructuras tridimensionalmente, se utilizará la herramienta **MatchMaker**, siguiendo los pasos descritos a continuación:

- I. Abrir la herramienta **Matchmaker**: ir al menú **Tools>Structure Comparison>MatchMaker**
- II. En la ventana emergente, elegir **6M2Q** como **Reference structure** (estructura fija) y **5C3N** como **Structure to match** (estructura móvil). Para ello, hacer clic izquierdo sobre cada nombre en el menú desplegable correspondiente
- III. Asegurarse de que todos los parámetros estén en sus valores predeterminados (por defecto)
- IV. Hacer clic en **OK** para iniciar el alineamiento estructural (ver Figura 2.6)
- V. Observar los resultados: la barra de comandos mostrará brevemente datos del ajuste, por ejemplo, número de residuos alineados, y el RMSD (en la siguiente sugerencia de actividades, se deberá investigar el significado y utilidad del RMSD)
- VI. Consultar el reporte completo del alineamiento: ir a **Favorites>Reply Log**, donde se mostrará información detallada del proceso, incluyendo métricas del ajuste estructural

Actividad 2.7. Describir lo que sucedió con las estructuras durante el alineamiento estructural. Identificar cuál subunidad del dímero de la Mpro del MERS quedó alineada con el monómero de la Mpro del SARS-CoV-2.

Además, existe una medida cuantitativa que indica qué tan bien se superponen dos objetos tridimensionales. Esta medida se denomina raíz de la desviación cuadrática media (RMSD, Root Mean Square Deviation). Se debe describir qué es el RMSD e interpretar los resultados mostrados en **Reply Log**.

Una forma eficiente de guardar una sesión de trabajo en UCSF Chimera, incluyendo el estado actual de la visualización y un registro completo de las acciones realizadas, es utilizar la opción **File>Save Session As...**

Al abrir una nueva sesión, el nombre del archivo de sesión guardado aparecerá en la parte superior derecha de la ventana. Para restaurar el entorno de trabajo, basta con hacer doble clic en dicho nombre, y toda la información se cargará exactamente como estaba antes de cerrar la sesión.

El menú **File** también incluye otras opciones útiles para guardar información. Dos de las más empleadas son:

- **Save PDB...**: permite guardar la estructura en formato PDB, incluyendo las modificaciones realizadas
- **Save Image...**: permite exportar la vista actual como imagen, con opciones para ajustar el formato (PNG, JPEG, TIFF, entre otros) y la resolución deseada



Figura 2.6. Alineamiento estructural entre las proteasas principales 5C3N (MERS-CoV) y 6M2Q (SARS-CoV-2). El homodímero de 5C3N se representa en color cian, mientras que el monómero de 6M2Q se muestra en color café. Imagen generada con UCSF Chimera version 1.18.

Actividad 2.8. Crear una figura que muestre las estructuras alineadas con las siguientes características:

- I. Fondo blanco
- II. Mostrar solo el monómero de 5C3N alineado, representado como cintas rojas
- III. Mostrar el monómero de 6M2Q en cintas de color gris

Sugerencia técnica:

- **Presets>Publication 3** proporcionará el fondo blanco
- Una vez seleccionado el monómero de MERS que no fue alineado, ocultarlo con **Actions>Ribbons>hide**



Bibliografia

- Brandt G. S.** & W. R. P. Novak (2021) SARS-CoV-2 virtual biochemistry labs on bioinformatics and drug design. *Biochem Mol Biol Educ.* 49:26-28.
- 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.
- Hu, Q.**, Xiong, Y., Zhu, G. H., Zhang, Y. N., Zhang, Y. W., Huang, P., & Ge, G. B. (2022). *The SARS-CoV-2 main protease (Mpro): Structure, function, and emerging therapies for COVID-19*. *MedComm*, 3(3): e151.
- Humphrey, W.**, Dalke, A., & Schulten, K. (1996). *VMD: visual molecular dynamics*. *Journal of molecular graphics*, 14(1): 33–28.
- Martinez, X.**, Krone, M., Alharbi, N., Rose, A. S., Laramée, R. S., O'Donoghue, S., Baaden, M., & Chavent, M. (2019). *Molecular Graphics: Bridging Structural Biologists and Computer Scientists*. *Structure* (London, England: 1993), 27(11): 1617–1623.
- Pettersen, E. F.**, Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). *UCSF Chimera a visualization system for exploratory research and analysis*. *Journal of computational chemistry*, 25(13): 1605–1612.
- Seeliger, D.**, & de Groot, B. L. (2010). *Ligand docking and binding site analysis with PyMOL and Autodock/Vina*. *Journal of computer-aided molecular design*, 24(5): 417–422.



Práctica 3. Examinando interacciones covalentes y no covalentes en un complejo proteína–ligando relacionado con el COVID-19

Objetivos

- Identificar y clasificar interacciones covalentes y no covalentes en un complejo proteína–ligando.
- Utilizar UCSF Chimera para visualizar distancias donador–aceptor en enlaces de hidrógeno y zonas ≤ 3.5 Å alrededor del ligando.
- Consultar bases de datos especializadas para recuperar propiedades fisicoquímicas del ligando y evaluar su consistencia.
- Integrar evidencia estructural y química para proponer el modo de unión del inhibidor y sugerencias para su optimización.

Software especializado requerido

- UCSF Chimera versión 1.18 o superior

El sitio activo de una enzima generalmente se localiza sobre su superficie o en una región cercana a ella, lo que facilita el acceso del sustrato. Esta zona puede adoptar la forma de un surco, hendidura o cavidad, que facilita el acoplamiento adecuado del sustrato a la enzima. En muchos casos, el sitio activo presenta una naturaleza más hidrofóbica que el resto de la superficie externa de la enzima, creando un entorno favorable para reacciones que serían difíciles o imposibles de llevar a cabo en medio puramente acuoso.

Los residuos de aminoácido que conforman el sitio activo suelen desempeñar funciones clave en la actividad enzimática. Para identificarlos, es común comparar la estructura primaria –la secuencia de aminoácidos– de la enzima en diferentes organismos mediante un alineamiento múltiple de secuencias. Algunos aminoácidos permanecen invariables independientemente del origen de la enzima, y se les conoce como *aminoácidos conservados*, y son cruciales para la función.

Estos aminoácidos conservados suelen participar en varias funciones importantes, como:

- I. Aminoácidos de unión: participan en la fijación el sustrato o algún cofactor al sitio activo
- II. Aminoácidos catalíticos: participan directamente en el mecanismo de reacción
- III. Aminoácidos estructurales o de estabilidad: contribuyen al mantenimiento de la estructura del sitio activo e incluso de la estructura general de la enzima

Afortunadamente, la SCV2–Mpro es una enzima estructuralmente sencilla, con un sitio activo fácilmente identificable. A continuación, se analizarán dos representaciones de su sitio activo. Para comenzar, se debe iniciar una nueva sesión en UCSF Chimera y cargar la estructura cristalina de la proteasa principal unida covalentemente a un inhibidor. El PDB ID de esta estructura es 2ALV.



Al cargar la estructura, UCSF Chimera desplegará automáticamente la proteasa en una representación tipo *cintas (ribbons)*. Por su parte, el inhibidor, junto con las cadenas laterales de los residuos de aminoácidos en contacto directo con él, se visualizarán como varillas, facilitando así la identificación del sitio activo y sus interacciones clave.

Actividad 3.1. Consulta de información en el PDB. Ir al sitio del PDB y recuperar el nombre de la estructura correspondiente al ID cargado y cita del artículo de investigación asociado.

¿Se trata de la estructura de la Mpro del SARS-CoV-1 o de la Mpro del SARS-CoV-2?

¿Cuál es el nombre corto del ligando y cuál es su nombre completo?

¿En qué año fue depositada en el PDB la primera estructura de una Mpro?

En la página del PDB, localizar la representación de la estructura del ligando, realizar una captura de pantalla y pegarla en la bitácora.

Es posible notar que la estructura 2ALV corresponde a la proteasa principal SARS-CoV-1 y no del SARS-CoV-2. Sin embargo, ambas comparten una alta similitud estructural.

Para explorar el sitio de unión del inhibidor, se debe girar la estructura del complejo con el ratón y ajustar el nivel de zoom según sea necesario, hasta obtener una visualización clara del complejo enzima–ligando.

Es importante identificar al ligando y el enlace covalente que forma con algún residuo en el sitio activo.

A continuación, se debe seleccionar al inhibidor y colorearlo de dorado mediante los siguientes pasos:

- I. **Select>Residue>CY6**
- II. **Actions>Color>all options...**
- III. En la ventana emergente, seleccionar **Show all colors>gold**

Si el inhibidor aún está seleccionado, utilizar **Actions>Focus** para mejorar la visualización del sitio de unión y facilitar sus análisis. Es importante mantener el inhibidor seleccionado para continuar con otras acciones o anotaciones.

Durante la epidemia del SARS-CoV-1 (2002-2004), se emprendieron diversos esfuerzos para desarrollar inhibidores dirigidos a esta proteasa. Varios de estos compuestos han servido como punto de partida para diseñar nuevos inhibidores contra la proteasa del SARS-CoV-2, dada su similitud estructural y funcional.

Para comprender mejor las interacciones moleculares que estabilizan al inhibidor en el sitio activo, es útil identificar los átomos específicos involucrados. Por ello, se recomienda utilizar la herramienta *color by heteroatom*, accediendo a **Actions>Color>by heteroatom**. Esta función asigna colores estándar a los átomos según su tipo químico



(por ejemplo, oxígeno en rojo, nitrógeno en azul), facilitando la interpretación visual de las interacciones no covalentes.

Actividad 3.2. Realizar una búsqueda en el PDB de otros tres inhibidores de la Mpro del SARS-CoV-1 y completar la información siguiente:

- I. PDB ID, fecha de publicación y nombre de la estructura
- II. Cita del artículo asociado a la estructura
- III. El nombre del inhibidor (corto y IUPAC)
- IV. Adicionalmente, para Cy6
- V. Se deben enlistar todos los grupos funcionales presentes
- VI. Se deben identificar cuántos centros quirales hay
- VII. Se debe determinar el peso molecular
- VIII. Registrar el número de enlaces rotables y grupos donadores y aceptores de enlace de hidrógeno

Ahora, se deben examinar las interacciones específicas que permiten que esta molécula se una al sitio activo. Asegurarse de que CY6 esté aún seleccionada. Si no es así, se debe volver a seleccionar mediante **Select>Residue>CY6**. Luego, ir a **Tools>Structure Analysis>FindHBond**. Esto abrirá la ventana **H-Bond Parameters** (ver Figura 3.1).

Algunos parámetros específicos de las interacciones intermoleculares, como los enlaces de hidrógeno, pueden personalizarse en esta ventana. Sin embargo, se deben mantener los valores predeterminados y verificar que estén activadas las siguientes casillas: **only find H bonds... with at least one end selected, if endpoint atom is hidden, show endpoint residue** y **write information to reply log**. Una vez configurados los parámetros, hacer clic en **Ok**.

Como resultado, se mostrarán líneas punteadas azules que representan los enlaces de hidrógeno entre el inhibidor y los residuos que lo rodean del sitio activo. Esta visualización permite identificar con precisión qué átomos y residuos participan en las interacciones clave que estabilizan la unión del inhibidor a la enzima.

Actividad 3.3. Análisis de enlaces de hidrógeno. Con base en el análisis realizado en UCSF Chimera, responder lo siguiente:

¿Cuántos enlaces de hidrógeno fueron identificados?

Describir cada uno de los enlaces de hidrógeno. ¿Quién es átomo donador (*D*) y quién el átomo aceptor (*A*)? ¿Cuál es la distancia entre *D* y *A*?

¿En alguno de los enlaces de participa la cadena lateral del residuo? Si es así, describirlo.

El panel **Reply Log** resulta útil para el análisis detallado de las interacciones: **Favorites>Reply Log**.

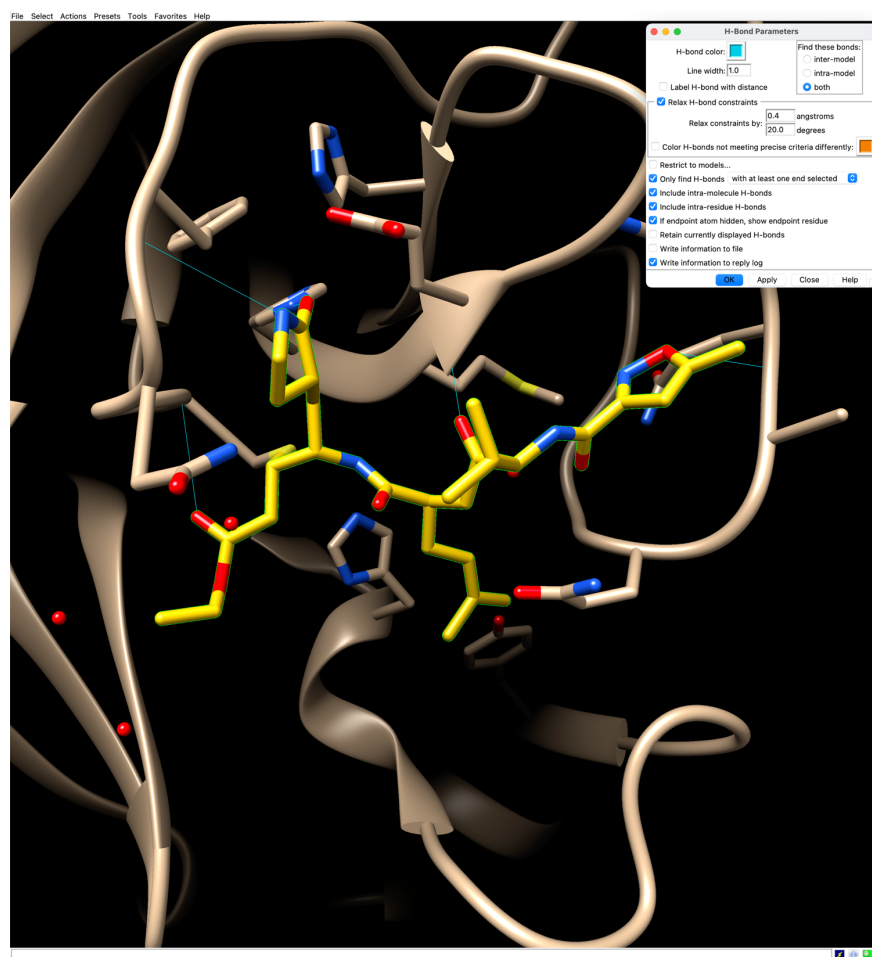


Figura 3.1. Representación del complejo proteína–ligando correspondiente a la Mpro del SARS-CoV-1 unida covalentemente a un inhibidor (PDB ID 2ALV). La enzima se muestra en formato de cintas, mientras que el inhibidor (en amarillo) y las cadenas laterales de los residuos de aminoácidos que interactúan directamente con él se representan en forma de varillas. Imagen generada con UCSF Chimera version 1.18.

El inhibidor CY6 se encuentra unido covalentemente a la Cys145, una interacción que puede observarse directamente en la representación estructural del complejo. En caso de que este enlace aún no sea evidente, se recomienda centrar el análisis en localizar visualmente dicha unión, prestando especial atención a la cercanía y continuidad entre los átomos del ligando y la cadena lateral de la cisteína.

Actividad 3.4. Describir el enlace covalente entre CY6 y Cys145:

¿Qué átomos específicos están participando en el enlace covalente?

¿Cuál es la distancia del enlace? En la **Práctica 2. Visualización de proteínas** se describió cómo medir distancias átomo–átomo.



Adicionalmente, investigar y proponer un mecanismo de reacción para explicar la formación del enlace covalente mediante la consulta en fuentes especializadas en química orgánica, en particular en la sección de reacciones de adición nucleofílica.

Hasta este punto del análisis se ha identificado la presencia de un enlace covalente entre el inhibidor CY6 y el residuo Cys145, así como varios enlaces de hidrógeno entre el ligando y residuos cercanos del sitio activo. La siguiente etapa consiste en determinar si existen interacciones adicionales, tales como puentes salinos, lo cual requiere un análisis estructural más detallado del complejo proteína–ligando.

Actividad 3.5. Incluir en la bitácora una imagen del complejo Mpro-CY6 que muestre de forma clara y visual las interacciones intermoleculares entre la proteasa principal y el ligando.

Identificar los grupos funcionales presentes en la molécula de CY6.

Con base en conocimientos de química general, relacionar estos grupos funcionales con la naturaleza química de cada interacción observada.

De acuerdo con los grupos químicos presentes en CY6, clasificar la molécula como polar o no polar, mencionando al menos tres grupos funcionales que sustenten dicha clasificación.

Determinar si CY6 posee cargas formales. En caso afirmativo, indicar su valor de la carga formal total del ligando, los átomos que portan dichas cargas y los residuos de aminoácidos involucrados en posibles interacciones electrostáticas.

Para obtener información adicional sobre el inhibidor, se sugiere consultar bases de datos especializadas como *ZINC database* (<https://zinc15.docking.org/>) y *ChemSpider* (<http://www.chemspider.com/>), ampliamente utilizadas por químicos, bioquímicos, médicos, biólogos, entre otros, para la obtención de propiedades fisicoquímicas de compuestos, particularmente en el contexto del Diseño Racional de Fármacos.

Al igual que en *GenBank*, *UniProt* y el PDB, estas bases permiten realizar búsquedas mediante distintos criterios. Se debe realizar la búsqueda del inhibidor covalente CY6 en ambas bases de datos empleando la palabra clave CY6.

En el caso de *ZINC database*, después de ingresar al sitio, se debe hacer clic en **Substances**. En esta sección se puede dibujar la estructura bidimensional del compuesto de interés o bien ingresar directamente el ZINC ID correspondiente: ZINC28526050, en el campo de búsqueda ubicado en la esquina superior derecha. Al hacer clic en **Search**, se mostrará un único coincidente con el ID ingresado. Hacer clic en dicho resultado y explorar la información disponible, tal y como se muestra en la Figura 3.2.

ZINC Substances Catalogs Tranches Biological + More + About +

ZINC28526050

Google Wikipedia PubMed

Added	Availability	Since	Mwt	logP	Download
2009-03-01	Annotated	2015-08-07	558.676	2.799	Download

Mol Formula	Rings	Heavy Atoms	Hetero Atoms	Fraction sp ²	Tranche
C28H42N4O7	2	40	11	0.59	KFAF

SMILES: CCOC(=O)C=C(C)[C@H](C)[C@H](C)CNC(=O)N(C)[C@H](C)C(=O)C(C)(C)C(=O)C[C@H](N)C(=O)C(C)C(=O)N1C(C)C

InChI: InChI=1S/C28H42N4O7/c1-7-39-25(35)11-10-22(15-21-12-13-30-27(136)31-28(37)30(9-8-17(2)316-24(34)20(18)4(5)32-29(38)23-14-19(6)40-33-23(4)8,10-11,14,18,26,27,36,37,38,39,40,41,42)/p-1/s

InChI Key: CSNQHKUCKPMZCY-YFLAUPXSA-N

Available 3D Representations

pH range	Net charge	H-bond donors	H-bond acceptors	TPSA	Rotatable bonds	Axial desolvation	Polar desolvation	Download
Reference	0	3	8	156	15	9.07	-18.6	Download

Vendors (0 Total)

This compound is not currently for sale. If you believe this to be in error, please write us at chemistry-biology@atg.net

Annotated Catalogs (10 Total)

CHEMBL20	CHEMBL196035
CHEMBL21	CHEMBL196035
CHEMBL23	CHEMBL196035
CHEMBL25	CHEMBL196035
CHEMBL27	CHEMBL196035
CHEMBL28	CHEMBL196035
CHEMBL29	CHEMBL196035
CHEMBL30	CHEMBL196035
CHEMBL33	CHEMBL196035
Protein Databank	CYS

10 Items Total

Figura 3.2. Captura de pantalla de la interfaz de la base de datos ZINC que muestra el resultado de la búsqueda del compuesto ZINC28526050, correspondiente al inhibidor CY6. Se presentan algunas de sus propiedades fisicoquímicas relevantes, tales como peso molecular, fórmula molecular, número de enlaces rotables, logP estimado, entre otros parámetros útiles en el análisis de fármacos.

Actividad 3.6. Describir al menos cinco propiedades fisicoquímicas de interés del ligando (por ejemplo, peso molecular, logP, número de enlaces rotables, número de donadores y aceptores de enlace de hidrógeno, área de superficie polar, etc.). Comparar la información entre ambas bases de datos para verificar si existe consistencia en, al menos, una propiedad seleccionada. Si no es así, tratar de encontrar la posible razón. Verificar si el peso molecular, el número de enlaces rotables y el número de grupos donadores/aceptores de enlace de hidrógeno descritos al inicio de esta práctica coinciden con los valores reportados por las dos bases de datos consultadas?

Además de los enlaces covalentes e interacciones electrostáticas, existen otras interacciones intermoleculares clave, como las fuerzas de dispersión o fuerzas *de van der Waals*, que también contribuyen al reconocimiento molecular y al acoplamiento del ligando al sitio activo. Para identificarlas visualmente, se recomienda realizar los siguientes pasos:

- I. Asegurarse de que solo el residuo **CY6** esté seleccionado
- II. Ir al menú superior y seleccionar **Select>Zone...**
- III. En la ventana emergente, marcar la casilla **angstroms from currently selected atoms**
- IV. Especificar el valor de distancia de **3.5 Å**
- V. Activar la opción **select all atoms/bonds of any residue within the selection zone**
- VI. Hacer clic en **OK** para confirmar y aplicar la selección



Una vez aplicados estos pasos, en la ventana principal se observará que varios átomos cercanos al inhibidor han sido seleccionados. Para visualizarlos, se deben colorear en azul mediante **Actions>Color>cornflower blue**. Si accidentalmente el inhibidor CY6 cambia de color y pierde su color dorado, se debe restaurar mediante **Actions>Color>gold**.

Es importante que el inhibidor CY6 permanezca claramente resaltado para el siguiente análisis, en el que se visualizará la isosuperficie de la Mpro. Este procedimiento es equivalente a delinear la silueta tridimensional de un objeto a partir de su radiografía, lo cual permite observar el grado de complementariedad geométrica entre la proteína y el ligando, aspecto clave en el reconocimiento molecular.

Para llevar a cabo este análisis, seleccionar únicamente la proteína, asegurándose de que el inhibidor CY6 no esté incluido en la selección. Si lo está, usar **Select>Invert (all models)**. A continuación, aplicar **Actions>Surface>show**.

La herramienta **Surface** de UCSF Chimera generará una representación superficial basada en los radios de *van der Waals*, aplicando automáticamente la coloración del residuo subyacente. Como resultado, cualquier residuo de aminoácido que tenga al menos un átomo a una distancia ≤ 3.5 Å del inhibidor se mostrará en color azul (según la configuración previa), mientras que el resto de la proteína conservará su color original.

La Figura 3.3A muestra esta configuración híbrida: la isosuperficie molecular de la Mpro se visualiza de forma continua, mientras que el inhibidor CY6 se representa en formato de varillas doradas, destacando su posición y sus posibles puntos de contacto.

Se recomienda girar la estructura tridimensional utilizando el ratón con el objetivo de identificar cavidades o hendiduras que puedan servir como sitios de unión para pequeñas moléculas. Durante esta inspección visual, se observará que el sitio activo de la Mpro es relativamente poco profundo, lo cual resulta coherente desde el punto de vista funcional: esta enzima procesa una poliproteína viral, por lo que necesita un sitio activo accesible que facilite la unión y el reconocimiento de un sustrato de gran tamaño.

Es importante resaltar que los sitios activos y sitios de unión de ligandos en proteínas suelen adoptar la forma de cavidades (espacios delimitados en la superficie), no de concavidades cerradas, ya que deben permitir la entrada, el acomodo y la interacción del ligando o sustrato.

Como último paso del análisis, se debe seleccionar nuevamente al inhibidor CY6 y cambiar su representación a esferas mediante **Actions>Atoms/Bonds>Sphere**. De manera opcional, renderizar su superficie molecular aplicando **Actions>Surface>show**. La Figura 3.3B ilustra esta configuración final, en la que el inhibidor CY6 se muestra en forma de esferas doradas, permitiendo apreciar mejor la complementariedad estérica con la superficie de la proteasa.

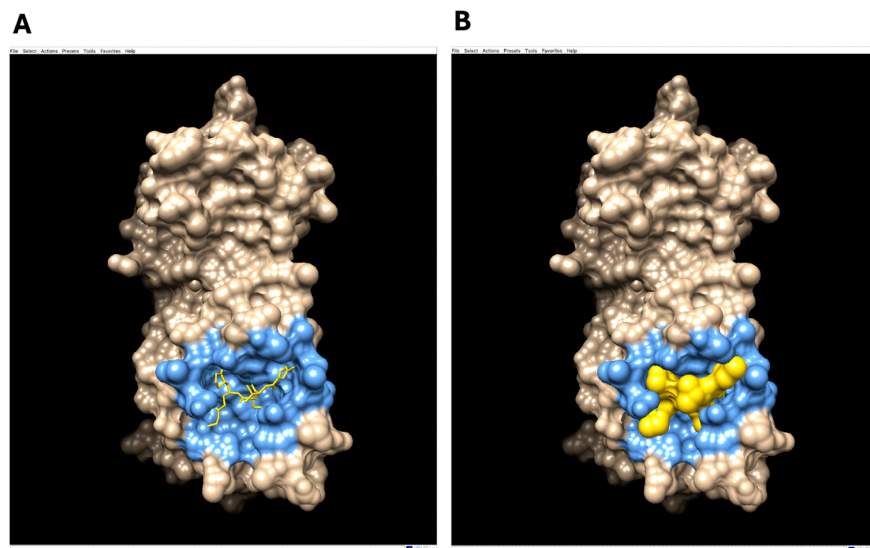


Figura 3.3. Representación tridimensional de la isosuperficie de la Mpro del SARS-CoV-1 basada en los radios de *van der Waals* de sus átomos, generada con UCSF Chimera versión 1.18. (A) El inhibidor CY6 se muestra en formato de varillas doradas, mientras que los residuos de aminoácido cuyos átomos se encuentran a ≤ 3.5 Å del ligando se visualizan mediante isosuperficie de color azul. (B) El mismo complejo se presenta con CY6 en formato de esferas, resaltando la complementariedad estérica entre el ligando y el sitio activo. Esta visualización permite identificar cavidades superficiales compatibles con la unión del inhibidor.

Actividad 3.7. Para inactivar la enzima eficazmente se requiere que los inhibidores se unan de manera estrecha y específica a su sitio. Con base a la estructura del ligando y el análisis de las interacciones observadas, responder lo siguiente:

¿Qué grupos funcionales podrían añadirse al ligando para mejorar su afinidad por el sitio activo?

¿Qué interacciones intermoleculares podrían explotarse o reforzarse mediante modificaciones químicas?

¿Qué partes del ligando podrían eliminarse sin afectar de manera significativa su unión a la enzima?

Finalmente, para facilitar el análisis visual del ligando y de su acomodo en el sitio activo, es posible alternar entre las representaciones de esqueleto y superficie del ligando y la proteína utilizando: **Model Panel: Favorites>ModelPanel.**

Bibliografía

Anighoro A. (2020). *Underappreciated Chemical Interactions in Protein-Ligand Complexes*. Methods in molecular biology (Clifton, N.J.) 2114, 75–86.

Brandt G. S. & W. R. P. Novak. (2021). SARS-CoV-2 virtual biochemistry labs on bioinformatics and drug design. Biochemistry and Molecular Biology Education, 49:26-28.



Irwin, J. J., & Shoichet, B. K. (2005). *ZINC--a free database of commercially available compounds for virtual screening*. Journal of chemical information and modeling 45(1): 177–182.

Irwin, J. J., Tang, K. G., Young, J., Dandarchuluun, C., Wong, B. R., Khurelbaatar, M., Moroz, Y. S., Mayfield, J., & Sayle, R. A. (2020). *ZINC20-A Free Ultralarge-Scale Chemical Database for Ligand Discovery*. Journal of chemical information and modeling 60(12): 6065–6073.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). *UCSF Chimera--a visualization system for exploratory research and analysis*. Journal of computational chemistry, 25(13): 1605–1612.



Práctica 4. Modelado de la estructura tridimensional de la SCV2-Mpro dimérica

Objetivos

- Generar un modelo tridimensional de la SCV2-Mpro mediante *AlphaFold2*, considerando su forma de homodímero funcional.
- Seleccionar una plantilla estructural adecuada a partir del análisis de homólogos mediante BLAST y *Clustal Omega*.
- Evaluar la calidad estereoquímica del modelo utilizando herramientas complementarias de validación estructural.
- Comparar el modelo con estructuras experimentales mediante alineamiento estructural y análisis de RMSD, discutiendo su confiabilidad en el contexto del diseño de fármacos.

Software especializado requerido

- UCSF Chimera versión 1.18 o superior

Tal como se ha señalado en prácticas previas, las proteínas son componentes esenciales de los seres vivos, y comprender su estructura tridimensional es fundamental para dilucidar su función a nivel molecular. A lo largo de los años, gracias a un enorme esfuerzo científico y tecnológico, se ha logrado determinar experimentalmente la estructura tridimensional de más de 210,000 proteínas (<https://www.rcsb.org/>). Sin embargo, esta cifra representa apenas una pequeña fracción del vasto universo de secuencias de proteínas conocidas o que se han predicho para diversos organismos.

Factores como la disponibilidad de cantidades suficientes de proteína pura, así como su solubilidad y estabilidad, son determinantes clave para poder resolver sus estructuras. Las principales técnicas utilizadas para estudiar la estructura de proteínas son la XRC, la EM y la NMR.

En el campo del Diseño de Fármacos Basado en la Estructura (Structure-Based Drug Design, SBDD), disponer de la estructura tridimensional de la proteína blanco es fundamental para facilitar procesos fundamentales como:

- I. Cribado virtual (*virtual screening*): consiste en identificar moléculas potencialmente activas mediante la simulación del acoplamiento (*docking*) de cientos o miles de compuestos, presentes en una quimioteca virtual, sobre sitios específicos predefinidos en la superficie de la proteína
- II. *Diseño de ligandos de novo*: en este enfoque, los ligandos se construyen directamente dentro del sitio de unión de la proteína, ensamblando progresivamente átomos o grupos funcionales. Las conformaciones generadas de este modo suelen ser novedosas, es decir, no están presentes en bases de datos existentes ni han sido previamente sintetizadas

- 
- III. *Optimización de ligandos conocidos*: se diseñan y prueban análogos estructurales de ligandos ya conocidos, con el objetivo de mejorar propiedades como la afinidad, especificidad o perfil farmacocinético frente a la proteína blanco

En caso de no contar con una estructura tridimensional experimental de la proteína de interés, es posible obtener un modelo estructural por homología, utilizando como plantilla la estructura experimental de una proteína relacionada evolutivamente. El proceso incluye las siguientes etapas:

- I. *Búsqueda y alineamiento de secuencias homólogas*: identificar proteínas con alta similitud de secuencia cuya estructura tridimensional haya sido resuelta
- II. *Construcción del modelo estructural*: se construye la estructura de la proteína blanco a partir de regiones conservadas, utilizando como referencia la proteína plantilla
- III. *Modelado de las regiones variables o ausentes*: las regiones no conservadas o faltantes (como bucles o extremos flexibles) se reconstruyen utilizando algoritmos especializados, que predicen su conformación probable

El modelado por homología requiere procedimientos sistemáticos y reproducibles, no arbitrarios. Para ello, se han desarrollado diversas herramientas especializadas como SWISS-MODEL, MODELLER y *AlphaFold2*, entre otras. Cada una de estas plataformas implementa algoritmos propios, los cuales son continuamente actualizados para mejorar su precisión y confiabilidad. La selección del programa adecuado depende de múltiples factores, como el objetivo del estudio, el nivel de experiencia del usuario, la calidad de las estructuras disponibles y los recursos computacionales. En todos los casos, es responsabilidad del investigador elegir la herramienta más adecuada para las necesidades de su proyecto.

AlphaFold2, desarrollado por DeepMind, ha revolucionado el campo de la predicción estructural de proteínas al alcanzar una precisión comparable con técnicas experimentales como la XRC y la NMR. A diferencia del modelado por homología tradicional, que requiere una plantilla estructural identificada a partir de secuencias homólogas, *AlphaFold2* puede predecir la estructura tridimensional de una proteína únicamente a partir de su secuencia de aminoácidos, mediante el uso de modelos de aprendizaje profundo entrenados con bases de datos masivas de secuencias y estructuras.

Una de las capacidades más destacadas de *AlphaFold2* es su extensión denominada AlphaFold-Multimer, que permite modelar complejos oligoméricos, es decir, estructuras formadas por múltiples cadenas polipeptídicas que interactúan funcionalmente. Esta herramienta facilita la predicción de dímeros, trímeros o complejos más grandes sin requerir necesariamente una plantilla estructural preexistente, lo cual representa una ventaja significativa frente a otros enfoques tradicionales.



En esta práctica, se construirá un modelo tridimensional por homología de la SCV2–Mpro. Posteriormente, el modelo obtenido será comparado con una estructura cristalográfica disponible en el PDB, lo cual permitirá validar el modelo y evaluar su precisión. Este ejercicio tiene como objetivo mostrar el potencial del modelado por homología como herramienta en el Diseño Racional de Fármacos, especialmente en aquellos escenarios en los que la estructura experimental de la proteína blanco aún no ha sido resuelta.

4.1 Identificación de la plantilla

Se parte del supuesto de que la **Práctica 1. Bioinformática: Utilización de las bases de datos para el estudio de proteínas** ya fue completada, y que la secuencia de aminoácidos de la SCV2–Mpro se encuentra disponible en la bitácora electrónica de trabajo. Con base en dicha secuencia, el siguiente paso consiste en buscar secuencias homólogas que cuenten con estructuras tridimensionales disponibles en el PDB.

Para ello, se utilizará la herramienta BLAST (Basic Local Alignment Search Tool), accesible desde el portal del NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). Esta herramienta permite realizar alineamientos locales entre una secuencia de entrada (ya sea de nucleótidos o aminoácidos) y una vasta colección de secuencias contenidas en bases de datos como *UniProt*, *GenBank* o PDB, con el fin de identificar homólogos evolutivos y, en el caso del PDB, estructuras tridimensionales asociadas. Procedimiento:

- I. Acceder al portal del NCBI y seleccionar la opción **Protein BLAST** en la página principal
- II. En la sección **Enter accession number(s), or FASTA sequence(s)** pegar la secuencia de la SCV2–Mpro sin incluir el *encabezado*
- III. En el campo **Job Title**, asignar un nombre descriptivo a la búsqueda (opcional)
- IV. En la sección **Database**, seleccionar la opción **Protein Data Bank proteins(pdb)**, con el fin de restringir la búsqueda únicamente a secuencias que cuenten estructuras tridimensionales depositadas en el PDB

Es importante tener en cuenta que, actualmente existen múltiples estructuras tridimensionales disponibles del SARS-CoV-2, incluyendo la de su Mpro. No obstante, para fines didácticos, en esta práctica se asumirá que la estructura de la SCV2–Mpro no se ha resuelto experimentalmente. Bajo este supuesto, se procederá a construir un modelo tridimensional por homología, utilizando como plantilla la estructura cuya secuencia tenga un alto porcentaje de identidad con la SCV2–Mpro. El porcentaje de identidad se refiere a la proporción de residuos de aminoácidos idénticos entre dos secuencias alineadas. Para evitar seleccionar por error la estructura exacta de la SCV2–Mpro como plantilla (la cual tendría 100 % de identidad), deben aplicarse las siguientes condiciones de exclusión:

- I. En la sección **Organism**, agregar los organismos indicados en la Figura 4.1, asegurándose de utilizar el botón **Add organism** tras ingresar cada uno
- II. Confirmar que las casilla de los organismos añadidos estén seleccionadas, lo cual limitará la búsqueda a secuencias provenientes de dichos organismos
- III. Mantener todos los parámetros con sus valores predeterminados
- IV. Finalmente, hacer clic en botón **BLAST** para iniciar la búsqueda

The screenshot shows the NCBI BLAST interface for a Standard Protein BLAST search. The 'Enter Query Sequence' field is populated with a long amino acid sequence. Below this, the 'Choose Search Set' section is expanded, showing the 'Database' set to 'Protein Data Bank proteins (pdb)'. Under the 'Organism' section, three organisms are listed: 'SARS-CoV-2 (taxid:2697049)', 'COVID-19 (taxid:2697049)', and 'Severe acute respiratory syndrome coronavirus 2 (taxid:2697049)'. Each organism has a checked box next to it, indicating it is selected for the search. The 'Program Selection' section shows 'blastp (protein-protein BLAST)' as the selected algorithm. The 'BLAST' button is located at the bottom of the form.

Figura 4.1. Captura de pantalla de la interfaz del servidor BLAST (NCBI), específicamente de la opción **Protein BLAST**. Se muestra el recuadro donde debe ingresarse la secuencia de aminoácidos de la SCV2–Mpro (sin *encabezado*), así como el campo **Organism**, en el cual se pueden añadir organismos de interés usando la opción **Add organism**.

Es fácil identificar que las primeras secuencias listadas por BLAST –las que presentan los valores más altos del porcentaje de identidad (*Per. Ident*)– corresponden a la misma SCV2–Mpro, ya que tienen una identidad del 100 % o muy cercana a este valor (ver Figura 4.2). Para los fines de esta práctica, dichas secuencias deben descartarse, ya que se asume que no existe una estructura tridimensional disponible para la SCV2–Mpro. Para ello, se aplicará un segundo filtro para eliminarlas. Los porcentajes de identidad (matriz de identidad) entre las Mpro de coronavirus relacionados, revisados



en la **Práctica 1. Bioinformática: Utilización de las bases de datos para el estudio de proteínas**, serán de utilidad. En la sección **Filter Results**, aplicar los valores de **Percent Identity** y **Query Coverage** indicados en la Figura 4.2 y hacer clic en **Filter**.

Actividad 4.1. Indicar la cantidad de resultados obtenidos en la búsqueda en BLAST. Comentar a qué organismos pertenecen las proteínas para las cuales se identificaron estructuras con secuencia homóloga. Para ello, buscar sus estructuras en el PDB y revisar el organismo de origen. Adicionalmente:

Responder si todas las estructuras corresponden a la forma silvestre (wild-type) de la Mpro o si algunas son variantes mutantes.

Indicar además si se trata de estructuras libres de ligando (*apo*) o si están unidas a algún ligando (*holo*).

Registrar el valor de la resolución (**Resolution**) de estructura.

Evaluar si los criterios de exclusión aplicados en la búsqueda fueron efectivos.

En la sección **Sequences producing significant alignments**, explicar el significado de los siguientes términos: *Query Cover*, *E value*, *Per. Ident* y *Acc Len*.

De entre los resultados restantes, se seleccionará como plantilla o molde aquella estructura cuya secuencia posea el mayor porcentaje de identidad con respecto a la SCV2-Mpro y cuya estructura cristalográfica tenga alta resolución (≤ 2.0 Å). Para este fin, se deberá calcular el porcentaje de identidad global entre la secuencia en formato *fasta* de la SCV2-Mpro y cada una de las secuencias de los resultados restantes. A continuación, se describen los pasos:

- I. Seleccionar las 32 entradas restantes (la cantidad puede variar con el tiempo debido al depósito de nuevas estructuras relacionadas en el PDB) haciendo clic en la opción **Select all**
- II. Hacer clic en el menú **Download**, seleccionar la opción **FASTA (completesequences)**. El archivo, con el nombre *seqdump.txt*, se descargará automáticamente. Guardarlo dentro de la carpeta de trabajo previamente creada para esta práctica
- III. Abrir el archivo *seqdump.txt* en un editor de texto (como Notepad o WordPad) y copiar en la parte superior del documento la secuencia en formato *fasta* de la SCV2-Mpro
- IV. Guardar los cambios. Este archivo contendrá todas las secuencias necesarias para realizar un alineamiento múltiple, del cual se obtendrá el porcentaje de identidad global entre la SCV2-Mpro y las demás proteínas

Como se mencionó al inicio, **BLAST protein** realiza alineamientos locales, enfocándose en las regiones más similares entre la secuencia problema y las secuencias depositadas en las bases de datos. Si bien esto es útil para detectar

dominios conservados, puede introducir un sesgo al no considerar la totalidad de las secuencias comparadas.

The screenshot displays the BLAST search results interface from the National Library of Medicine. The search was performed for 'SARS-CoV-2 Mpro-like sequences' using the query 'JY27KWF3016'. The results are filtered to show sequences with a Percent Identity between 50 and 97, and Query Coverage between 99 and 100. The table lists 32 sequences, including various SARS-CoV-2 Mpro-like sequences and other related proteins. The table columns include: select all, Description, Scientific Name, Max Score, Total Score, Query Cover, E value, Per Ident, Acc Len, and Accession.

select all	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per Ident	Acc Len	Accession
☒	Chain A, 3C-like proteinase [Severe acute respiratory syndrome-related coronavirus]	Severe acute respiratory syndrom...	623	623	100%	0.0	96.08%	308	3M3V_A
☒	Chain A, 3C-like proteinase [Severe acute respiratory syndrome-related coronavirus]	Severe acute respiratory syndrom...	623	623	100%	0.0	96.08%	307	2A5K_A
☒	Chain A, 3C-like proteinase [Severe acute respiratory syndrome-related coronavirus]	Severe acute respiratory syndrom...	623	623	100%	0.0	96.08%	306	3E91_A
☒	Chain A, 3C-like proteinase [Severe acute respiratory syndrome-related coronavirus]	Severe acute respiratory syndrom...	623	623	100%	0.0	96.08%	306	1UJ1_A
☒	Chain A, 3C-like proteinase [Severe acute respiratory syndrome-related coronavirus]	Severe acute respiratory syndrom...	623	623	100%	0.0	96.08%	311	1WOF_A
☒	Chain A, 3C-like proteinase [Severe acute respiratory syndrome-related coronavirus]	Severe acute respiratory syndrom...	622	622	100%	0.0	96.08%	307	6XHL_A
☒	Chain A, 3C-like proteinase [Severe acute respiratory syndrome-related coronavirus]	Severe acute respiratory syndrom...	622	622	100%	0.0	95.75%	308	3F9E_A
☒	Chain A, 3C-like proteinase [SARS coronavirus BJ01]	SARS coronavirus BJ01	622	622	100%	0.0	95.75%	308	2PWX_A
☒	Chain A, 3C-like proteinase [Severe acute respiratory syndrome-related coronavirus]	Severe acute respiratory syndrom...	621	621	100%	0.0	95.75%	316	5B6Q_A
☒	Chain A, Replicase polyprotein 1a [Severe acute respiratory syndrome coronavirus]	Severe acute respiratory syndrom...	621	621	100%	0.0	96.07%	325	8ULD_A
☒	Chain A, 3C-like proteinase [unidentified]	unidentified	621	621	100%	0.0	95.75%	308	2QC2_A
☒	Chain A, 3C-like proteinase [Severe acute respiratory syndrome-related coronavirus]	Severe acute respiratory syndrom...	621	621	100%	0.0	95.75%	308	3M3T_A
☒	Chain A, 3C-like proteinase [Severe acute respiratory syndrome-related coronavirus]	Severe acute respiratory syndrom...	620	620	100%	0.0	95.75%	306	3M3S_A
☒	Chain A, 3C-like proteinase [Severe acute respiratory syndrome-related coronavirus]	Severe acute respiratory syndrom...	620	620	100%	0.0	95.75%	315	4H13_A
☒	Chain A, 3C-like proteinase [Severe acute respiratory syndrome-related coronavirus]	Severe acute respiratory syndrom...	620	620	100%	0.0	95.75%	306	3E9A_A
☒	Chain A, 3C-like proteinase [Severe acute respiratory syndrome-related coronavirus]	Severe acute respiratory syndrom...	620	620	100%	0.0	95.75%	306	1Z1J_A
☒	Chain A, 3C-like proteinase [Severe acute respiratory syndrome-related coronavirus]	Severe acute respiratory syndrom...	620	620	100%	0.0	95.75%	308	3F9F_A
☒	Chain A, 3C-like proteinase [Severe acute respiratory syndrome-related coronavirus]	Severe acute respiratory syndrom...	620	620	100%	0.0	95.75%	306	2QCY_A
☒	Chain A, severe acute respiratory syndrome coronavirus (SARS-CoV) [SARS coronavirus BJ01]	SARS coronavirus BJ01	620	620	100%	0.0	95.75%	306	2Q6G_A
☒	Chain A, 3C-Like Proteinase [Severe acute respiratory syndrome-related coronavirus]	Severe acute respiratory syndrom...	620	620	100%	0.0	95.75%	306	3ATW_A

Figura 4.2. Captura de pantalla del servidor BLAST que muestra los resultados de la búsqueda de secuencias homólogas a SCV2-Mpro con estructura tridimensional resuelta. Se aplicó un segundo criterio de filtrado, consistente en descartar las secuencias con **Percent Identity** mayor a 97 % y menor a 50 %. Para cada resultado se indican una breve descripción del sistema molecular, el nombre científico del organismo, el porcentaje de identidad con la SCV2-Mpro, el valor de *E-value* y del *Query Coverage*.

Por el contrario, *Clustal Omega* (<https://www.ebi.ac.uk/jdispatcher/msa/clustalo>) realiza alineamientos múltiples de tipo global, es decir, toma en cuenta toda la longitud de cada secuencia, lo que permite una mejor evaluación de la similitud estructural y funcional entre proteínas homólogas. Este tipo de alineamiento es más adecuado para calcular el porcentaje de identidad global, necesario para seleccionar una buena plantilla para modelado por homología.



A continuación, se detallan los pasos para ejecutar el alineamiento múltiple:

- I. Acceder al sitio web de *Clustal Omega*, <https://www.ebi.ac.uk/jdispatcher/msa/clustalo>
- II. En la sección **Input sequence**, pegar todas las secuencia contenidas en el archivo *seqdump.txt*, asegurándose de mantener el formato *fasta*
- III. Expandir la sección **More options** haciendo clic en el título
- IV. En el parámetro **ORDER**, seleccionar la opción **input**, lo cual mantendrá el orden de entrada original de las secuencias en el alineamiento resultante
- V. Verificar que todos los demás parámetros permanezcan con sus valores predeterminados
- VI. Hacer clic en **Submit** para iniciar el alineamiento (ver Figura 4.3)

Una vez finalizado el proceso, se mostrará la página de resultados. En la sección **Result Files**, se debe hacer clic en el enlace correspondiente a **Percent Identity Matrix**, que contiene la matriz de porcentajes de identidad global entre todas las secuencias alineadas.

Esta matriz permitirá identificar cuál de las proteínas candidatas posee la mayor similitud global con la SCV2-Mpro, lo que la convierte en la mejor plantilla para modelado por homología.

Actividad 4.2. Análisis de porcentajes de identidad. Tomar una captura de pantalla de la matriz de identidad y pegarla en la bitácora electrónica.

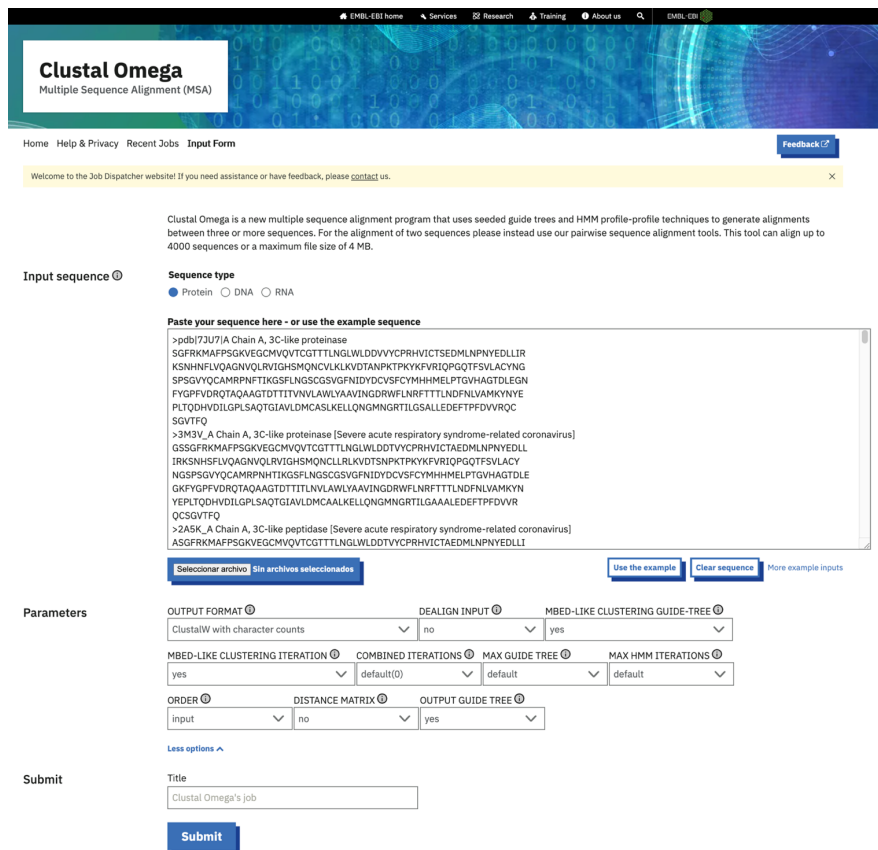
Determinar cuántas secuencias presentan un porcentaje de identidad mayor al 90 % con respecto a la SCV2-Mpro. Registrar estos valores en la bitácora electrónica.

Comparar los porcentajes de identidad reportados por *BLAST* y por *Clustal Omega* e identificar cualquier discrepancia. En caso de diferencias, comentar posibles causas (método de alineamiento, regiones truncadas, etc.).

Identificar los PDB ID de tres estructuras en forma libre (*apo*) que presenten el mayor porcentaje de identidad global y que correspondan a la forma nativa (no mutantes) de la Mpro. Indicar también la resolución reportada para cada una de esta estructuras.

¿Cuál estructura es más adecuada para utilizarse como plantilla (*template*) para el modelado por homología de la SCV2-Mpro?

El siguiente paso consiste en descargar, desde el portal del PDB, las estructuras tridimensionales identificadas. Adicionalmente, descargar el archivo con PDB ID 7NTT, que corresponde a la estructura de la SCV2-Mpro obtenida mediante XRC con resolución igual a 1.74 Å. Esta estructura será utilizada en las etapas posteriores de comparación estructural y validación del modelo por homología obtenido. Guardar las estructuras en la carpeta de trabajo.



Clustal Omega
Multiple Sequence Alignment (MSA)

Home Help & Privacy Recent Jobs Input Form Feedback

Welcome to the Job Dispatcher website! If you need assistance or have feedback, please [contact us](#).

Clustal Omega is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between three or more sequences. For the alignment of two sequences please instead use our pairwise sequence alignment tools. This tool can align up to 4000 sequences or a maximum file size of 4 MB.

Input sequence

Sequence type
☒ Protein ☐ DNA ☐ RNA

Paste your sequence here - or use the example sequence

```
>pdb173U71A Chain A, 3C-like proteinase
SGFRKMAFPGSGKEGCMQVTCGTTTLNGLWLDVYCPRHVICTSEMLNPNYEDLLR
KSNHNFVLQAGNVQLRVIGHSMQNCVLKLVDTANPKTKYKVRIPQGTFSVLACNG
SPSGVYQCAMRPNFTIKGSLNGSCGVGNIDYDCVFCYMHMELPTGVHAGTDLEGN
FYGPFDVDRQTAQAAGTDTTITNLVLAWYAAVINGDRWFLNRFTTLDNDFNLVAMKYN
PLTQDHVDILGLPSAQGTGLAVLDMCASLKELLQNGMNGRTILGSALEDEFTPFDDVRQC
SGVTFQ
>3M3V_A Chain A, 3C-like proteinase [Severe acute respiratory syndrome-related coronavirus]
GSSGFRKMAFPGSGKEGCMQVTCGTTTLNGLWLDVYCPRHVICTSEMLNPNYEDLL
IRKSNHNFVLQAGNVQLRVIGHSMQNCVLKLVDTANPKTKYKVRIPQGTFSVLACNG
NGSPGVYQCAMRPNFTIKGSLNGSCGVGNIDYDCVFCYMHMELPTGVHAGTDLE
GKPYGPFVDVDRQTAQAAGTDTTITNLVLAWYAAVINGDRWFLNRFTTLDNDFNLVAMKYN
YEPKQDHVDILGLPSAQGTGLAVLDMCAALKELLQNGMNGRTILGSALEDEFTPFDDVR
QCSGVTFQ
>2ASK_A Chain A, 3C-like peptidase [Severe acute respiratory syndrome-related coronavirus]
ASGFRKMAFPGSGKEGCMQVTCGTTTLNGLWLDVYCPRHVICTSEMLNPNYEDLLI
```

[Seleccionar archivo](#) [Sin archivos seleccionados](#) [Use the example](#) [Clear sequence](#) [More example inputs](#)

Parameters

OUTPUT FORMAT **DEALIGN INPUT** **MBED-LIKE CLUSTERING GUIDE-TREE**
 ClustalW with character counts no yes

MBED-LIKE CLUSTERING ITERATION **COMBINED ITERATIONS** **MAX GUIDE TREE** **MAX HMM ITERATIONS**
 yes default(0) default default

ORDER **DISTANCE MATRIX** **OUTPUT GUIDE TREE**
 input no yes

[Less options](#)

Submit

Title
 Clustal Omega's job

Submit

Figura 4.3. Captura de pantalla de la interfaz del servidor *Clustal Omega* tras introducir la secuencia de la SCV2-Mpro y 32 secuencias homólogas obtenidas mediante *BLAST*. En el recuadro **Input sequence** se muestran las tres primeras secuencias en formato *fasta*. En el apartado **More options** fue seleccionada la opción **ORDER = input** para mantener el orden original de ingreso de las secuencias. Al presionar el botón **Submit**, se inicia el proceso de alineamiento múltiple, que generará la **Percent Identity Matrix**, útil para comparar globalmente las secuencias y seleccionar la mejor plantilla para modelado por homología.

4.2 Construcción del modelo de la SCV2-Mpro dimérica

A continuación, se utilizará un programa de inteligencia artificial que permite predecir estructuras tridimensionales de proteínas mediante un sistema de aprendizaje profundo, denominado *AlphaFold2*. Esta herramienta fue desarrollada por DeepMind (filial de Alphabet/Google) y ha superado considerablemente en precisión a otros algoritmos de modelado estructural de proteínas. Se puede acceder gratuitamente a *AlphaFold2* a través de:

<https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb>

Considerando que la forma activa de la SCV2-Mpro es un homodímero, lo adecuado es contar con un modelo que refleje esta circunstancia. Para ello se deberá crear un

archivo de texto que contenga dos veces la secuencia de la SCV2–Mpro, separadas con el símbolo de dos puntos (:) (Figura 4.4). Esto permite a *AlphaFold2* interpretar que se desea modelar un dímero homólogo (A:A). La estructura del archivo debe ser la siguiente:

```
SGFRKMAFSPGKVEGCMVQVTCGTTTLNGLWLDVVYCPRAVICTSEDMLNPNYEDLLIRKSNHNFLVQAGNVQL
RVIGHSMQNCVLKLVDTANPKTPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMPNFTIKGSFLNGSCGSGVGF
NIDYDCVSFCYMHMELPTGVHAGTDLEGNFYGPFVDRQTAQAAGTDTTITVNVLAWLAAVINGDRWFLNRFTT
TLNDFNLVAMKYNIEPLTDHVDILGPLSAQTGIAVLDMCASLKELLQNGMNGRTILGSALLEDEFTPFDDVVRQC
SGVTFO:SGFRKMAFSPGKVEGCMVQVTCGTTTLNGLWLDVVYCPRAVICTSEDMLNPNYEDLLIRKSNHNFLV
QAGNVQLRVIGHSMQNCVLKLVDTANPKTPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMPNFTIKGSFLNG
SCGSGVGFNIDYDCVSFCYMHMELPTGVHAGTDLEGNFYGPFVDRQTAQAAGTDTTITVNVLAWLAAVINGDRW
FLNRFTTTLNDFNLVAMKYNIEPLTDHVDILGPLSAQTGIAVLDMCASLKELLQNGMNGRTILGSALLEDEFTP
FDVVRQCSGVTFO
```

Figura 4.4. Ejemplo de cómo ingresar la secuencia para construir el modelo de la SCV2–Mpro en su forma de homodímero. Se muestra el formato correcto en el que deben introducirse dos copias consecutivas de la secuencia de aminoácidos, separadas por el símbolo de dos puntos (:), según los requerimientos del servidor ColabFold basado en *AlphaFold2*. Este formato indica que se desea predecir la estructura tridimensional de un homodímero de la proteasa principal del SARS-CoV-2. El color del texto ha sido utilizado con fines ilustrativos para señalar el inicio y final de cada secuencia.

Esta información debe ser ingresada en el campo **query_sequence**, localizado en la sección **Input protein sequences (s), then...** (ver Figura 4.5). A continuación, deben ajustarse los parámetros del **num_relax** a 5 y **template_mode** a *custom*. El parámetro *num_relax* indica cuántas estructuras de las estructuras predichas serán refinadas energéticamente mediante AMBER, un software de simulación molecular que permite corregir problemas estéricos menores sin modificar el plegamiento global de la proteína.

Por su parte, *template_mode* define la fuente de plantillas estructurales homólogas que *AlphaFold2* empleará para mejorar la predicción. Al seleccionar la opción *custom*, *AlphaFold2* utilizará exclusivamente la plantilla personalizada proporcionada por el usuario. El resto de los parámetros deben permanecer sin modificaciones.

En la parte superior de la página, hacer clic en **Entorno de ejecución** y luego en **Ejecutar todo**.

La pantalla permanecerá prácticamente igual al principio, excepto por la activación de la casilla **Elegir archivos**, ubicadas justo debajo del parámetro **template_mode**, donde se deberá subir el archivo correspondiente al *template*.

Seleccionar el archivo *1q2w.pdb*, la estructura más adecuada identificada en la sección anterior (las otras dos estructuras fueron *1uj1.pdb* y *5b6o.pdb*), para iniciar el modelado. Es fundamental que el PDB ID del archivo esté en minúsculas.

Después de algunos segundos, el ícono de **Play** situado junto a la sección **Input protein sequences(s), then...** cambiará a un ícono de **Stop**, el cual comenzará a girar, indicando que la tarea está en ejecución. Una vez que el proceso haya terminado, el

ícono regresará a **Play**. Eventualmente, lo mismo sucederá con el resto de las secciones. El proceso completo de modelado requiere aproximadamente entre 20 y 30 minutos. Al finalizar, se descargará automáticamente al equipo del usuario una carpeta con los cinco modelos tridimensionales generados y optimizados, ordenados según una métrica interna de calidad, así como archivos con información sobre la calidad del modelo. El modelo con mejor puntuación (usualmente etiquetado como nombre_relaxed_rank_001_alphafold2_multimer_v3_model_5_seed_000.pdb) se considera el más confiable y será el que se utilice en análisis posteriores.

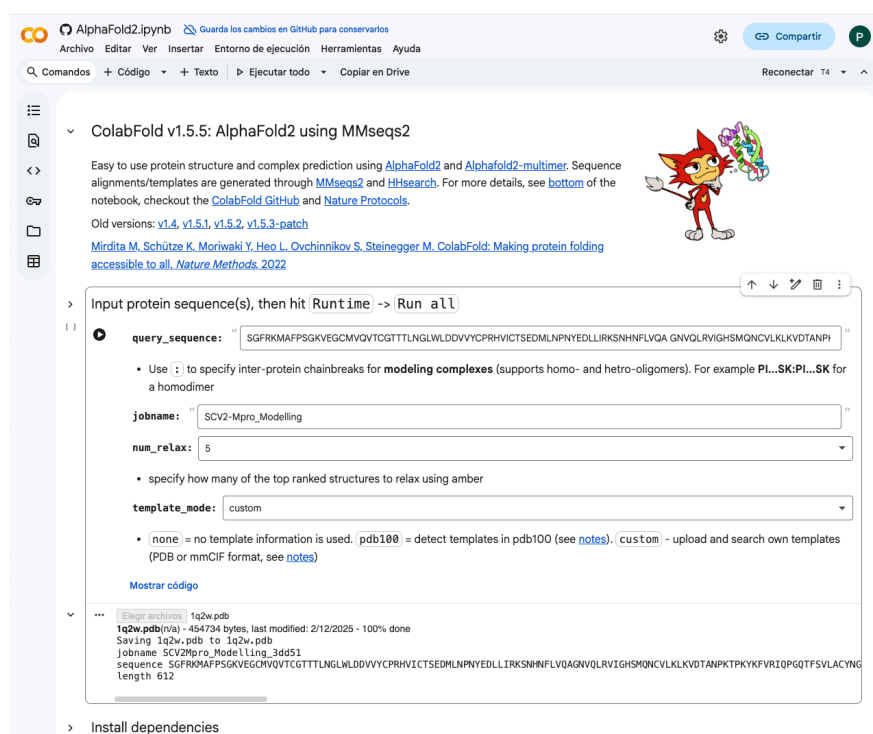


Figura 4.5. Captura de pantalla de la interfaz del servidor *AlphaFold2* que muestra la primera sección en la cual se pueden modificar los parámetros para la construcción del modelo de la SCV2-Mpro, así como la ubicación del enlace utilizado para cargar el archivo *1q2w.pdb*.

4.3 Validación y comparación del modelo

La calidad de la estructura tridimensional de una proteína ya sea experimental o modelada, puede evaluarse utilizando el servidor en línea *SAVES v6.0* (<https://saves.mbi.ucla.edu/>), el cual integra varias herramientas de validación estructural. Su utilización es directa:

- I. Cargar un archivo en formato *.pdb* haciendo clic en **Select File**. Se debe descomprimir la carpeta descargada desde *AlphaFold2* y seleccionar el archivo cuyo nombre contiene el termino *rank_001* (por ejemplo,



SCV2Mpro_Modelling_3dd51_relaxed_rank_001_alphafold2_multimer_v3_model_5_seed_000), para su análisis en *SAVES v6.0*

II. Iniciar el análisis haciendo clic en **Run programs**

La página de resultados será similar a la presentada en la Figura 4.6. Para visualizar los distintos resultados del análisis, se debe hacer clic en **Start** en las subrutinas *ERRAT*, *Verify3D*, *WHATCHECK* y *PROCHECK*.

Brevemente:

- I. *ERRAT*: evalúa la calidad de modelos tridimensionales de proteínas mediante el análisis de interacciones entre átomos no enlazados. Detecta regiones con desviaciones significativas del comportamiento esperado, lo que resulta útil para identificar errores en la geometría local de la estructura
- II. *Verify3D*: compara la estructura tridimensional del modelo con su secuencia primaria de aminoácidos. Evalúa si cada residuo se encuentra en un entorno estructural apropiado para su tipo, según datos estadísticos derivados de estructuras experimentales depositadas en el PDB
- III. *WHATCHECK*: realiza un análisis exhaustivo de los parámetros estructurales, incluyendo ángulos diedros, longitudes de enlace, geometría de grupos funcionales, residuos anómalos, entre otros. Esta herramienta es útil para detectar errores sutiles durante la construcción del modelo tridimensional
- IV. *PROCHECK*: evalúa la calidad estereoquímica de la proteína, prestando especial atención a los ángulos ϕ (phi) y ψ (psi) de la cadena principal. Su resultado más importante es el diagrama de Ramachandran, el cual indica qué proporción de residuos se encuentra en conformaciones energéticamente permitidas o prohibidas

Actividad 4.3. Interpretar los resultados obtenidos en las diferentes subrutinas y elaborar un breve informe sobre la calidad del modelo generado con *AlphaFold2*. Con base en dichos resultados, justificar si el modelo puede considerarse adecuado o si presenta limitaciones importantes que deban señalarse.

Para comprobar la validez estructural del modelo generado mediante *AlphaFold2*, se realizará un alineamiento estructural con la estructura experimental de proteasa principal del SARS-CoV-2 con PDB ID 7NTT, previamente descargada, utilizando UCSF Chimera mediante la herramienta **MatchMaker**, tal como se describe en la **Práctica 2. Visualización de proteínas**.

Una vez completado el alineamiento, se debe registrar el valor del RMSD parcial y global, lo cual permitirá evaluar el grado de similitud entre el modelo generado y la estructura experimental.

UCLA-DOE LAB — SAVES v6.1

UCLA

Job 393066 has been created

[New Job](#)

job #393066:

SCV2Mpro_Modelling_3dd51_relaxed_rank_001_alphafold2_multimer_v3_model_5_seed_000.pdb

[\[job link\]](#) [\[3D Viewer\]](#)

ERRAT Analyzes the statistics of non-bonded interactions between different atom types and plots the value of the error function versus position of a 9-residue sliding window, calculated by a comparison with statistics from highly refined structures. Start	Verify3D Determines the compatibility of an atomic model (3D) with its own amino acid sequence (1D) by assigning a structural class based on its location and environment (alpha, beta, loop, polar, nonpolar etc) and comparing the results to good structures. Start	PROVE Temporarily down at the moment
WHATCHECK Derived from a subset of protein verification tools from the WHATIF program (Vriend, 1990), this does extensive checking of many stereochemical parameters of the residues in the model. Start	PROCHECK Checks the stereochemical quality of a protein structure by analyzing residue-by-residue geometry and overall structure geometry. Start	OPEN We are open to suggestions for a 6th program to operate in this window. If you know of a program that would be most useful, please let us know: email holton at mbi dot ucla dot edu with your suggestion

References

ERRAT

- Reference: Verification of protein structures: patterns of nonbonded atomic interactions, Colovos C and Yeates TO, 1993.
- C++ software

VERIFY 3D

- Profile Search Software [Bowie et al., 1991, Luethy et al., 1992].
- DSSP original and Wikipedia

PROVE

- Reference: Deviations from standard atomic volumes as a quality measure for protein crystal structures, Pontius J, Richelle J, Wodak SJ, 1996

PROCHECK

- PROCHECK source information
- Result analysis
- from Protein Structures
- Original references

WHATCHECK

- WHATCHECK documentation and source
- For questions about WHATCHECK results, please email the creator: Prof. Gert Vriend vriendgert@gmail.com

Figura 4.6. Captura de pantalla de la interfaz de servidor *SAVES* v6.0 después de cargar el archivo *.pdb* y ejecutar el análisis. El servidor utiliza cuatro algoritmos para validar la calidad estructural del modelo tridimensional: *ERRAT*, *Verify3D*, *WHATCHECK* y *PROCHECK*. Cada recuadro de la imagen describe brevemente el propósito de cada uno de estos programas en la evaluación estructural del modelo generado.

Actividad 4.4. Repetir el alineamiento utilizando una segunda estructura, pero ahora empleando como referencia la estructura de la SCV2-Mpro unida a un inhibidor, por ejemplo, PDB ID 8DZB). Comparar los valores de RMSD obtenidos y discutir las posibles diferencias entre ambos alineamientos. La calidad del modelo de la SCV2-Mpro debe evaluarse con base en los valores del RMSD y en el alineamiento estructural observado.



Como referencia, en análisis previos se han obtenido resultados como: RMSD de 0.707 Å considerando 298 de los 306 residuos alineados, y RMSD de 1.34 Å considerando todos los residuos.

Bibliografía

- Jumper, J.**, Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589.
- Bitencourt-Ferreira, G** & W. F., Jr de Azevedo. (2019). *Homology Modeling of Protein Targets with MODELLER*. *Methods in molecular biology* (Clifton, N.J.), 2053: 231–249.
- Colovos, C** & T. O. Yeates. (1993). *Verification of protein structures: patterns of nonbonded atomic interactions*. *Protein science* 2(9), 1511-1519.
- Laskowski, R. A.**, MacArthur, M. W., Thornton J. M. (2001). *PROCHECK: validation of protein structure coordinates*, In: *International Tables of Crystallography*, Volume F. *Crystallography of Biological Macromolecules*, eds. Rossmann, M. G & Arnold, E. Dordrecht. Kluwer Academic Publishers. The Netherlands pp. 722-725.
- Muhammed, M. T.** & E. Aki-Yalcin. (2019). *Homology modeling in drug discovery: Overview, current applications, and future perspectives*. *Chemical biology & drug design*, 93(1): 12–20.
- Pitman, M. R.** & R. I. Menz. (2006). *Methods for protein homology modelling*. In. *Applied Mycology and Biotechnology*. Elsevier. 6, 37-59.
- Pontius, J.**, Richelle, J & S. J. Wodak. (1996). *Deviations from standard atomic volumes as a quality measure for protein crystal structures*. *Journal of molecular biology* 264(1): 121-136.



Práctica 5. Construcción y análisis de la flexibilidad conformacional del 1,3-dimetilciclohexano

Objetivos

- Construir un modelo tridimensional del 1,3-dimetilciclohexano e identificar su isomería cis/trans y la orientación axial/ecuatorial de sus grupos metilo.
- Analizar su flexibilidad conformacional mediante simulaciones breves de dinámica molecular y describir los cambios observados.
- Relacionar la flexibilidad estructural con los modelos de reconocimiento molecular proteína-ligando (cerradura-llave, ajuste inducido, selección conformacional) y su relevancia en el diseño de ligandos.

Software especializado requerido

- UCSF Chimera versión 1.18 o superior

Visualizar una estructura tridimensional a partir de un dibujo plano puede ser complejo, incluso para personas con formación química avanzada, especialmente al considerar las múltiples conformaciones que una molécula puede adoptar en el espacio. Este reto se acentúa al analizar compuestos flexibles o cíclicos, donde los efectos estéricos locales y la orientación axial o ecuatorial influyen en la estabilidad. Un ejemplo representativo es el 1,3-dimetilciclohexano, cuya inversión de silla y variantes conformacionales permiten explorar conceptos como interacciones 1,3-diaxiales (ver Figura 5.1), tensión conformacional y diferencias de energía libre entre estados alternantes.

Este ejercicio permite construir una intuición geométrica y energética aplicable a sistemas más complejos, como la unión de un fármaco a su blanco o los cambios estructurales que experimenta una proteína durante su dinámica. Los principios abordados aquí son fundamentales para interpretar el muestreo conformacional en dinámica molecular, analizar trayectorias mediante RMSD/RMSF, y diseñar fármacos a través de acoplamiento molecular, cribado virtual o fragment-based drug design. Cuando sea posible, se recomienda construir modelos moleculares físicos para facilitar la visualización de la geometría, estereoquímica y flexibilidad antes del análisis computacional.

Actividad 5.1. Analizar cuál de las conformaciones mostradas es la más estable y justificar la elección con base en criterios estructurales y energéticos (por ejemplo, disposición axial o ecuatorial de los sustituyentes y efectos estéricos). Asimismo, Investigar y reportar el nombre de cada una de las conformaciones mostradas.

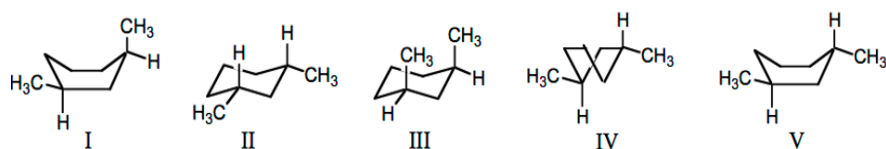


Figura 5.1. Modelos tridimensionales del 1,3-dimetilciclohexano que ilustran las conformaciones más comunes adoptadas por esta molécula. Se muestran las formas tipo silla con las posibles orientaciones axial y ecuatorial de los grupos metilo en las posiciones 1 y 3 del anillo ciclohexano.

5.1 Construyendo la molécula en UCSF Chimera

Para comenzar, inicie UCSF Chimera y diríjase a **Tools>Structure Editing>Build Structure**. En la ventana que aparece, seleccionar las opciones **atom** y **Center of view**, y hacer clic en **Apply** sin modificar los parámetros predeterminados. En la ventana principal aparecerá un átomo de helio (He) representado dentro del modelo *scratch*. En la parte superior de la ventana **Build Structure**, localizar el menú desplegable **Start Structure** y cámbielo a **Modify Structure**. A continuación, en **Change selected atoms to...**, seleccione **C**, con **4 Bonds** y **Geometry: tetrahedral**, y presionar **Apply**.

El átomo de He se convertirá en molécula de metano. Seleccionar uno de los átomos de hidrógeno y, utilizando nuevamente la opción **Modify Structure**, sustituirlo por un átomo de carbono tetraédrico haciendo clic **Apply**. Repetir este procedimiento con los hidrógenos adecuados hasta construir el anillo 1,3-dimetilciclohexano. Si los átomos fueron colocados correctamente, el enlace que cierra el anillo de seis carbonos se formará automáticamente (ver Figura 5.2). En caso contrario, utilizar la opción **Adjust Bonds** desde el mismo menú desplegable para completar la estructura manualmente.

Para mejorar la visualización de la geometría molecular, se recomienda cambiar la representación a *ball-and-stick* (esferas y varillas).

Una vez construida la molécula, se debe determinar si la conformación generada corresponde al isómero *cis* o *trans* del 1,3-dimetilciclohexano. En caso de que la conformación no sea la deseada, esta deberá ajustarse utilizando las herramientas de eliminación de átomos o edición de enlaces disponibles en UCSF Chimera. Es importante asegurarse de trabajar con el isómero *cis*, correspondiente a la conformación II mostrada en la Figura 5.1.

5.2 Flexibilidad estructural del 1,3-dimetilciclohexano

Un aspecto en química orgánica consiste en determinar cuál es la conformación de menor energía del isómero *cis*-1,3-dimetilciclohexano. Aunque esta pregunta tiene una respuesta estructural bien definida, es importante recordar que las moléculas no son entidades rígidas, sino que, en realidad, presentan una dinámica conformacional constante, adoptando múltiples conformaciones a lo largo del tiempo.

UCSF Chimera permite explorar esta flexibilidad estructural mediante una herramienta de dinámica molecular (DM, en la siguiente práctica daremos una breve explicación sobre ese tema). Para llevar a cabo esta simulación:

- I. Acceder a **Tools>MD&Ensemble Analysis>Molecular Dynamics Simulation**
- II. Utilizar la configuración predeterminada del módulo
- III. En la pestaña **Run Parameters**, hacer clic en **Run** para iniciar la simulación

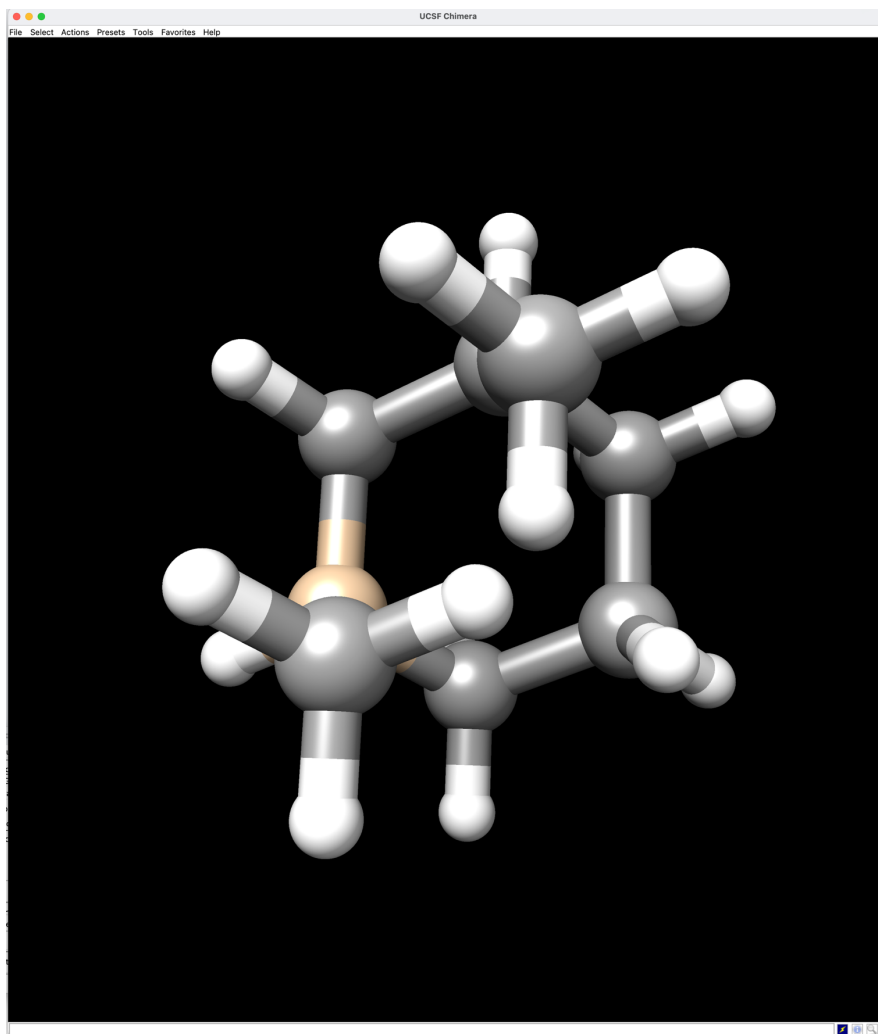


Figura 5.2. Modelo tridimensional del 1,3-dimetilciclohexano en representación en esferas y varillas. Imagen generada con UCSF Chimera version 1.18.

En caso de que el programa solicite agregar átomos de hidrógeno o asignar cargas parciales a la molécula, aceptar los valores predeterminados y hacer clic **Ok** en las tres ventanas emergentes consecutivas. La simulación de DM se ejecutará en pocos segundos. Si fuera el caso, cerrar la ventana de advertencia **Chimera Warning**. Una vez finalizada la simulación, se abrirá la ventana **MD Movie: Dynamic Trajectory**.



Presionar **Play** para visualizar el movimiento conformacional de la molécula a lo largo del tiempo.

Durante la ejecución de la DM, UCSF Chimera genera varios archivos temporales en el escritorio, entre ellos *prod.log*, *prod.nc*, *beat_res.nc*, *beating.log* y *beating.nc*. Se recomienda eliminarlos antes de repetir una nueva simulación para evitar conflictos o sobreescritura de datos.

Actividad 5.2. Avanzar hasta el último cuadro (*frame*) de la simulación y evaluar si la conformación obtenida corresponde a la de menor energía del *cis*-1,3-dimetilciclohexano. Explicar qué criterios se utilizaron para establecerlo, considerando aspectos como la energía potencial, la disposición axial o ecuatorial de los sustituyentes y la estabilidad conformacional observada.

La simulación de DM permite ilustrar conceptos fundamentales sobre la interacción entre proteínas y ligandos, como los modelos de *Cerradura-Llave* o *Ajuste Inducido*. Tradicionalmente, se ha concebido al ligando (llave) y al sitio activo de la enzima (cerradura) como estructuras rígidas que encajan perfectamente. No obstante, este enfoque simplifica en exceso la realidad molecular. En verdad, las moléculas no son estructuras estáticas: están en constante movimiento, adoptando diversas conformaciones. Algunas enzimas, por ejemplo, unen al sustrato en una conformación “tensa”, forzándolo en la dirección deseada de la reacción que catalizan. Además, existe un mecanismo conocido como *selección conformacional*, en el cual el sitio activo de una enzima fluctúa entre múltiples estados estructurales, pero solo una de estas conformaciones es capaz de reconocer y alojar al sustrato. Este fenómeno contribuye de forma crucial a la especificidad y eficiencia catalítica de muchas enzimas. En conjunto, estos conceptos evidencian la importancia de considerar la flexibilidad estructural tanto del ligando como de la proteína en el diseño de fármacos y en la comprensión de mecanismos enzimáticos.

El modelo de *Ajuste Inducido* ofrece una descripción más precisa del reconocimiento proteína–ligando que el modelo de *Cerradura-Llave*. A diferencia de este último, que asume estructuras rígidas y perfectamente complementarias, el modelo de Ajuste Inducido considera que ambas moléculas son flexibles y capaces de adaptarse mutuamente para lograr una complementariedad geométrica óptima. Esta perspectiva dinámica nos recuerda que muchas moléculas, en particular los compuestos orgánicos como el 1,3-dimetilciclohexano, presentan una amplia flexibilidad conformacional. Esta capacidad de adoptar diversas conformaciones implica que, durante una interacción molecular, una molécula puede inducir cambios estructurales en la otra, facilitando el reconocimiento y la unión específica. Para explorar esta flexibilidad de manera práctica, se debe seleccionar el enlace que une uno de los grupos metilo al anillo y dirigirse a **Tools>Structure Editing>Adjust**



Torsions. En la ventana que se despliega, hacer clic en **Activate selected bond as torsion** para habilitar la rotación en ese enlace específico.

Actividad 5.3. Localizar el ícono con apaciencia de velocímetro ubicado en el panel derecho de la interfaz. Observar y describir el efecto de modificar el ángulo de torsión utilizando el control de rotación.

En el menú desplegable situado debajo de **Bond**, seleccionar la opción **reverse** y arrastrar la aguja alrededor del velocímetro. Describir qué ocurre al emplear la función *reverse* y explicar su efecto sobre la rotación del enlace y la conformación molecular resultante.

5.3 Obtención del conformero de menor energía

Con el objeto de identificar la conformación más estable del *cis*-1,3-dimetilciclohexano, se simulará una transición estructural desde la conformación *cis* (II) hacia la conformación *trans* (III) como se muestra en la Figura 5.1. Se debe seleccionar un enlace C–C del anillo que sea adyacente a un carbono que posea un grupo metilo. Para ello accede a **Tools>Structure Editing>Build Structure>Adjust Bonds**, utilizar la opción **Delete** para eliminar ese enlace. La estructura del ciclohexano deberá visualizarse abierta.

Ahora se debe identificar y seleccionar el enlace C-C que permanece intacto, ubicado entre el carbono con grupo metilo y su carbono adyacente (ver Figura 5.3 como referencia).

Luego, abrir **Tools>Structure Editing>AdjustTorsions** y hacer clic en **Activate selected bond as torsion**. Se debe utilizar el control de rotación para ajustar la posición del grupo metilo a una posición axial. Repetir este procedimiento para el segundo grupo metilo, asegurándose también de colocarlo en posición axial. A pesar de tener enlaces rotos o posibles superposiciones entre átomos, esta modificación es temporal y tiene como fin lograr la conformación deseada del isómero (conformación III, Figura 5.1).

Para cerrar nuevamente el anillo de ciclohexano, seleccionar los dos pares de átomos que deben enlazarse (ver Figura 5.4) y dirigirse nuevamente a **Adjust Bonds**. En cada caso, seleccionar la opción **Add all possible bonds between selected atoms** para formar los enlaces faltantes. Para una mejor apreciación de la geometría final, se debe cambiar la representación de la molécula a esferas seleccionando **Actions>Atoms/Bonds>sphere**.

En química orgánica, las interacciones 1,3-diaxiales, son un fenómeno clave en el análisis de conformaciones del ciclohexano. Estas interacciones ocurren cuando sustituyentes en posición axial se encuentran en carbonos 1 y 3, generando una repulsión estérica debido a la proximidad espacial de los grupos, lo cual puede afectar significativamente la estabilidad de la molécula.

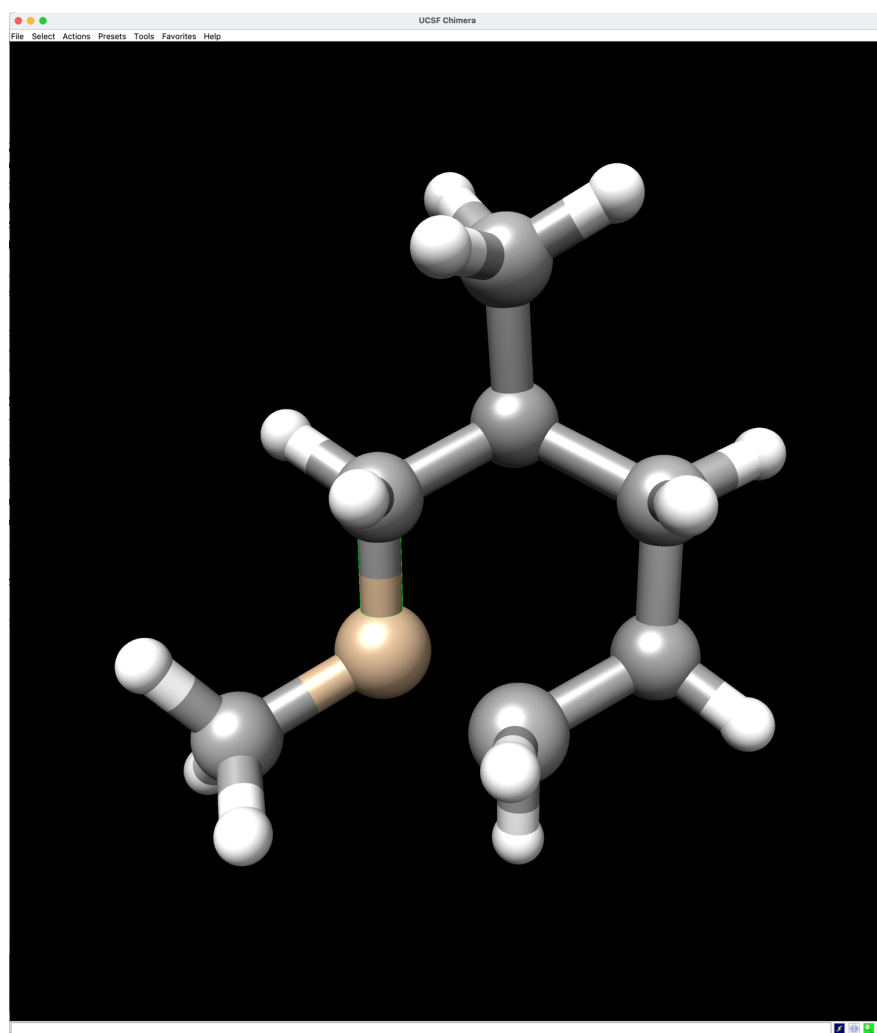


Figura 5.3. Modelo tridimensional del 1,3-dimetilciclohexano con el anillo abierto. Se resalta el enlace intacto entre un carbono del anillo y uno de los grupos metilo. Imagen generada con UCSF Chimera version 1.18.

Para explorar si en la conformación III del cis-1,3-dimetilciclohexano se produce colisión entre los dos grupos metilo en posición axial, realiza los siguientes pasos:

- I. Modificar la visualización: cambiar la representación de la molécula a *ball-and-stick* para facilitar la observación de la geometría molecular: **Actions>Atoms/Bonds >ball&stick**
- II. Simulación de DM: acceder a **Tools>MD & Ensemble Analysis>Molecular Dynamics Simulation**
- III. Configuración de la simulación: en la pestaña **Prep Structure**, hay que asegurar que esté seleccionada la opción **Neither memorize nor use memorized options**. Luego, dirigirse a la pestaña **Run Parameters** y desactivar la casilla **Minimize before MD** (esto permitirá simular la conformación tal como fue construida, sin minimizarla previamente)

IV. Ejecutar la simulación: hacer clic en **Run** para iniciar la simulación

Representar nuevamente la molécula como *ball-and-stick* y vayan a **Tools>MD&EnsembleAnalysis>Molecular Dynamics Simulation**. Asegúrense de que en la pestaña **Prep Structure** esté seleccionado **Neither memorize nor use memorized options**. En la pestaña **Run Parameters** desactiven la opción **Minimize before MD** y luego presionen **Run**.

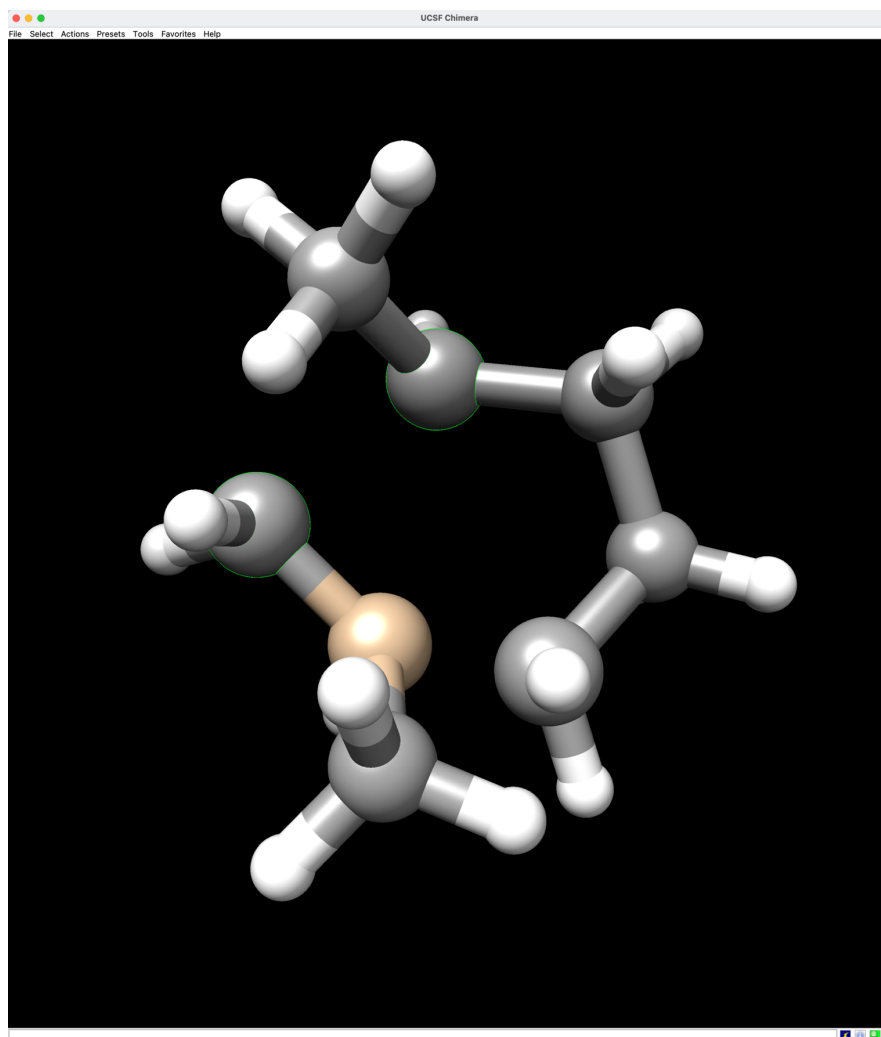


Figura 5.4. Modelo del isómero cis-1,3-dimetilciclohexano con ambos grupos metilo en posición axial. Se muestran dos enlaces del anillo ciclohexano aún sin cerrar. Imagen generada con UCSF Chimera version 1.18.

Durante la simulación, se debe observar cuidadosamente si los dos grupos metilo en posición axial se aproximan lo suficiente como para generar choques estéricos, tensión conformacional o incluso distorsiones en la geometría del anillo ciclohexánico (por ejemplo, desviaciones del arreglo tipo silla). Este análisis permite



visualizar de manera dinámica cómo las interacciones estéricas influyen en la estabilidad relativa de las conformaciones y favorecen aquellas con menor repulsión entre sustituyentes. En conjunto, el estudio conformacional de moléculas pequeñas constituye una base conceptual para comprender la flexibilidad estructural de macromoléculas como las proteínas, la cual se abordará en las siguientes prácticas mediante simulaciones de dinámica.

Actividad 5.4. En el **MD Movie viewer** (si la ventana está cerrada, localizarla en la parte inferior del menú **Tools**), observar el primer y último paso de la simulación y describir en qué se diferencian.

Evaluar si UCSF Chimera permitió alcanzar la conformación totalmente ecuatorial (conformación II de la Figura 5.1).

Posteriormente, mediante alguna combinación adecuada de las herramientas **Built Structure** y **Minimize before MD**, crear una representación de *cis*-1,3-dimetilciclohexano en su conformación de menor energía.

Representar el modelo como una estructura de esferas y varillas, en color gris oscuro, coloreada por heteroátomo, sobre un fondo blanco.

(Opcional). Si tiene acceso a modelos moleculares físicos, construir el modelo del *cis*-1,3-dimetilciclohexano para comparar la conformación obtenida.

Bibliografía

Brandt G. S. & Novak W. R. P. (2021) SARS-CoV-2 virtual biochemistry labs on bioinformatics and drug design. *Biochem Mol Biol Educ.* 49:26-28.

Pettersen, E. F., T. D. Goddard., C. C. Huang., G. S. Couch., D. M. Greenblatt., E. C. Meng., T. E. Ferrin, (2004). *UCSF Chimera—a visualization system for exploratory research and analysis.* *Journal of Computational Chemistry* 25(13):1605-12.



Práctica 6. Generación de conformaciones de una proteína mediante simulación de dinámica molecular

Objetivos

- Preparar un sistema proteico para simulación de dinámica molecular mediante GROMACS.
- Ejecutar la minimización, estabilización (NVT/NPT) y producción del sistema proteico, verificando su estabilidad con métricas termodinámicas.

Software especializado requerido

- GROMACS versión 2024.0 o posterior
- XMGrace versión 5.1 o posterior
- Python versión 3 o posterior

El acoplamiento molecular, también conocido como *docking*, es una técnica computacional que simula la interacción entre una molécula pequeña (el ligando) y el sitio de unión de una proteína. Este proceso implica generar múltiples conformaciones del ligando y evaluarlas en diferentes orientaciones (*poses*) dentro del sitio activo de la proteína. El objetivo es identificar la pose que maximice las interacciones intermoleculares, como enlaces de hidrógeno, fuerzas de *van der Waals* y enlaces iónicos, con los residuos clave del sitio de unión.

Una limitante común de este enfoque es la suposición de que la proteína permanece rígida durante la simulación. Esta simplificación contrasta con la realidad fisiológica, en la que las proteínas son entidades dinámicas que experimentan fluctuaciones conformacionales, incluso en ausencia del ligando. Esta rigidez asumida puede afectar la precisión de las predicciones sobre afinidad y modo de unión.

Para representar la flexibilidad intrínseca de la proteína, se recurre a la simulación de DM. Esta es una técnica computacional que permite estudiar el comportamiento de sistemas moleculares –como proteínas, ácidos nucleicos, fármacos o solventes– a lo largo del tiempo, modelando el movimiento de los átomos según las leyes de la física clásica, particularmente mediante la segunda ley de Newton.

En una simulación de DM:

- Cada átomo se describe por su posición y velocidad en el espacio tridimensional
- Las fuerzas y energías que actúan sobre los átomos se calculan a partir de un campo de fuerzas (*force field*), el cual define parámetros para enlaces, ángulos, torsiones y fuerzas no covalentes
- El sistema evoluciona en el tiempo mediante la integración numérica en pasos muy pequeños (del orden de picosegundos, 10^{-12} s)

Como resultado, se obtiene una trayectoria temporal del sistema, que permite visualizar y analizar cómo cambian las posiciones atómicas a lo largo de la simulación, en escalas que van típicamente desde nanosegundos (ns) hasta microsegundos (μ s).



La simulación de DM es una herramienta clave en Diseño de Fármacos, ya que permite incorporar el comportamiento dinámico y realista de los sistemas moleculares. Algunas de sus principales aplicaciones en este contexto incluyen:

- Estudiar la flexibilidad estructural de las proteínas blanco, lo que permite identificar conformaciones alternativas funcionalmente relevantes
- Observar cómo interactúa un ligando con su receptor más allá de una única estructura estática, como ocurre en el acoplamiento molecular tradicional
- Evaluar la estabilidad del complejo proteína–ligando, observando si el ligando permanece unido o se disocia durante la simulación
- Explorar sitios de unión ocultos o transitorios, que pueden no ser evidentes en estructuras cristalinas, pero que se revelan al considerar la dinámica conformacional
- Estimar propiedades termodinámicas, como la energía libre de unión, mediante enfoques como MM-PBSA, MM-GBSA o FEP
- Reducir el número de falsos positivos provenientes de cribados virtuales, descartando ligandos inestables o con interacciones poco realistas

Existen múltiples programas de simulación de DM, cada uno con ventajas particulares según el tipo de sistema, el nivel de experiencia del usuario y los recursos computacionales disponibles. A continuación, se listan algunos de los más utilizados en investigación biomolecular y en el diseño de fármacos:

- GROMACS: software de alto rendimiento, especialmente eficiente para estudiar proteínas y otros sistemas biológicos. Es gratuito, de código abierto y ampliamente documentado, lo que lo hace ideal para uso académico
- AMBER: muy utilizado en el estudio de biomoléculas. Destaca por sus herramientas especializadas para el cálculo de energía libre de unión entre fármacos y proteínas. Requiere licencia comercial, aunque accesible para universidades
- NAMD: diseñado para operar en supercomputadoras, con excelente escalabilidad y rendimiento. Es gratuito para uso académico y compatible con múltiples plataformas
- CHARMM: uno de los programas más completos y robustos en bioquímica computacional. Su uso requiere una licencia comercial
- Desmond: software rápido y con una interfaz gráfica amigable (Maestro, de Schrödinger). Es de uso comercial, pero se ofrece gratuitamente para instituciones educativas con licencia
- OpenMM: diseñado para aprovechar la aceleración por GPU. Es fácil de programar en Python, gratuito y de código abierto, ideal para usuarios con conocimientos básicos de programación en Python (scripting)



En esta práctica se utilizará GROMACS, debido a su rapidez, versatilidad, amplia documentación disponible y el hecho de ser software libre. Por lo tanto, deberá ser instalado en una computadora personal para el desarrollo del trabajo.

En el Apéndice I se proporcionan las instrucciones detalladas para la instalación de GROMACS en el sistema operativo Linux.

6.1 Ajuste de los estados de protonación

Antes de ejecutar una simulación de DM, es fundamental definir correctamente los estados de protonación de los grupos funcionales ionizables de la proteína blanco. Este paso es esencial para asegurar una representación realista de las condiciones fisiológicas, ya que una protonación incorrecta puede afectar las interacciones electrostáticas y la estabilidad del sistema.

Este procedimiento puede efectuarse mediante el servidor en línea *APBS* (Adaptative Poisson-Boltzmann Solver) disponible en <https://server.poissonboltzmann.org/>. Dentro de este servidor, se utiliza la subrutina PDB2PQR, una herramienta bioinformática que convierte automáticamente archivos en formato *.pdb* a *.pqr*, el cual contiene información adicional necesaria para cálculos electrostáticos, como cargas parciales y radios atómicos.

Durante el proceso de preparación, PDB2PQR realiza automáticamente las siguientes tareas:

- I. Incorporación de átomos pesados faltantes: se modelan los átomos ausentes en cadenas laterales o residuos incompletos, asegurando una estructura íntegra
- II. Estimación de los valores de pK_a : utiliza el algoritmo *PropKa* para predecir el estado de ionización de los grupos ácidos y básicos (como Asp, Glu, Lys, His, entre otros) en función del entorno estructural. Aunque *PropKa* asume un pH fisiológico (≈ 7.3), es posible modificar este parámetro para simular otras condiciones experimentales
- III. Asignación de cargas parciales: se asignan cargas atómicas y radios de *van der Waals* en función del campo de fuerzas seleccionado (por ejemplo, AMBER, CHARMM, PARSE), generando un archivo *.pqr* listo para análisis posteriores o simulaciones

Acceder al portal de *APBS* y hacer clic *PDB2PQR*. En la nueva ventana (ver Figura 6.1A), seleccionar la opción **Upload a PDB file** y hacer clic en **Select File** para cargar el archivo *.pdb* correspondiente al modelo tridimensional de la SCV2-Mpro obtenido previamente en la **Práctica 3. Modelado por homología**.

A

1 PDB2PQR Configuration > 2 PDB2PQR Job Status > 3 APBS Configuration > 4 APBS Job Status > 5 Visualization

PDB Selection

PDB Source

PDB ID [Upload a PDB file](#)

Please upload a PDB file

[Select File](#)

[mpo_model.pdb](#)

For continued support of this server, please register your use of this software:

[Register Here](#)

pKa Options

pH:

☐ No pKa calculation

☒ Use PROPKA to assign protonation states at provided pH

Forcefield Options

Please choose a forcefield to use

[AMBER](#) [CHARMM](#) [PEOEPB](#) [PARSE](#) [SWANSON](#) [TYL06](#) [User-defined Forcefield](#)

Please choose an output naming scheme to use

[Internal naming scheme](#) [AMBER](#) [CHARMM](#) [PARSE](#) [PEOEPB](#) [SWANSON](#) [TYL06](#)

Additional Options

- ☒ Ensure that new atoms are not rebuilt too close to existing atoms
- ☒ Optimize the hydrogen bonding network
- ☐ Assign charges to the ligand specified in a MOL2 file
- ☒ Create an APBS input file
- ☐ Add/keep chain IDs in the PQR file
- ☐ Insert whitespaces between atom name and residue name, between x and y, and between y and z
- ☐ Make the protein's N-terminus neutral (requires PARSE forcefield)
- ☐ Make the protein's C-terminus neutral (requires PARSE forcefield)
- ☒ Remove the waters from the output file

[Start Job](#)

B

1 PDB2PQR Configuration > 2 PDB2PQR Job Status > 3 APBS Configuration > 4 APBS Job Status > 5 Visualization

To return to your results after leaving, save this page. [🔗](#)

Job ID: **a8xkn4z4_20251209** Job Type: **PDB2PQR** Time Elapsed: **00:00:48** Next: [Use results with APBS >](#)

☒ Submitted ☒ Pending Job Start ☒ Running ☒ Complete

PDB2PQR Input Files

mpo_model.pdb	740.87 KB	Download
-------------------------------	-----------	--------------------------

PDB2PQR Output Files

pdb2pqr-metrics.json	621 Bytes	Download
a8xkn4z4.in	432 Bytes	Download
pdb2pqr.stderr.txt	61.28 KB	Download
pdb2pqr.stdout.txt	0 Bytes	Download
a8xkn4z4.pqr	640.13 KB	Download
a8xkn4z4.log	62.91 KB	Download

If you haven't already, please remember to register your use of this software:

[Register Here](#)

Log Preview

☐ Show line numbers

- > Log (a8xkn4z4.log) [Download](#)
- > Stdout (pdb2pqr.stdout.txt)
- > Stderr (pdb2pqr.stderr.txt) [Download](#)

Figura 6.1. Captura de pantalla de la interfaz del servidor *APBS*, sección *PDB2PQR*. (A) Interfaz inicial que muestra los parámetros ajustables para el cálculo de los valores de pK_a de los residuos ionizables en el modelo de la SCV2-Mpro. (B) Ventana de resultados que muestra la lista de archivos generados tras completar el cálculo de los valores de pK_a .

Se debe fijar el valor del pH a 7.3 y seleccionar *AMBER* como *campo de fuerzas*. El resto de los parámetros pueden ser los predeterminados. Una vez configurados los parámetros, hacer clic en **Start Job** para iniciar el procesamiento.

Al finalizar el cálculo, el servidor desplegará una página con los resultados, similar a la mostrada en la Figura 6.1B, desde donde se podrá descargar el archivo *pdb2pqr.stderr.txt*. Abrir este archivo y ubicar la sección **SUMMARY OF THIS PREDICTION** (ver Figura 6.2), donde se resumen los valores del pK_a predichos.

```
propka3.5.1
2025-12-09

-----
-- PROPKA: A PROTEIN PKA PREDICTOR
--
-- VERSION 1.0, 04/25/2004, IOWA CITY
-- BY HUI LI
--
-- VERSION 2.0, 11/05/2007, IOWA CITY/COPENHAGEN
-- BY DELPHINE C. BAS AND DAVID M. ROGERS
--

:
:

INFO: Coupling map:
INFO: Calculating pKas for Conformation container 1A with 5632 atoms and 1510 groups
INFO: pKa determinants:

-----
RESIDUE      pKa    BURIED    DESOLVATION  EFFECTS  SIDECHAIN  BACKBONE  COULOMBIC
          REGULAR    RE    HYDROGEN BOND  HYDROGEN BOND  INTERACTION
-----
ASP 33 A    3.89    5 %    0.49 296    0.00 0    -0.50 TYR 101 A    0.00 XXX 0 X    0.10 ASP 34 A
ASP 34 A    3.63    0 %    0.38 267    0.00 0    0.00 XXX 0 X    -0.26 ASP 34 A    -0.02 LYS 88 A
ASP 34 A    3.63    0 %    0.38 267    0.00 0    0.00 XXX 0 X    0.00 XXX 0 X    -0.28 LYS 90 A
ASP 48 A    2.51    0 %    0.48 271    0.00 0    -0.82 THR 45 A    -0.80 THR 45 A    -0.06 ARG 60 A
ASP 48 A    2.51    0 %    0.48 271    0.00 0    0.00 XXX 0 X    0.00 XXX 0 X    -0.10 LYS 61 A

:
:

SUMMARY OF THIS PREDICTION
Group      pKa    model-pKa    ligand atom-type
ASP 33 A    3.89    3.80
ASP 34 A    3.63    3.80
ASP 48 A    2.51    3.80
ASP 56 A    3.88    3.80
ASP 92 A    2.99    3.80
ASP 153 A    4.28    3.80
ASP 155 A    2.75    3.80
ASP 176 A    3.65    3.80
ASP 187 A    2.88    3.80
ASP 197 A    3.99    3.80
ASP 216 A    4.24    3.80
ASP 229 A    3.04    3.80
ASP 245 A    3.99    3.80
ASP 248 A    3.41    3.80
ASP 263 A    3.52    3.80
ASP 289 A    3.45    3.80
ASP 295 A    3.47    3.80
```

Figura 6.2. Captura de pantalla de la sección **SUMMARY OF THIS PREDICTION** del archivo *pdb2pqr.stderr.txt*, generado por el servidor *PDB2PQR*. Se muestran los resultados del cálculo aplicados a la estructura del modelo de la SCV2-Mpro, incluyendo los valores estimados de pK_a (columna *pKa*) para los grupos ionizables de residuos ácidos y básicos, así como los valores de referencia correspondientes al aminoácido libre (columna *model-pKa*). Esta información permite evaluar el efecto del entorno estructural sobre el comportamiento ácido-base de los residuos.

Deben revisarse los valores y comparar los pK_a estimados con el pH fisiológico de trabajo (pH 7.3), con el fin de identificar qué residuos podrían requerir un ajuste en su estado de protonación. Por ejemplo, si un residuo como GLU14 presenta un pK_a



estimado cercano a 7.5, esto sugiere que podría encontrarse protonado a pH 7.3, a diferencia de su forma desprotonada habitual en condiciones estándar.

Para validar estos ajustes, se recomienda utilizar UCSF Chimera para inspeccionar el entorno estructural del residuo, considerando factores como interacciones de puentes de hidrógeno, proximidad a otros grupos cargados o su exposición al solvente. Esta evaluación permitirá justificar si un cambio en el estado de protonación está sustentado por el contexto tridimensional del modelo.

Actividad 6.1. Discutir los resultados obtenidos a partir del análisis con *PropKa* para el modelo de la SCV2-Mpro.

Elaborar una lista de los residuos, así como de los grupos amino o carboxilo terminales, cuyos valores de pK_a predichos sugieran un cambio en su estado de protonación canónico.

Con el uso de UCSF Chimera, evaluar si existe alguna correlación entre la variación del valor de pK_a y el entorno de los grupos funcionales afectados.

Después de analizar los resultados obtenidos con *PropKa*, no se identificaron valores de pK_a anómalos ni predicciones que sugieran estados de protonación atípicos bajo las condiciones consideradas. En consecuencia, se procederá a asignar los estados de protonación canónicos (según pH fisiológico) y a generar la topología correspondiente utilizando GROMACS.

6.2 Preparación del sistema para la simulación dinámica molecular

La simulación se realizará utilizando GROMACS, un programa de código abierto ampliamente validado y empleado para estudios de simulación de DM de sistemas biomoleculares como proteínas, lípidos y ácidos nucleicos. En este ejercicio, se configurará un sistema que incorpore la SCV2-Mpro inmersa en una caja periódica solvatada con agua e iones, seguido de una simulación bajo condiciones fisiológicas controladas.

Para ello, se requiere tener GROMACS instalado en Linux o macOS, preferentemente en una versión igual o posterior a la serie 2018. La estructura de la proteína, previamente modelada mediante *AlphaFold2*, será simulada en medio acuoso a pH 7.3, 310 K, 1 bar de presión y 0.15 M de NaCl.

6.2.1 Conversión del formato *.pdb* a *.gmx* y generación de la topología

El primer paso en la preparación del sistema para simulación de DM consiste en generar la topología de la proteína a partir de su estructura tridimensional en formato PDB. Para ello se utiliza el comando `pdb2gmx` de GROMACS, que convierte el archivo *.pdb* en un archivo *.gro* (formato interno de GROMACS) y, además, genera los archivos de topología necesarios para la simulación.



Antes de ejecutar el comando, es importante verificar que el archivo *.pdb* contenga únicamente átomos pertenecientes a la proteína, ya que la presencia de otras entidades (ligandos, iones, moléculas de agua, entre otros) podría ocasionar errores durante el proceso de conversión.

Se recomienda crear una carpeta de trabajo específica para esta práctica y colocar en ella el archivo del modelo tridimensional de la SCV2-Mpro, renombrándolo como *mpro_model.pdb*. A continuación, se debe abrir una terminal dentro de dicho directorio de trabajo y ejecutar el siguiente comando:

```
$ gmx pdb2gmx -f mpro_model.pdb -o mpro_processed.gro -water  
spce -ignh
```

Este comando realiza las siguientes tareas:

- `-f mpro_model.pdb`: indica el archivo de entrada con la estructura de la proteína
- `-o mpro_processed.gro`: especifica el nombre del archivo de salida con las coordenadas en formato GROMACS
- `-water spce`: define el modelo de agua a utilizar; define el modelo de agua a utilizar; en este caso, SPC/E (Extended Simple Point Charge), un modelo ampliamente empleado y compatible con varios campos de fuerza.
- `-ignh`: indica a GROMACS que ignore los hidrógenos del archivo PDB, de modo que el programa los reconstruya automáticamente según el campo de fuerzas elegido.

Durante la ejecución del comando, GROMACS solicitará seleccionar un *campo de fuerzas* para generar la topología de la proteína. Este paso es fundamental, ya que el *campo de fuerzas* determina los parámetros fisicoquímicos que describen las interacciones entre los átomos del sistema (enlaces, ángulos, torciones, cargas, *van der Waals*, entre otros). Para este trabajo, se empleará el *campo de fuerzas* OPLS-AA/L all-atom, ampliamente validado para simulaciones de proteínas. Cuando aparezca la lista de opciones en la terminal, introducir el número **15** (correspondiente a OPLS-AA/L) y presionar **Return** para seleccionarlo.

Una vez ejecutado el comando, GROMACS generará en el directorio de trabajo seis archivos:

- I. *mpro_processed.gro*: archivo de coordenadas compatible con GROMACS, que contiene la estructura de la proteína con todos los átomos definidos de acuerdo con el campo de fuerzas seleccionado
- II. *posre_Protein_chain_A.itp* y *posre_Protein_chain_B.itp*: archivos de restricciones de posición aplicadas a los átomos pesados (no hidrógenos) de la cadena A y B, respectivamente. Se utiliza durante las etapas iniciales de minimización y equilibración

- III. *topol_Protein_chain_A.itp* y *topol_Protein_chain_B.itp*: archivos que contienen la topología detallada de la cadena A y B, respectivamente, incluyendo información sobre átomos, enlaces, ángulos, torsiones, y parámetros del *campo de fuerzas* (ver Figura 6.3)
- IV. *topol.top*: archivo principal de topología del sistema completo, que incluye referencias a las topologías individuales de cada cadena, las restricciones, el campo de fuerzas utilizado, y los parámetros globales (moléculas presentes, cantidades, entre otros, ver Figura 6.4)

Una vez completada esta etapa y generados los archivos descritos previamente, se podrá continuar con la siguiente fase de preparación del sistema, que incluye la definición de la caja de simulación, solvatación y neutralización iónica del sistema, como se describe a continuación.

```
;
;
; Created by:
;               :-) GROMACS - gmx pdb2gmx, 2025.3-Homebrew (-:
;
; Executable: /opt/homebrew/bin/./Cellar/gromacs/2025.3_1/bin/gmx
; Data prefix: /opt/homebrew/bin/./Cellar/gromacs/2025.3_1
; Command line:
; gmx pdb2gmx -f mpro_model.pdb -o mpro_processed.gro -water spce -ignh
; Force field was read from the standard GROMACS share directory.
;

[ moleculetype ]
; Name          nrexcl
Protein_chain_A 3

[ atoms ]
; nr      type  resnr residue atom  cgnr  charge      mass  typeB  chargeB  massB
; residue 1 SER  rtp SER  q +1.0
1  op1s_287  1  SER  N  1  -0.3  14.0027
2  op1s_290  1  SER  H1  1  0.33  1.008
3  op1s_290  1  SER  H2  1  0.33  1.008
4  op1s_290  1  SER  H3  1  0.33  1.008
5  op1s_293B 1  SER  CA  1  0.25  12.011
6  op1s_140  1  SER  HA  1  0.06  1.008
7  op1s_157  1  SER  CB  2  0.145 12.011
8  op1s_140  1  SER  HB1 2  0.06  1.008
9  op1s_140  1  SER  HB2 2  0.06  1.008
10 op1s_154  1  SER  OG  3 -0.683 15.9994
11 op1s_155  1  SER  HG  3  0.418  1.008
12 op1s_235  1  SER  C  4  0.5  12.011
13 op1s_236  1  SER  O  4 -0.5  15.9994 ; qtot 1
; residue 2 GLY rtp GLY  q 0.0
14 op1s_238  2  GLY  N  5 -0.5  14.0067
15 op1s_241  2  GLY  H  5  0.3  1.008
16 op1s_223B 2  GLY  CA  5  0.08 12.011
17 op1s_140  2  GLY  HA1 5  0.06  1.008
18 op1s_140  2  GLY  HA2 5  0.06  1.008
19 op1s_235  2  GLY  C  6  0.5  12.011
20 op1s_236  2  GLY  O  6 -0.5  15.9994 ; qtot 1
; residue 3 PHE rtp PHE  q 0.0
21 op1s_238  3  PHE  N  7 -0.5  14.0067
22 op1s_241  3  PHE  H  7  0.3  1.008
23 op1s_224B 3  PHE  CA  7  0.14 12.011
24 op1s_140  3  PHE  HA  7  0.06  1.008
25 op1s_149  3  PHE  CB  8 -0.005 12.011
26 op1s_140  3  PHE  HB1 8  0.06  1.008
27 op1s_140  3  PHE  HB2 8  0.06  1.008
28 op1s_145  3  PHE  CG  8 -0.115 12.011
29 op1s_145  3  PHE  CD1 9 -0.115 12.011
30 op1s_146  3  PHE  HD1 9  0.115  1.008
31 op1s_145  3  PHE  CD2 10 -0.115 12.011
32 op1s_146  3  PHE  HD2 10  0.115  1.008
;
;
```

Figura 6.3. Captura de pantalla de la visualización del archivo de topología correspondiente a la cadena A de la SCV2-Mpro (*topol_Protein_chain_A.itp*), generado mediante el módulo *pdb2gmx* de GROMACS. Este archivo contiene la parametrización detallada de la cadena polipeptídica de acuerdo con el campo de fuerzas OPLS-AA/L

all-atom. En particular, la sección [atoms] especifica, para cada átomo: su identificador, tipo atómico, residuo asociado, masa atómica, y carga parcial, información esencial para el cálculo de interacciones durante la simulación molecular.

```
; File 'topol.top' was generated
; By user: pgg (501)
; On host: ppgs-MacBook-Pro.local
; At date: Tue Dec 9 12:06:46 2025
;
; This is a standalone topology file
;
; Created by:
;               :-) GROMACS - gmxdump, 2025.3-Homebrew (-:
;
; Executable: /opt/homebrew/bin/./Cellar/gromacs/2025.3_1/bin/gmxdump
; Data prefix: /opt/homebrew/bin/./Cellar/gromacs/2025.3_1
; gmxdump -f mpro_model.pdb -o mpro_processed.gro -water spce -ignh
; Force field was read from the standard GROMACS share directory.
;

; Include forcefield parameters
#include "oplsaa.ff/forcefield.itp"

; Include chain topologies
#include "topol_Protein_chain_A.itp"
#include "topol_Protein_chain_B.itp"

; Include water topology
#include "oplsaa.ff/spce.itp"

#ifdef POSRES_WATER
; Position restraint for each water oxygen
[ position_restraints ]
; i funct fcx fcy fcz
; 1 1 1000 1000 1000
#endif

; Include topology for ions
#include "oplsaa.ff/ions.itp"

[ system ]
; Name
Protein

[ molecules ]
; Compound #mols
Protein_chain_A 1
Protein_chain_B 1
```

Figura 6.4. Visualización digital del archivo de topología general *topol.top*, generado con el módulo *gmxdump* de GROMACS. Este archivo integra los parámetros del campo de fuerzas seleccionado (OPLS-AA/L all-atom), las restricciones de posición sobre átomos pesados, la topología del modelo de agua TIP3P, así como las secciones [system] y [molecules], que especifican la composición completa del sistema molecular (proteína, solvente, iones) a simular con DM.

6.2.2 Construcción de la caja de simulación y solvatación

Una vez generada la topología del sistema, el siguiente paso consiste en definir la forma y volumen de la caja de simulación que contendrá a la proteína y, posteriormente, solvatarla con moléculas de agua. Para ello, se construye una caja de simetría dodecaédrica, centrando la proteína en su interior y estableciendo una distancia mínima de 2.0 nm entre la superficie molecular y las paredes de la caja. Este margen garantiza que la proteína pueda moverse libremente durante la dinámica sin interactuar consigo misma a través de las condiciones periódicas de contorno (PBC).

La caja se genera mediante el siguiente comando:

```
$ gmx editconf -f mpro_processed.gro -o mpro_box.gro -c -d 2.0 -bt dodecahedron
```

Donde:

- `-f mpro_processed.gro`: especifica el archivo de entrada (estructura procesada de la proteína)
- `-o mpro_box.gro`: define el archivo de salida con la caja ya definida
- `-c`: Centra automáticamente la proteína dentro de la caja
- `-d 2.0`: establece una distancia mínima de 2.0 nm entre la proteína y las paredes de la caja
- `-bt dodecahedron`: indica el uso de una caja dodecaédrica truncada, geoméricamente más eficiente que una caja cúbica, ya que requiere menos moléculas de solvente y, por tanto, reduce el costo computacional

El volumen y las dimensiones de la caja deberían ser similares a las mostradas en la Figura 6.5. Una vez construida la caja, se procede a su solvatación.

```
gmx editconf -f mpro_processed.gro -o mpro_box.gro -c -d 2.0 -bt dodecahedron
Note that major changes are planned in future for editconf, to improve usability and utility.
Read 9364 atoms
Volume: 398.365 nm^3, corresponds to roughly 179200 electrons
No velocities found
  system size : 6.712 8.148 7.284 (nm)
  diameter    : 8.471 (nm)
  center      : 0.065 -0.040 0.029 (nm)
  box vectors : 6.712 8.148 7.284 (nm)
  box angles  : 90.00 90.00 90.00 (degrees)
  box volume  : 398.36 (nm^3)
  shift       : 9.289 9.394 4.380 (nm)
  new center  : 9.354 9.354 4.409 (nm)
  new box vectors : 12.471 12.471 12.471 (nm)
  new box angles  : 60.00 60.00 90.00 (degrees)
  new box volume  : 1371.61 (nm^3)
GROMACS reminds you: "printf("%d is the year of the linux desktop", year+1);" (Anonymous)
```

Figura 6.5. Captura de pantalla de la Terminal donde se muestran las dimensiones de la caja de simulación construida con el módulo *editconf* para el sistema SCV2-Mpro.

6.2.3 Solvatación del proteína y llenado de la caja con moléculas de agua

Una vez definida la caja de simulación, el siguiente paso corresponde a la solvatación del sistema, es decir, llenar el volumen libre con moléculas de agua para simular un entorno biológico realista. En este ejercicio se empleará el modelo de agua spc216.gro, seleccionado de forma implícita al usar el modelo predeterminado SPC/E, que es compatible con el campo de fuerzas OPLS-AA/L.

La solvatación se realiza mediante el siguiente comando:

```
$ gmx solvate -cp mpro_box.gro -cs spc216.gro -o mpro_solvate.gro -p topol.top
```

Donde:

- **solvate:** calcula el número de moléculas de agua que deben adicionarse según el espacio vacío en la caja de simulación
- **-cp mpro_box.gro:** indica el archivo de entrada con la proteína centrada en la caja (*mpro_box.gro*)
- **-cs spc216.gro:** especifica el modelo de agua a utilizar (*spc216.gro*), incluido por defecto en GROMACS
- **-o mpro_solvate.gro:** define el archivo de salida que contendrá la proteína inmersa en un medio acuoso (*mpro_solvate.gro*)
- **-p topol.top:** señala el archivo de topología global del sistema (*topol.top*), el cual será actualizado automáticamente para reflejar el número de moléculas de agua añadidas

El programa calcula automáticamente el número adecuado de moléculas de agua requeridas para llenar el volumen libre de la caja, evitando cualquier superposición con los átomos de la proteína. Una vez finalizada la solvatación, se recomienda revisar el archivo *topol.top*, específicamente la sección [molecules], donde debe aparecer una línea correspondiente al solvente, generalmente indicada como SOL. En dicha línea se reporta la cantidad exacta de moléculas de agua que fueron añadidas durante el proceso (ver Figura 6.6). Esta información es fundamental para asegurar que la topología del sistema esté actualizada antes de continuar con las siguientes etapas de preparación para la simulación.

```
Generating solvent configuration
Will generate new solvent configuration of 7x7x5 boxes
Solvent box contains 149301 atoms in 49767 residues
Removed 15615 solvent atoms due to solvent-solvent overlap
Removed 8691 solvent atoms due to solute-solvent overlap
Sorting configuration
Found 1 molecule type:
  SOL ( 3 atoms): 41665 residues
Generated solvent containing 124995 atoms in 41665 residues
Writing generated configuration to mpro_solvate.gro

Output configuration contains 134359 atoms in 42277 residues
Volume      : 1371.61 (nm^3)
Density     : 996.122 (g/L)
Number of solvent molecules: 41665
Processing topology
Adding line for 41665 solvent molecules with resname (SOL) to topology file (topol.top)

Back Off! I just backed up topol.top to ./#topol.top.1#

GROMACS reminds you: "Nullis in verba [Nobody's word is final]." (Motto of the Royal Society)
```

Figura 6.6. Captura de pantalla de la Terminal donde se especifica el número de moléculas de agua adicionadas a la caja de simulación por GROMACS mediante el módulo *solvate*. Este número también queda registrado en la sección [molecules] del archivo *topol.top*. Esta información es esencial para verificar que la caja ha sido correctamente solvatada y que el sistema está preparado para las siguientes etapas de preparación, como la neutralización iónica y la minimización de energía.

Este procedimiento requiere la creación de un archivo de parámetros denominado *ions.mdp*, el cual debe contener las configuraciones mínimas necesarias para preparar



el sistema para la inserción de iones. La configuración recomendada se muestra en la Tabla 6.1. Se debe copiar el texto en el archivo correspondiente y guardarse en el directorio de trabajo.

Para proceder, se debe generar un archivo binario de entrada (*ions.tpr*) utilizando el siguiente comando:

```
$ gmx grompp -f ions.mdp -c mpro_solvate.gro -p topol.top -o ions.tpr
```

Donde:

- `-c mpro_solvate.gro`: indica la configuración estructural actual del sistema (la caja ya solvatada)
- `-o ions.tpr`: especifica el archivo de salida en formato TPR, que contiene la información del sistema preparada para el paso de neutralización e incorporación de iones.

Una vez generado el archivo *ions.tpr*, se procede a añadir los iones mediante el siguiente comando:

```
$ gmx genion -s ions.tpr -o mpro_solvate_ions.gro -p topol.top -neutral -conc 0.15
```

Donde:

- `-neutral`: instruye al programa a neutralizar la carga del sistema
- `-conc 0.15`: especifica la concentración deseada de NaCl (0.15 M)

Durante la ejecución del comando *genion*, el programa solicitará seleccionar el componente del sistema en el que se realizará el reemplazo. En este caso, se debe elegir el grupo **SOL (grupo 13)**, correspondiente a las moléculas de agua presentes en la caja de simulación. GROMACS seleccionará aleatoriamente algunas de estas moléculas y las sustituirá por iones sodio (Na^+) y cloruro (Cl^-).

Actividad 6.2. Reportar en la bitácora el número de iones sodio y de iones cloruro adicionados a la caja de simulación. Con base a esta información, determinar la carga eléctrica total de la proteína. Finalmente, considerando el volumen de la caja y el número de iones agregados al sistema, evaluar si la concentración teórica del NaCl corresponde a 0.15 M?

El archivo de salida, denominado *mpro_solvate_ions.gro*, contendrá la configuración final del sistema: la proteína embebida en un entorno acuoso con fuerza iónica adecuada, lista para proceder a la etapa de minimización de energía.



```

; ions.mdp - used as input into grompp to generate ions.tpr
; Parámetros: qué hacer, cuándo parar y qué guardar
integrator          = steep ; Algoritmo (steep = steepest descent
minimization)
emtol                = 1000.0 ; Detener cuando la fuerza máxima < 1000.0
kJ/mol/nm
emstep              = 0.01  ; Tamaño del paso de energía
nsteps              = 50000 ; Número máximo de pasos de minimización
; Parameters for neighbors and interactions
nstlist             = 10    ; Frecuencia para actualizar la lista de vecinos
cutoff-scheme       = Verlet
coulombtype         = cutoff ; Tratamiento de interacciones electrostáticas
rcoulomb            = 1.2   ; Corte electrostático (nm)
rvdw                = 1.2   ; Corte de Van der Waals (nm)
pbc                 = xyz   ; Condiciones periódicas de frontera

```

Tabla 6.1. Visualización del archivo *ions.mdp*, el cual define los parámetros mínimos necesarios para la generación del archivo binario de entrada *ions.tpr*, requerido durante la adición de iones al sistema de simulación. Este archivo establece condiciones básicas de preprocesamiento energético, como el tipo de simulación, el esquema de control de temperatura y la selección del integrador. Su contenido puede adaptarse según las necesidades del sistema y se basa en los ejemplos proporcionados en la documentación oficial de GROMACS (<https://manual.gromacs.org/>).

6.2.5 Minimización de la energía

Antes de iniciar la simulación de DM, es fundamental realizar una minimización de la energía. Esta etapa elimina contactos atómicos poco realistas y reduce las tensiones estructurales introducidas durante la preparación del sistema (modelado, solvatación y adición de iones). La minimización asegura que el sistema comience la simulación desde una conformación físicamente estable y energéticamente favorable.

Para ejecutar esta etapa, se requiere un archivo de parámetros llamado *minim.mdp*, el cual debe crearse y guardarse en el directorio de trabajo. Este archivo contiene los parámetros específicos del proceso de minimización, como el algoritmo empleado, el número máximo de pasos, los criterios de convergencia y el tipo de restricciones aplicadas. El contenido básico de *minim.mdp* puede basarse en configuraciones estándar ampliamente utilizadas en GROMACS. En la Tabla 6.2 se muestra un ejemplo representativo de dicho archivo, el cual puede adaptarse según las características del sistema y los requerimientos del usuario. Para los fines de esta práctica, puede emplearse tal cual.

El archivo binario de entrada para la minimización de la energía se genera utilizando siguiente comando:

```
$ gmx grompp -f minim.mdp -c mpro_solvate_ions.gro -p topol.top -o em.tpr
```

Posteriormente, se ejecuta la minimización mediante el comando:

```
$ gmx mdrun -v -deffnm em
```


Este procedimiento genera varios archivos, entre los que destacan:

- *em.gro*: estructura del sistema con coordenadas minimizadas
- *em.edr*: archivo binario que contiene la evolución de la energía durante la minimización
- *em.log*: registro completo del proceso de ejecución

La minimización se considera exitosa si la energía potencial alcanza un valor negativo y estable, lo cual indica que el sistema está adecuadamente relajado (ver Figura 6.7). Para confirmar este resultado, se recomienda utilizar el siguiente comando:

```
$ gmx energy -f em.edr -o potential.xvg
```

Este comando permite visualizar la evolución de la energía potencial y verificar que el sistema se ha relajado correctamente antes de continuar con las etapas de equilibración. Al realizar el análisis de la energía, se debe seleccionar la opción **Potential** (opción 10) cuando se presenten las opciones disponibles en la terminal. La gráfica generada (archivo *potential.xvg*) puede visualizarse mediante *XMGrace* o cualquier otro programa compatible con archivos *.xvg*.

```
$ xmgrace potential.xvg
```

```
; minim.mdp - used as input into grompp to generate em.tpr
integrator          = steep    ; Algorithm (steep = steepest descent minimization)
emtol               = 1000.0   ; Stop minimization when the maximum force < 1000.0
                        kJ/mol/nm
emstep              = 0.01     ; Energy step size
nsteps              = 50000    ; Maximum number of (minimization) steps to perform
; Parameters describing how to find the neighbors of each atom and how to calculate
the interactions
nstlist             = 10      ; Frequency to update the neighbor list and long
range forces
cutoff-scheme       = Verlet
ns_type             = grid    ; Method to determine neighbor list (simple, grid)
coulombtype         = PME     ; Treatment of long range electrostatic
interactions
rcoulomb            = 1.2     ; Short-range electrostatic cut-off
rvdw                = 1.2     ; Short-range Van der Waals cut-off
pbc                 = xyz     ; Periodic Boundary Conditions (yes/no)
```

Tabla 6.2. Visualización del archivo de parámetros para una minimización de la energía (*minim.mdp*). Este archivo especifica los valores necesarios para llevar a cabo el proceso de minimización, incluyendo el tipo de integrador (en este caso, *steep* para el método de descenso de energía más empinado), el número máximo de pasos de minimización, el tamaño del paso de integración, así como las condiciones de frontera periódicas y otros parámetros relevantes. Este tipo de archivo puede ser descargado y adaptado a partir de ejemplos disponibles en la documentación oficial de GROMACS (<https://manual.gromacs.org/>).

Se espera obtener una curva descendente que refleje la disminución progresiva de la energía potencial, hasta alcanzar un mínimo estable. Este comportamiento indica que el sistema ha alcanzado un estado energéticamente favorable, condición necesaria

para continuar con las etapas posteriores. Un ejemplo representativo de esta gráfica se muestra en la Figura 6.8.

```

Step= 2247, Dmax= 8.0e-03 nm, Epot= -2.43768e+06 Fmax= 6.75070e+03, atom= 7952
Step= 2248, Dmax= 9.6e-03 nm, Epot= -2.43769e+06 Fmax= 8.01514e+03, atom= 7952
Step= 2249, Dmax= 1.1e-02 nm, Epot= -2.43770e+06 Fmax= 9.64699e+03, atom= 7952
Step= 2251, Dmax= 6.9e-03 nm, Epot= -2.43776e+06 Fmax= 9.76877e+02, atom= 7952
writing lowest energy coordinates
Steepest Descents converged to Fmax < 1000 in 2252 steps
Potential Energy = -2.437755e+06
Maximum force = 9.768773e+02 on atom 7952
Norm of force = 1.1122015e+01
GROMACS reminds you: "It's Not Dark Yet, But It's Getting There" (Bob Dylan)

```

Figura 6.7. Captura de pantalla del mensaje de salida generado por GROMACS tras finalizar la etapa de minimización de la energía del sistema que contiene a la SCV2-Mpro. El mensaje resume parámetros clave del cálculo, incluyendo el número total de pasos ejecutados, el criterio de convergencia alcanzado y el valor final de la energía potencial. Esta información permite evaluar si el sistema ha alcanzado un estado energéticamente estable, condición indispensable para continuar con la equilibración.

6.2.6 Estabilización del sistema

Una vez completada la minimización de la energía, el siguiente paso consiste en equilibrar el sistema bajo condiciones termodinámicas controladas. Esta etapa es crucial para asegurar que la proteína y el entorno solvente alcancen un estado estable en cuanto a temperatura, presión y densidad, previo al inicio de la simulación de producción.

El proceso de estabilización consta de dos fases sucesivas: estabilización NVT (número de partículas, volumen y temperatura constantes) y estabilización NPT (número de partículas, presión y temperatura constantes).

En la estabilización NVT se mantiene fijo el volumen de la caja de simulación mientras se ajusta la temperatura del sistema a 310 K. El objetivo es estabilizar las velocidades moleculares de acuerdo con una distribución de Maxwell-Boltzmann, en un entorno térmicamente definido. Esta fase permite adaptar al sistema a la temperatura objetivo sin afectar su densidad.

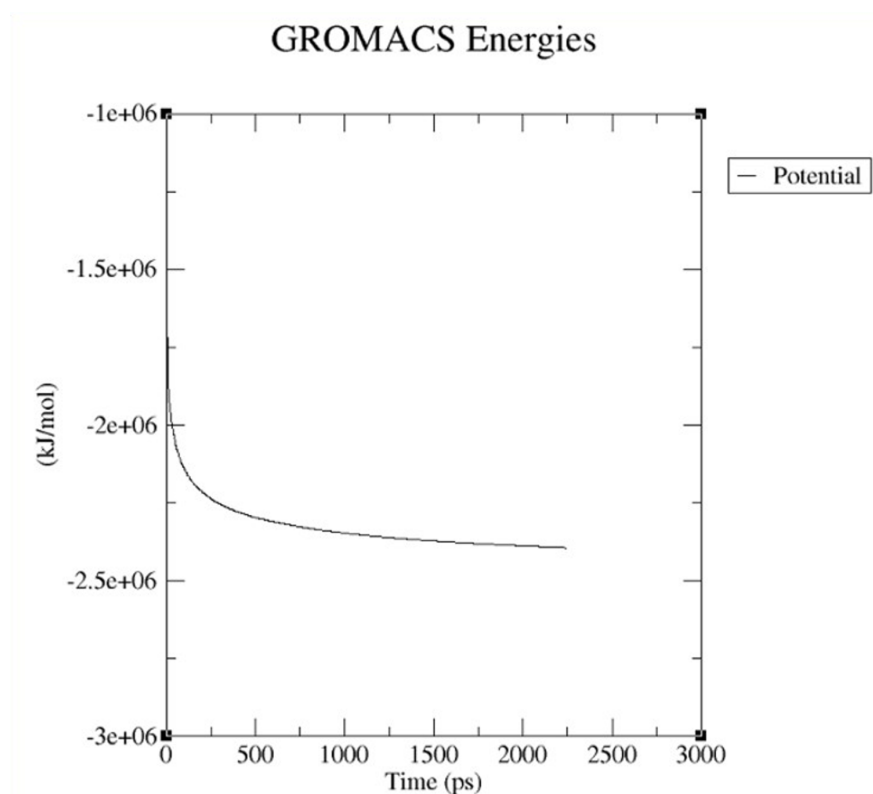


Figura 6.8. Evolución de la energía potencial durante la minimización del sistema que contiene la SCV2-Mpro, obtenida mediante GROMACS y visualizada con XMGrace. La gráfica muestra una disminución progresiva de la energía potencial conforme avanza el número de pasos de minimización, lo cual indica que el sistema alcanza un estado energéticamente más estable. Esta tendencia descendente valida la correcta relajación del sistema antes de iniciar las etapas de equilibración.

Para llevar a cabo esta etapa, es necesario crear el archivo de parámetros *nvt.mdp* con la configuración apropiada. Este archivo debe contener las opciones relacionadas con el control de temperatura, el integrador (tipo de algoritmo utilizado para actualizar las posiciones y velocidades de las partículas durante la simulación), los pasos de simulación (son el número total de iteraciones que GROMACS ejecutará para integrar las ecuaciones de movimiento y simular el comportamiento temporal del sistema), por mencionar algunas. Un ejemplo de configuración puede consultarse en la Tabla 6.3, la cual debe copiarse para crear el archivo correspondiente y guardarse en el directorio de trabajo con el nombre indicado.

El archivo binario de entrada binario para la fase de estabilización NVT se genera con el siguiente comando:

```
$ gmx grompp -f nvt.mdp -c em.gro -r em.gro -p topol.top -o nvt.tpr
```

A continuación, se ejecuta la simulación mediante:

```
$ gmx mdrun -v -deffnm nvt
```

```

title                = OPLS Mpro NVT equilibration
define               = -DPOSRES    ; position restrain the protein
; Run parameters
integrator           = md          ; leap-frog integrator
nsteps               = 50000       ; 2 * 50000 = 100 ps
dt                  = 0.002       ; 2 fs
; Output control
nstxout              = 500         ; save coordinates every 1.0 ps
nstvout              = 500         ; save velocities every 1.0 ps
nstenergy            = 500         ; save energies every 1.0 ps
nstlog               = 500         ; update log file every 1.0 ps
; Bond parameters
continuation         = no          ; first dynamics run
constraint_algorithm = lincs        ; holonomic constraints
constraints           = h-bonds     ; bonds involving H are constrained
lincs_iter           = 1           ; accuracy of LINCS
lincs_order          = 4           ; also related to accuracy
; Neighborsearching
cutoff-scheme        = Verlet
ns_type              = grid        ; search neighboring grid cells
nstlist              = 10          ; 20 fs, largely irrelevant with Verlet
rcoulomb              = 1.2         ; short-range electrostatic cutoff (in nm)
rvdw                  = 1.2         ; short-range van der Waals cutoff (in nm)
DispCorr              = EnerPres   ; account for cut-off vdW scheme
; Electrostatics
coulombtype           = PME         ; Particle Mesh Ewald for long-range electrostatics
pme_order             = 4           ; cubic interpolation
fourierspacing        = 0.16       ; grid spacing for FFT
; Temperature coupling is on
tcoupl                = V-rescale   ; modified Berendsen thermostat
tc-grps               = Protein Non-Protein ; two coupling groups-more accurate
tau_t                 = 0.1 0.1    ; time constant, in ps
ref_t                 = 310 310     ; reference temperature, one for each group, in K
; Pressure coupling is off
pcoupl                = no          ; no pressure coupling in NVT
; Periodic boundary conditions
pbc                   = xyz         ; 3-D PBC
; Velocity generation
gen_vel               = yes         ; assign velocities from Maxwell distribution
gen_temp              = 310         ; temperature for Maxwell distribution
gen_seed              = -1          ; generate a random seed

```

Tabla 6.3. Visualización del archivo *nvt.mdp* con los parámetros para el equilibrio NVT, empleado para fijar la temperatura del sistema de simulación en aproximadamente 310 K. Este archivo está disponible públicamente de los manuales de GROMACS (<https://manual.gromacs.org/>).

Para evaluar si el sistema ha alcanzado una temperatura estable cercana a 310 K, se debe analizar el archivo de energía generado durante la simulación con el siguiente comando:

```
$ gmx energy -f nvt.edr -o temperatura.svg
```

En el menú interactivo, se debe seleccionar la opción **Temperature**. La gráfica resultante, que puede visualizarse mediante *XMGrace*, debe mostrar una oscilación controlada alrededor del valor objetivo de temperatura. Un comportamiento estable,

como el ilustrado en la Figura 6.9, indica que el sistema está térmicamente equilibrado y permite avanzar a la siguiente fase de equilibración.

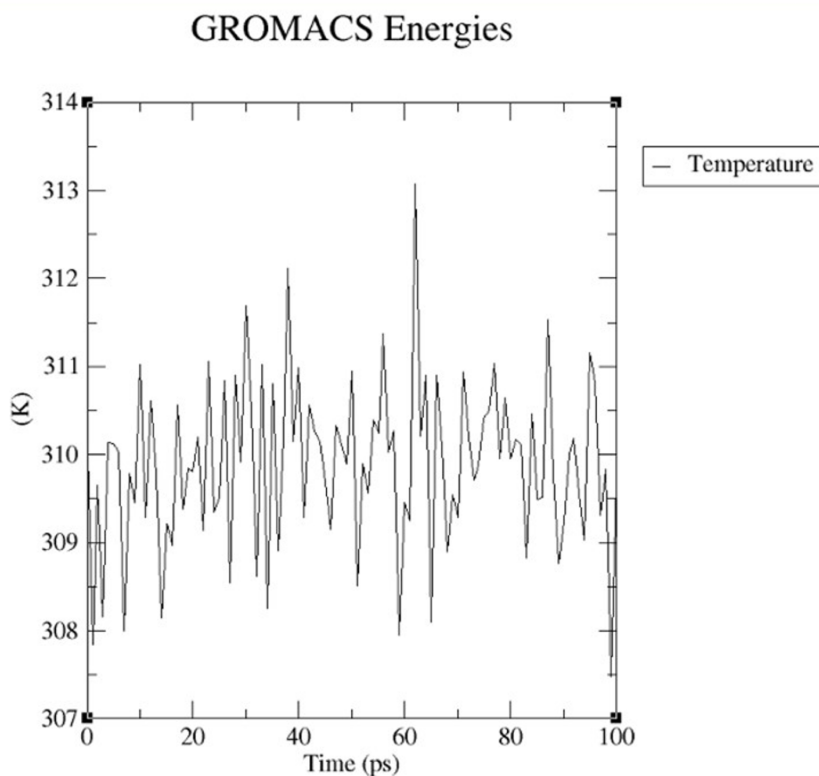


Figura 6.9. Gráfica de la evolución de la temperatura durante la fase de equilibrio NVT, obtenida mediante el módulo *energy* de GROMACS y visualizada con *XMGrace*. Se observa la estabilización de la temperatura en torno a 310 K, lo cual indica un adecuado control térmico del sistema y una distribución cinética conforme a la temperatura objetivo.

La segunda fase de equilibración, conocida como NPT, tiene como objetivo estabilizar la presión del sistema a 1 bar, manteniendo constante la temperatura alcanzada previamente en la fase NVT. Durante esta etapa se permite la variación del volumen de la caja de simulación, lo que posibilita que el sistema alcance una densidad adecuada. Esta fase es crucial para garantizar un entorno de simulación equilibrado antes de iniciar la etapa de producción.

Para ello, se debe generar un archivo de parámetros llamado *npt.mdp*, el cual debe contener la configuración mostrada en la Tabla 6.4 y guardarse en el mismo directorio de trabajo.

El archivo binario *.tpr* necesario para esta fase se genera con el siguiente comando:

```
$ gmx grompp -f npt.mdp -c nvt.gro -r nvt.gro -t nvt.cpt -p  
topol.top -o npt.tpr
```




```

title                = OPLS Mpro NPT equilibration
define               = -DPOSRES    ; position restrain the protein
; Run parameters
integrator           = md          ; leap-frog integrator
nsteps               = 50000       ; 2 * 50000 = 100 ps
dt                  = 0.002       ; 2 fs
; Output control
nstxout              = 500         ; save coordinates every 1.0 ps
nstvout              = 500         ; save velocities every 1.0 ps
nstenergy            = 500         ; save energies every 1.0 ps
nstlog               = 500         ; update log file every 1.0 ps
; Bond parameters
continuation         = yes         ; Restarting after NVT
constraint_algorithm = lincs       ; holonomic constraints
constraints           = h-bonds    ; bonds involving H are constrained
lincs_iter           = 1           ; accuracy of LINCS
lincs_order          = 4           ; also related to accuracy
; Neighborsearching
cutoff-scheme        = Verlet
ns_type              = grid        ; search neighboring grid cells
nstlist              = 10         ; 20 fs, largely irrelevant with Verlet
rcoulomb             = 1.2         ; short-range electrostatic cutoff (in nm)
rvdw                 = 1.2         ; short-range van der Waals cutoff (in nm)
DispCorr             = EnerPres    ; account for cut-off vdW scheme
; Electrostatics
coulombtype          = PME         ; Particle Mesh Ewald for long-range electrostatics
pme_order            = 4           ; cubic interpolation
fourierspacing       = 0.16       ; grid spacing for FFT
; Temperature coupling is on
tcoupl               = V-rescale   ; modified Berendsen thermostat
tc-grps              = Protein Non-Protein ; two coupling groups-more accurate
tau_t                = 0.1 0.1    ; time constant, in ps
ref_t                = 310 310    ; reference temperature, one for each group, in
K
; Pressure coupling is on
pcoupl               = Parrinello-Rahman ; Pressure coupling in NPT
pcoupltype           = isotropic    ; uniform scaling of box vectors
tau_p                = 2.0         ; time constant, in ps
ref_p                = 1.0         ; reference pressure, in bar
compressibility       = 4.5e-5     ; isothermal compressibility of water, bar^-1
refcoord_scaling      = com
; Periodic boundary conditions
pbc                  = xyz         ; 3-D PBC
; Velocity generation
gen_vel              = no         ; Velocity generation is off

```

Tabla 6.4. Archivo de parámetros *npt.mdp* utilizado para la fase de estabilización del sistema bajo condiciones de presión, temperatura y número de partículas constantes (ensamblaje NPT). Este archivo define los parámetros necesarios para estabilizar la presión a 1 bar, mientras se mantiene la temperatura previamente establecida durante la etapa NVT.

La simulación se ejecuta mediante:

```
$ gmx mdrun -v -deffnm npt
```

Para evaluar si la presión del sistema se ha estabilizado correctamente, se analiza el archivo de energía con:

```
$ gmx energy -f npt.edr -o pressure.xvg
```

Cuando se despliegue el menú interactivo, se debe seleccionar **Pressure** (opción 18). La gráfica resultante, visualizada con *XMGrace*, debe mostrar oscilaciones en torno a 1 bar, como se muestra en la Figura 6.10.

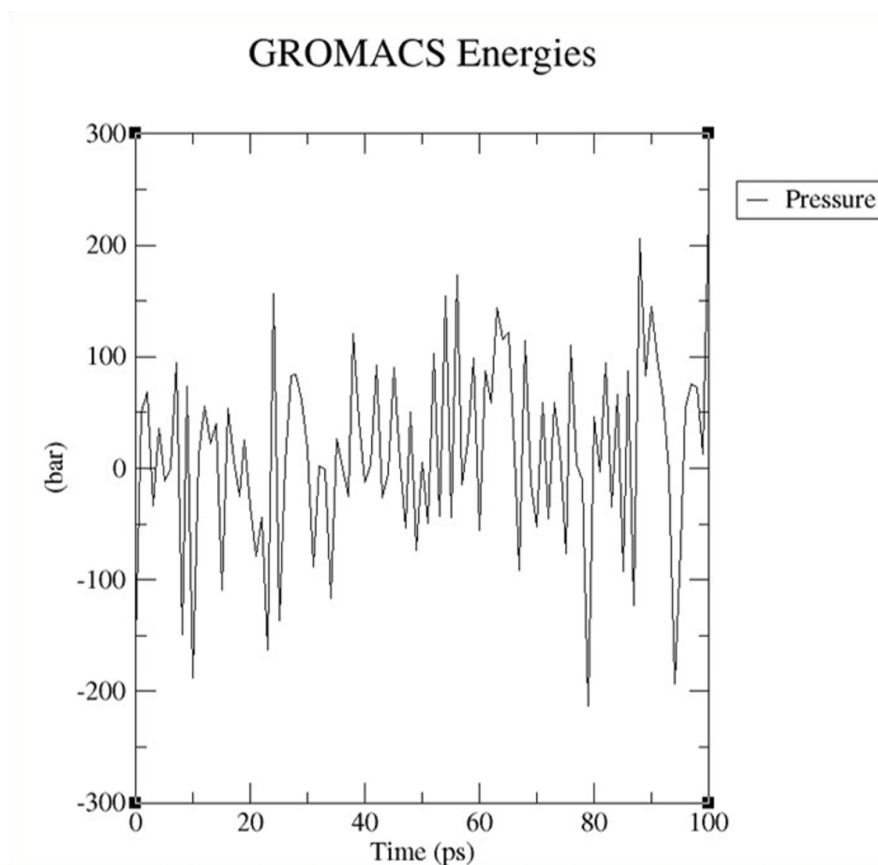


Figura 6.10. Evolución de la presión durante la fase de estabilización NPT, obtenida con el módulo *energy* de GROMACS y visualizada mediante *XMGrace*. La gráfica muestra oscilaciones en torno a 1 bar, en concordancia con las condiciones establecidas, lo que indica una adecuada estabilización del sistema bajo control barostático.

Estas variaciones son esperadas en simulaciones de DM y pueden dificultar una evaluación directa de su correspondencia con el valor de referencia de 1 bar. Para complementar esta verificación, se recomienda realizar también el análisis de la densidad del sistema, cuya estabilidad representa un indicador indirecto, pero robusto, del equilibrio termodinámico. La Figura 6.11 muestra un ejemplo de este análisis. La gráfica correspondiente se genera mediante la ejecución del siguiente comando:

```
$ gmx energy -f npt.edr -o density.xvg
```

En el menú interactivo, se debe seleccionar **Density** (opción 24). Posteriormente, la gráfica puede visualizarse mediante:

\$ xmgrace density.xvg

Si la densidad se mantiene estable a lo largo del tiempo, puede concluirse que el sistema ha alcanzado condiciones adecuadas de equilibrio termodinámico; en ese caso, puede iniciarse la simulación de producción.

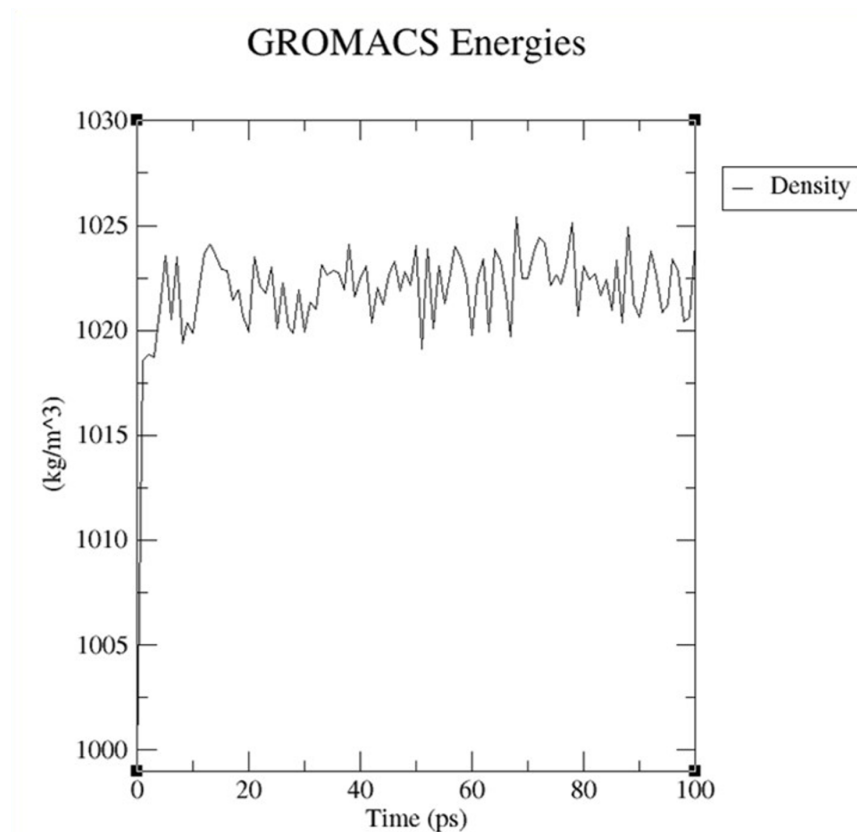


Figura 6.11. Gráfica de la evolución de la densidad del sistema durante la fase de equilibrio NPT, generada con el módulo energy de GROMACS y visualizada mediante *XMGrace*. Se observa que la densidad se mantiene próxima al valor esperado para el agua (alrededor de 1000 kg/m³), lo que indica una adecuada estabilidad volumétrica y respalda la validez de las condiciones simuladas.

6.2.7 Ejecución de la simulación de dinámica molecular

La simulación de producción constituye la etapa final del protocolo y tiene como objetivo analizar el comportamiento dinámico de la proteína en un entorno acuoso, bajo condiciones fisiológicas controladas. A diferencia de las fases de minimización y equilibración, esta etapa está orientada a generar datos estructurales (la conformación y organización tridimensional) y dinámicos (cómo cambia el sistema en el tiempo) representativos del sistema en equilibrio.

Para llevar a cabo esta simulación, se requiere un archivo de parámetros denominado *md.mdp*, el cual debe especificar las condiciones particulares para la integración

temporal, los esquemas de acoplamiento térmico y barostático, la duración total de la simulación, así como otros parámetros relevantes. Este archivo debe elaborarse conforme a la configuración mostrada en la Tabla 6.5 y guardarse en el directorio de trabajo.

```

title                = OPLS Mpro NPT equilibration
; Run parameters
integrator           = md                ; leap-frog integrator
nsteps              = 500000             ; 2 * 500000 = 1000 ps (1 ns)
dt                 = 0.002               ; 2 fs
; Output control
nstxout             = 0                  ; save coordinates every 0.0 ps
nstvout             = 0                  ; save velocities every 0.0 ps
nstfout             = 0                  ; nstfout, and nstfout
nstenergy           = 5000               ; save energies every 1000.0 ps
nstlog              = 5000               ; update log file every 1000.0 ps
nstxout-compressed   = 5000             ; save compressed coordinates every 1000.0
ps
compressed-x-grps    = System            ; save the whole system

; Bond parameters
continuation         = yes               ; Restarting after NPT
constraint_algorithm = lincs             ; holonomic constraints
constraints          = h-bonds           ; bonds involving H are constrained
lincs_iter           = 1                 ; accuracy of LINCS
lincs_order          = 4                 ; also related to accuracy
; Neighborsearching
cutoff-scheme        = Verlet
ns_type              = grid              ; search neighboring grid cells
nstlist              = 20                ; 20 fs, largely irrelevant with Verlet
scheme
rlist                = 1.2               ; short-range neighborlist cutoff (in nm)
rcoulomb              = 1.2              ; short-range electrostatic cutoff (in nm)
rvdw                  = 1.2              ; short-range van der Waals cutoff (in nm)
; Electrostatics
coulombtype           = PME              ; Particle Mesh Ewald for long-range
electrostatics
pme_order             = 4                 ; cubic interpolation
fourierspacing        = 0.16             ; grid spacing for FFT
; Temperature coupling is on
tcoupl               = V-rescale          ; modified Berendsen thermostat
tc-grps               = Protein Non-Protein ; two coupling groups - more
accurate
tau_t                 = 0.1 0.1          ; time constant, in ps
ref_t                 = 310 310          ; reference temperature, one for each group,
in K
; Pressure coupling is on
pcoupl               = Parrinello-Rahman ; Pressure coupling on in NPT
pcoupltype            = isotropic         ; uniform scaling of box vectors
tau_p                 = 2.0               ; time constant, in ps
ref_p                 = 1.0               ; reference pressure, in bar
compressibility        = 4.5e-5           ; isothermal compressibility of water, bar^-1
1
; Periodic boundary conditions
pbc                   = xyz               ; 3-D PBC
; Dispersion correction
DispCorr              = EnerPres          ; account for cut-off vdW scheme
; Velocity generation
gen_vel               = no               ; Velocity generation is off

```

Tabla 6.5. Ejemplo del archivo *md.mdp*, el cual contiene los parámetros requeridos para ejecutar la simulación de DM de producción. Este archivo define las condiciones de integración temporal, control de temperatura y presión, duración total de la simulación, así como otros aspectos fundamentales del protocolo computacional.



El archivo binario de entrada (*md.tpr*) se genera con el siguiente comando:

```
$ gmx grompp -f md.mdp -c npt.gro -t npt.cpt -p topol.top -o md.tpr
```

Donde:

- -f md.mdp: archivo de parámetros de simulación de producción
- -c npt.gro: estructura final obtenida tras la fase de estabilización NPT
- -t npt.cpt: archivo de punto de control que conserva velocidades y parámetros del equilibrio
- -p topol.top: archivo de topología del sistema
- -o md.tpr: archivo binario de salida listo para ejecutar la simulación de producción

La simulación se ejecuta mediante el siguiente comando:

```
$ gmx mdrun -v -deffnm md
```

Donde:

- -v: activa la visualización detallada del progreso de la simulación en tiempo real
- -deffnm 1ns: define el prefijo para los archivos generados, asociados a una simulación de 1 nanosegundo, aunque dicho valor puede ajustarse según los objetivos experimentales

Debe considerarse que esta fase puede representar una demanda computacional significativa. El tiempo de ejecución dependerá tanto de la duración configurada para la simulación como de las capacidades del hardware disponible (particularmente el uso de CPU y/o GPU).

Una vez finalizada la simulación de producción, se generarán los siguientes archivos clave:

- md.gro: contiene las coordenadas finales del sistema tras la simulación
- md.xtc: archivo comprimido que almacena la trayectoria completa del sistema (coordenadas a lo largo del tiempo)
- md.edr: archivo con los datos energéticos registrados durante la simulación
- md.log: bitácora detallada del proceso

6.3 Análisis posterior a la simulación

Entre los análisis más comunes se incluyen:

- I. RMSD (*Root Mean Square Deviation*): evalúa la desviación cuadrática media de la estructura proteica a lo largo del tiempo con respecto a una conformación de

- 
- referencia (usualmente el primer frame). Es un indicador global de estabilidad estructural
- II. RMSF (*Root Mean Square Fluctuation*): mide la flexibilidad de cada residuo, lo que permite identificar regiones móviles o rígidas de la proteína
 - III. R_g (*Radius of gyration*): informa sobre la compactación estructural del sistema durante la simulación
 - IV. E_{pot} (*Potential Energy*) y temperatura: estos parámetros permiten complementar la evaluación del equilibrio termodinámico del sistema. La energía potencial debe mantenerse en un valor estable y negativo, y la temperatura debe oscilar alrededor del valor objetivo (310 K)
 - V. Análisis de trayectorias y contactos moleculares: permite estudiar interacciones específicas, como enlaces de hidrógeno, puentes salinos, contactos hidrofóbicos o interacciones proteína–ligando. También puede emplearse para observar transiciones conformacionales y movimientos colectivos relevantes

Cada uno de estos análisis puede llevarse a cabo mediante módulos específicos de GROMACS, y sus resultados pueden ser visualizados utilizando herramientas especializadas como *XMGrace*, VMD o PyMOL, dependiendo del tipo de dato generado y el propósito del análisis.

En un contexto de investigación aplicada, los resultados derivados de estos análisis pueden ser fundamentales para:

- Proponer estrategias de inhibición específicas,
- Realizar una validación estructural pos-modelado
- bien, contribuir al Diseño Racional de Fármacos, mediante la identificación de residuos clave involucrados en la función o interacción molecular

Bibliografía

- Lemkul, J. A.** (2018). *From Proteins to Perturbed Hamiltonians: A Suite of Tutorials for the GROMACS-2018 Molecular Simulation Package*, v1.0 Living Journal of Computational Molecular Science 1(1): 5068.
- Hunkler, S.,** Diederichs, K., Kukharensko, O., & Peter, C. (2023). *Fast conformational clustering of extensive molecular dynamics simulation data*. The Journal of Chemical Physics, 158(14):14410
- Tubiana, T.,** Carvaille, J. C., Boulard, Y., & Bressanelli, S. (2018). *TTClust: a versatile molecular simulation trajectory clustering program with graphical summaries*. Journal of chemical information and modeling, 58(11): 2178-2182.
- Karplus, M.,** & Kuriyan, J. (2005). *Molecular dynamics and protein function*. Proceedings of the National Academy of Sciences, 102(19): 6679-6685.



Garduño-Juárez, R., Tovar Anaya, D. O., Pérez-Aguilar, J. M., Lozano-Aguirre Beltrán, L. F., Zubillaga, R. A. Álvarez Pérez, M. A. & Villarreal-Ramírez E. (2024). *Molecular dynamic simulations for biopolymers with biomedical applications*. Polymers 16, 1864.



Práctica 7. Análisis estructural de la trayectoria de simulación de dinámica molecular de una proteína

Objetivos

- Procesar la trayectoria de simulación de dinámica molecular, corrigiendo efectos de frontera periódica y alineando la proteína para su análisis estructural.
- Cuantificar e interpretar la estabilidad y la dinámica conformacional mediante métricas estructurales tales como RMSD, RMSF, radio de giro, enlaces de hidrógeno.
- Identificar y caracterizar estados conformacionales mediante *clustering* estructural, seleccionando conformaciones representativas para cribado virtual.

Software especializado requerido

- GROMACS versión 2024.0 o posterior
- XMGrace versión 5.1 o posterior
- VMD versión 1.8 o posterior
- TTKlust versión 10.3 o posterior

Una vez finalizada la simulación de DM, el siguiente paso corresponde al procesamiento de la trayectoria generada para facilitar su análisis estructural. Dicho procesamiento incluye la eliminación de discontinuidades causadas por las condiciones de frontera periódicas –una estrategia computacional que simula un sistema infinito replicando la caja de simulación en todas direcciones– y la alineación de la proteína en el espacio para remover traslaciones y rotaciones globales.

Para comenzar, debe eliminarse el efecto de *saltos* en la trayectoria generados cuando átomos o cadenas cruzan los límites de la caja periódica. Esto se logra mediante el siguiente comando:

```
$ gmx trjconv -f md.xtc -s md.tpr -o md_NJMP.xtc -pbc nojump -ur compact
```

Donde:

- `-pbc nojump`: corrige los saltos causados por la caja periódica
- `-ur compact`: genera una representación más eficiente del sistema

Al ejecutar el comando, se solicitará seleccionar un grupo de átomos. Debe seleccionarse **System** (grupo 0).

Este paso no altera la geometría interna de la proteína ni sus interacciones, pero permite representar de manera continua su movimiento a lo largo de la simulación. El archivo resultante (*md_NJMP.xtc*) se utilizará en las siguientes etapas de análisis estructural.

El siguiente comando permite ajustar la trayectoria simulada para centrar y superponer la proteína en cada fotograma, eliminando así los desplazamientos globales (traslaciones y rotaciones) que podrían interferir con el análisis estructural:



```
$ gmx trjconv -s md.tpr -f md_NJMP.xtc -o md_NJMP_FIT.xtc -fit  
rot+trans
```

Durante la ejecución, GROMACS solicitará dos selecciones sucesivas:

- I. Para definir el grupo de referencia: ingresar **4 (C-alpha)** (corresponde a los átomos C α de la proteína, utilizados comúnmente para alinear la estructura)
- II. Para el grupo a transformar: ingresar **0 (System)** (se alineará todo el sistema con base en la referencia)

Una vez generada la trayectoria alineada (*md_NJMP_FIT.xtc*), puede visualizarse la simulación de DM de la SCV2-Mpro utilizando VMD- Visual Molecular Dynamics (VMD).

Abrir VMD y en **VMD Main**, ir a **File>New Molecule...** En la ventana **Molecule File Browser**, hacer clic **Browse**, seleccionar el archivo *md.gro* ubicado en el directorio de trabajo y luego hacer clic en **Load**. Posteriormente, definir la representación visual de la proteína: en **VMD Main**, acceder a **Graphics>Representations...** En la pestaña **Selected Atoms**, eliminar la selección por defecto y escribir **protein**.

Esto mostrará exclusivamente la proteína, con una visualización estática basada en la estructura final de la simulación (ver Figura 7.1).

A continuación, cargar la trayectoria: en **VMD Main**, hacer clic derecho sobre la molécula cargada (identificada como Mpro) y seleccionar **Load Data Into Molecule...**

En la nueva ventana, seleccionar el archivo *md_NJMP_FIT.xtc*. Hacer clic en **Load**.

Una vez cargada la trayectoria, la ventana OpenGL Display mostrará la dinámica de la SCV2-Mpro a lo largo del tiempo. Es posible observar cómo cambia su conformación y movilidad durante la simulación. La reproducción de la trayectoria puede controlarse mediante los botones ubicados en la parte inferior de la ventana **VMD Main** (play, stop, rewind, entre otros), facilitando la identificación de eventos dinámicos relevantes, como movimientos de dominios, apertura de cavidades, o reorientación de cadenas laterales.

Sobre la trayectoria, pueden aplicarse diversos análisis para evaluar la evolución estructural de la proteína durante la simulación. Uno de los análisis más comunes es el cálculo del RMSD, el cual mide la desviación promedio de las posiciones atómicas respecto a una estructura de referencia (generalmente la conformación inicial). En el contexto de simulaciones, este análisis permite evaluar la estabilidad global del sistema y detectar posibles transiciones conformacionales significativas.

Para calcular el RMSD, utilizar el siguiente comando:

```
$ gmx rms -s em.tpr -f md_NJMP_FIT.xtc -o rmsd.svg -tu ns
```

Donde:

- -tu ns: ajusta las unidades de tiempo a nanosegundos

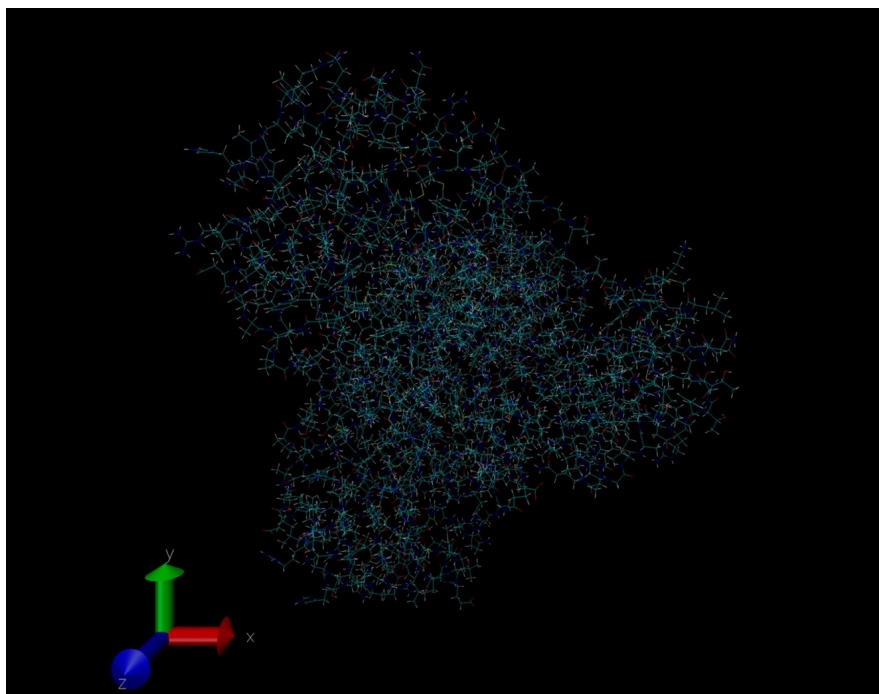


Figura 7.1. Representación en varillas de la SCV2-Mpro visualizada con el programa VMD. La imagen corresponde a la estructura cargada desde el archivo *md.gro*, utilizando el criterio *protein* en el campo **Selected Atoms** del menú **Graphical Representations**. Esta visualización muestra la conformación tridimensional de la proteína en su estado final tras la simulación, antes de reproducir la trayectoria dinámica.

Cuando se solicite la selección de grupos para el análisis, seleccione **4:4 (C-alpha: C-alpha)**, lo cual permite calcular la desviación estructural considerando exclusivamente los átomos $C\alpha$, que representan el esqueleto de la cadena polipeptídica y son apropiados para evaluar la estabilidad estructural general.

Para visualizar la gráfica resultante, utilizar el siguiente comando:

```
$ xmgrace rmsd.svg
```

La gráfica obtenida muestra la evolución del RMSD en función del tiempo, permitiendo identificar regiones de estabilidad estructural o fluctuaciones conformacionales. Una curva que asciende rápidamente y luego se estabiliza en una meseta indica que el sistema ha alcanzado un estado estable tras un periodo inicial de relajación. Por el contrario, una curva con variaciones amplias o tendencias crecientes podría reflejar una transición conformacional, inestabilidad o artefactos en la simulación. Un resultado característico se muestra en la Figura 7.2.

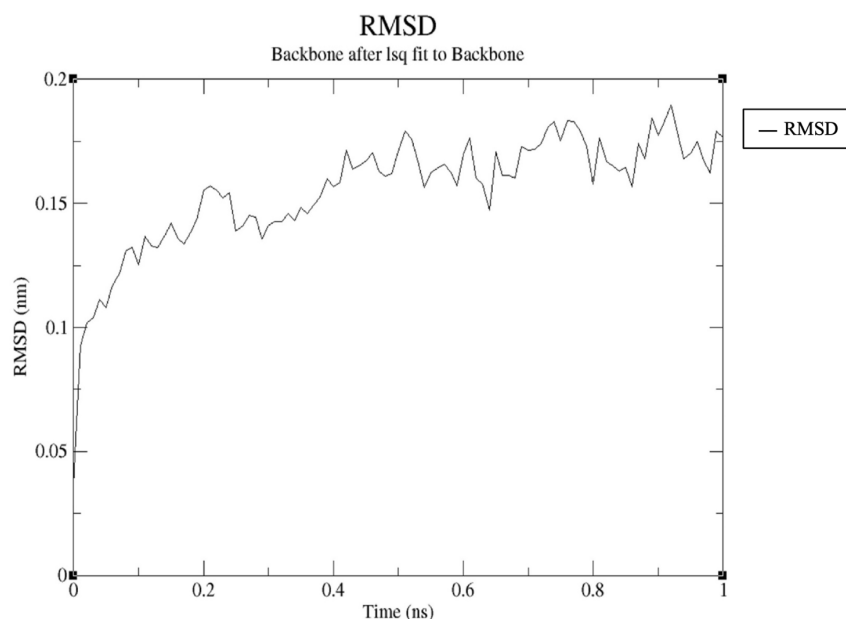


Figura 7.2. Gráfica de la evolución del RMSD a lo largo del tiempo para la SCV2-Mpro, obtenida con el módulo *rms* de GROMACS y visualizada mediante *XMGrace*. El eje Y muestra la desviación estructural promedio en nanómetros (nm) respecto a la estructura de referencia, mientras que el eje X indica el tiempo de simulación en nanosegundos (ns). Esta gráfica permite evaluar la estabilidad conformacional de la proteína durante la simulación de DM.

La estructura de referencia utilizada para calcular el RMSD de cada uno de los *frames* –configuración atómica registrada en un tiempo determinado– de la simulación corresponde a la obtenida tras la etapa de minimización de la energía. Como se observa en la Figura 7.2, el valor del RMSD alcanza, hacia el final de la simulación, un promedio cercano a 2.0 Å, lo cual indica una desviación moderada respecto a la conformación minimizada. Este comportamiento sugiere que la proteína adopta una nueva conformación estable, consistente con un proceso natural de relajación estructural.

El análisis posterior de una simulación de DM dependerá de los objetivos particulares del estudio. Sin embargo, la mayoría de los trabajos reportan parámetros clave como el R_g , RMSF (Root Mean Square Fluctuation) y la dinámica de los enlaces de hidrógeno, ya que estos permiten caracterizar el comportamiento conformacional, la estabilidad estructural y la red de interacciones internas de la proteína. Particularmente, R_g es un indicador de la compacidad de una proteína. Una proteína plegada de manera estable suele mantener un valor de R_g relativamente constante, mientras que, si comienza a desplegarse, este valor tiende a variar a lo largo del tiempo. Por su parte, el RMSF es una medida que describe la movilidad local de una proteína durante una simulación de dinámica molecular. Permite cuantificar cuánto fluctúa cada átomo o residuo respecto a su posición promedio, lo que ayuda a



identificar regiones rígidas y estables, así como segmentos flexibles o altamente móviles, como loops o extremos terminales. En general, valores bajos de RMSF indican zonas estructuralmente estables, mientras que valores altos señalan regiones dinámicas cuya movilidad puede estar relacionada con la función biológica o con cambios inducidos por mutaciones, interacción con ligandos u otras condiciones del sistema.

Actividad 7.1. Utilizar los siguientes comandos para calcular, respectivamente, el R_g y el RMSF de la SCV2-Mpro a partir de la trayectoria de simulación de DM. Generar las gráficas correspondientes, interpretar los resultados e incluir los comentarios en la bitácora electrónica.

Cálculo del R_g :

```
$ gmx gyrate -s em.tpr -f md_NJMP_FIT.xtc -o gyrate.xvg -tu ns
$ 1
$ xmgrace gyrate.xvg
```

Cálculo del RMSF por residuo

```
$ gmx rmsf -s em.tpr -f md_NJMP_FIT.xtc -o rmsf.xvg -res
$ 1
$ xmgrace rmsf.xvg
```

A continuación, se analizará la evolución del número de enlaces de hidrógeno formados por la SCV2-Mpro durante la simulación. Para ello, se utiliza el siguiente comando:

```
$ gmx hbond -f md_NJMP_FIT.xtc -s em.tpr -num hbond.xvg
```

Cuando se solicite la selección de grupos, debe elegirse la opción 1:1, lo que indica que se analizarán exclusivamente los enlaces de hidrógeno intra-proteicos.

Una vez generado el archivo *hbond.xvg*, se debe visualizar mediante *XMGrace*:

```
$ xmgrace hbond.xvg
```

La gráfica resultante –como se muestra en la Figura 7.3– representa la cantidad de enlaces de hidrógeno formados a lo largo del tiempo. Se observa que el número promedio de estas interacciones tiende a disminuir ligeramente conforme avanza la simulación, posiblemente asociado al proceso de relajación estructural comentado, o reacomodo interno de la proteína bajo condiciones dinámicas.

Hydrogen Bonds

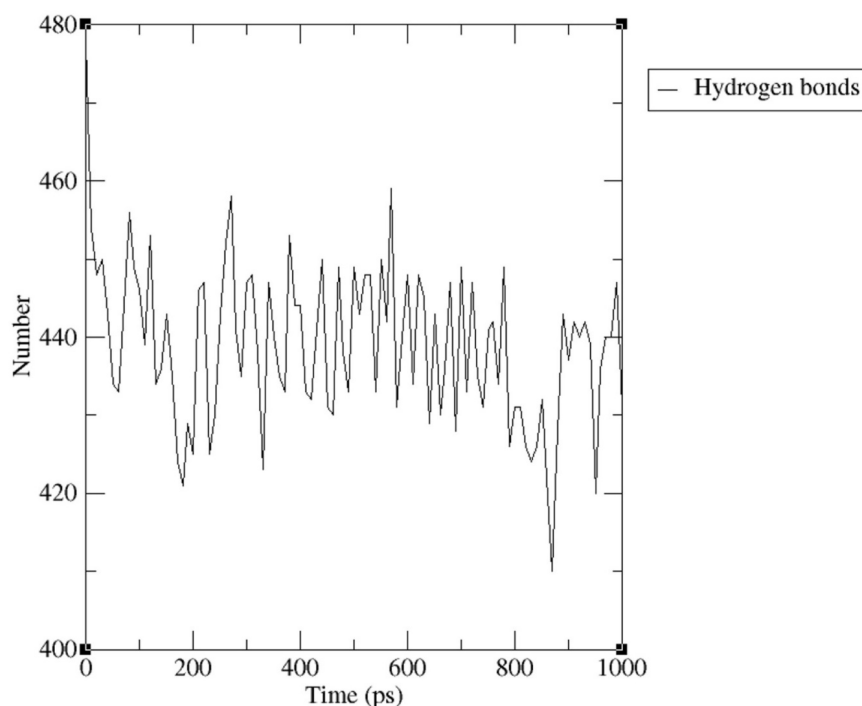


Figura 7.3. Evolución temporal del número de enlaces de hidrógeno intramoleculares formados en la SCV2-Mpro, obtenida mediante el módulo *hbond* de GROMACS y graficada con *XMGrace*. El eje Y representa el número total de enlaces de hidrógeno detectados en cada cuadro de la simulación, mientras que el eje X muestra el tiempo transcurrido en ns. Este análisis se realizó a partir de la trayectoria dinámica del sistema previamente preparado y equilibrado mediante simulación de DM, y permite evaluar la estabilidad de las interacciones internas a lo largo del tiempo.

7.1 Generación de clústeres

La agrupación de conformaciones generadas durante una simulación de DM en función de su similitud estructural constituye una herramienta clave para analizar la flexibilidad de una proteína y la seleccionar racional de estructuras representativas. Este enfoque, conocido como agrupamiento conformacional o *clustering*, y permite identificar los estados estructurales más poblados a lo largo de la trayectoria, facilitando el análisis de estados metaestables, transiciones conformacionales, y la representación de la diversidad estructural del sistema.

Aunque GROMACS incluye algoritmos básicos para realizar *clustering* (como el comando *gmx cluster*), en este caso emplearemos TtClust (consultar el Apéndice I para su instalación), una herramienta externa escrita en Python que permite agrupar conformaciones, generar centroides representativos y producir reportes estadísticos y figuras de apoyo. TtClust admite trayectorias en formatos estándar (*.xtc*, *.pdb*, *.dcd*) y



produce estructuras representativas de cada grupo, útiles para análisis posteriores como el *docking*.

En versiones anteriores, TTClust incluía un procedimiento automático para seleccionar el número óptimo de clústeres mediante el método del codo (*elbow analysis*). No obstante, a partir de la versión 4.10.1, esta función dejó de estar disponible, por lo que el usuario debe especificar manualmente el número de clústeres a generar mediante la opción `-ng` (por ejemplo, `-ng 3`, `-ng 4`, `-ng 5`). En esta práctica se utilizará `-ng 3`, valor que ha sido determinado previamente como adecuado para describir la variabilidad conformacional de la trayectoria de simulación de DM obtenida. Aún así, en el Apéndice II se presentan instrucciones paso a paso para construir el gráfico de codo a partir de la trayectoria de simulación de DM y justificar de forma objetiva la selección de `-ng`.

Para comenzar, se debe abrir una terminal en el directorio de trabajo donde se encuentran la trayectoria centrada (*md_NJMP_FIT.xtc*) y la estructura de referencia (*md.gro*). Ejecutar el siguiente comando:

```
$ ttclust -f md_NJMP_FIT.xtc -t md.gro -ng 3
```

Donde:

- `-f`: archivo de trayectoria (formato *.xtc*)
- `-t`: estructura de referencia (en formato *.gro* o *.pdb*)
- `-ng 5`: indica a TTClust el número de clústeres deseados para describir la trayectoria.

Una vez finalizada la ejecución, TTClust generará una subcarpeta denominada *clustering* que contendrá los archivos de salida del análisis, incluyendo archivos *.pdb* representativos del centroide de cada clúster, archivos de *log* y métricas de agrupamiento y gráficas de análisis de codo y dendrogramas.

Cada uno de los cinco clústers agrupa las conformaciones más frecuentes muestreadas durante la trayectoria (ver Figura 7.4).

Uno de los archivos generados, por ejemplo *C4_f57_s40.pdb*, corresponde a una estructura consenso que representa el *centroide* del clúster más poblado, el cual en este caso agrupa 40 *frames* de la simulación. Esta conformación puede considerarse altamente representativa del comportamiento estructural de la proteína bajo las condiciones simuladas. En conjunto, los cinco centroides obtenidos (uno por clúster) constituyen un conjunto reducido de conformaciones representativas que capturan los estados conformacionales principales de la SCV2-Mpro a lo largo de la trayectoria de DM. Cada una de estas estructuras puede emplearse como receptor en estudios posteriores, por ejemplo, para delimitar y comparar cavidades o sitios potenciales de unión (incluyendo sitios transitorios o alostéricos) y para realizar *docking* con bibliotecas de ligandos, evaluando cómo la variabilidad conformacional de la proteína

influye en la accesibilidad del sitio, la geometría de la cavidad y la plausibilidad de modos de unión.

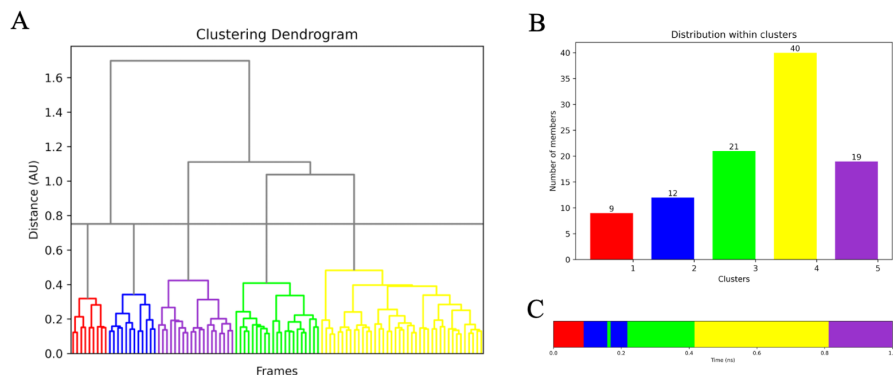


Figura 7.4. Resultados del análisis de agrupamiento conformacional realizado mediante la herramienta TTClust en Python. (A) Matriz de distancias entre clústeres basada en el valor promedio del RMSD, que refleja la similitud estructural entre los grupos identificados. (B) Distribución de estructuras intermedias por clúster, donde se observa la población relativa de cada grupo conformacional, lo que permite identificar el estado más representado. (C) Representación lineal de los clústeres a lo largo de la trayectoria simulada, en la que se visualiza la alternancia y persistencia de diferentes estados conformacionales. El análisis permitió identificar cinco clústeres principales, representativos de los estados estructurales predominantes de la SCV2-Mpro durante la simulación de DM.

Bibliografía

- Bekker, H.**, Berendsen, H.J.C., Dijkstra, E.J., Achterop, S., van Drunen, R., van der Spoel, D., Sijbers, A and Keegstra H. *et al.*, (1993). *Gromacs: A parallel computer for molecular dynamics simulations*. In Physics computing 92. Edited by R. A. de Groot and J. Nadrchal. World Scientific, Singapore.
- Humphrey, W.**, Dalke, A. and Schulten, K.(1996). *VMD - Visual Molecular Dynamics*. Journal of Molecular Graphics and Modelling 14: 33-38.
- Tubiana, T.**, Carvaillo, J. C., Boulard, Y., & Bressanelli, S. (2018). *TTClust: a versatile molecular simulation trajectory clustering program with graphical summaries*. Journal of chemical information and modeling 58(11): 2178-2182.
- van Gunsteren, W.F.** and Berendsen, H.J.C. (1990). *Computer simulation of molecular dynamics: Methodology, applications, and perspectives in chemistry*. Angewandte Chemie International Edition in English 29: 992–1023.



Práctica 8. Acoplamiento molecular y cribado virtual

Objetivos

- Comprender y aplicar los fundamentos del cribado virtual como herramienta de diseño racional de fármacos, mediante la simulación de interacciones proteína–ligando.
- Preparar adecuadamente la proteína blanco y la quimioteca de compuestos *drug-like* para acoplamiento molecular.
- Ejecutar un protocolo automatizado de cribado virtual con AutoDock Vina, integrando herramientas auxiliares para la preparación y procesamiento masivo de ligando.
- Interpretar y comparar los resultados obtenidos a partir de distintas conformaciones del receptor, explorando la influencia de la flexibilidad estructural sobre la predicción de afinidad.

Software especializado requerido

- AutoDock Vina versión 1.1.2 o posterior
- AutoDock Tools versión 1.5.7 o posterior
- Open Babel versión 2.4.1 o posterior

Una vez identificadas las conformaciones más representativas de la proteína blanco, el siguiente paso corresponde a la delimitación de las regiones de su superficie susceptibles de interacción con posibles ligandos. Estas regiones funcionales, como sitios activos o interfaces alostéricas, constituyen los objetivos primarios para el diseño y evaluación de compuestos bioactivos.

A continuación, se realizará la simulación del acoplamiento molecular de un conjunto amplio de compuestos con diversidad estructural, provenientes de quimiotecas: colecciones de moléculas con propiedades fisicoquímicas compatibles con características tipo fármaco (*drug-like*), obtenidas de catálogos comerciales o bases de datos moleculares.

Los programas de *docking* simulan la interacción entre la proteína y cada uno de los ligandos, evaluando múltiples orientaciones y conformaciones del complejo. A partir de ello, se calcula un puntaje que refleja la afinidad teórica, considerando tanto la energía de interacción proteína–ligando como la complementariedad estérica y electrostática. Al finalizar el proceso, se genera una lista ordenada de ligandos en función de su afinidad de unión predicha, lo cual permite seleccionar un subconjunto de compuestos con mayor potencial de interacción, que pueden ser priorizados para ensayos experimentales o etapas posteriores de validación *in silico* (como simulaciones de DM, predicción de propiedades ADMET, entre otras).

Este proceso de filtrado computacional es lo que se conoce como cribado virtual (*virtual screening*), y constituye una herramienta fundamental en las etapas iniciales del Diseño Racional de Fármacos.

El cribado virtual se estructura típicamente en tres etapas fundamentales:

- 
- I. Preparación de la molécula blanco: se requiere contar con la estructura tridimensional de la proteína blanco. Esta puede obtenerse a través de bases de datos experimentales, como el PDB, o mediante métodos de modelado computacional (*in silico*), tales como la predicción basada en inteligencia artificial (por ejemplo, *AlphaFold2*), el modelado por homología, el modelado *ab initio* (*de novo*), o estrategias de ensamblaje de fragmentos (*threading*)
 - II. Delimitación del sitio de unión: es necesario definir con precisión los sitios potenciales de interacción, que pueden corresponder a sitios activos (en el caso de enzimas), sitios alostéricos o interfaces proteína–proteína. Estos sitios guían el posicionamiento de los ligandos durante la simulación de *docking*
 - III. Preparación de la quimioteca de ligandos: los compuestos se seleccionan a partir de bibliotecas moleculares disponibles (por ejemplo, ZINC, ChEMBL, DrugBank, Enamine) y deben ser procesados individualmente para asegurar su compatibilidad con el motor de *docking*. Para cada molécula se debe eliminar contraiones y sales, añadir átomos de hidrógeno faltantes (especialmente en grupos funcionales polares), ajustar estados de protonación a pH 7.4, si es posible (por ejemplo, usando herramientas como Open Babel), minimizar su energía conformacional utilizando campos de fuerza clásicos (por ejemplo, MMFF94, UFF)
 - IV. Simulación del acoplamiento molecular: durante esta etapa, un algoritmo de docking posiciona cada ligando dentro del sitio de unión de la proteína. El motor explora múltiples conformaciones del ligando (rotaciones de enlaces simples), orientaciones relativas proteína–ligando y modos de interacción estérica y electrostática. Cada posible complejo proteína–ligando es evaluado mediante una función de energía o puntuación de afinidad, y se asigna un puntaje de *binding*. Al finalizar, se genera una lista de candidatos ordenados por afinidad teórica. Los ligandos mejor posicionados se consideran potenciales inhibidores o moduladores del blanco proteico y pueden ser seleccionados para validación experimental (*in vitro* o *in vivo*), o bien para simulaciones de refinamiento mediante DM

Entre los programas de uso común para realizar cribado virtual se encuentran: AutoDock 4, AutoDock Vina, MOE, GOLD, FlexX, LUDI, ICM, HADDOCK y Pro_LEDS. Cada uno de estos programas implementa diferentes algoritmos de búsqueda conformacional y funciones de puntuación, por lo que su elección dependerá del tipo de sistema, la disponibilidad de licencias y los objetivos específicos del estudio.

En particular, AutoDock Vina, se ha consolidado como una herramienta muy popular gracias a las siguientes características:

- Es de acceso libre y de código abierto
- Posee una interfaz simple y documentación extensa

- 
- Ofrece una alta velocidad de ejecución en comparación con su predecesor (AutoDock 4)
 - Mantiene un buen equilibrio entre precisión y eficiencia computacional
 - Permite realizar cribado virtual de cientos a miles de ligandos
 - Es compatible con interfaces gráficas como PyRx o AutoDock Tools (ADT), lo que facilita su uso sin conocimientos avanzados en programación

8.1 Preparación del receptor

Para iniciar el proceso de cribado virtual, el primer paso corresponde a la selección y preparación de la estructura tridimensional de la proteína blanco, que fungirá como receptor en el protocolo de acoplamiento molecular. Como se ha discutido a lo largo de este manual, existen múltiples estructuras cristalográficas disponibles de la SCV2-Mpro, tanto en estado libre (*apo*) como en complejos con inhibidores (*holo*). Sin embargo, en la **Práctica 7. Análisis estructural de la trayectoria de simulación de dinámica molecular de una proteína**, se identificaron clústeres conformacionales a partir de una trayectoria de simulación de DM, los cuales representan los estados estructurales más frecuentes de la proteína bajo condiciones fisiológicas simuladas. Entre ellos, el clúster más poblado resultó ser el más representativo, al agrupar 40 conformaciones (frames), lo que sugiere que su estructura consenso (centroide) constituye una opción particularmente adecuada como modelo receptor para realizar acoplamiento molecular con receptor rígido (rigid receptor docking), ya que captura de forma robusta la conformación predominante de la proteína durante la simulación. Para comenzar, se debe crear una carpeta de trabajo exclusiva para esta práctica. Esto permitirá mantener organizados los archivos del análisis de *docking*. En dicha carpeta copiar, se debe copiar el archivo *.pdb* correspondiente a la estructura representativa del clúster más poblado generada con TtClust (por ejemplo, *C4_f57_s40.pdb*). Abrir una terminal en el directorio de trabajo y verificar que la ruta activa sea la correcta mediante el siguiente comando:

```
$ pwd
```

Este comando imprimirá la ruta del directorio actual, la cual será útil más adelante para establecer correctamente las rutas de salida de los archivos generados durante el proceso de acoplamiento molecular. A continuación, inicie ADT desde la terminal utilizando alguno de los siguientes comandos, dependiendo de la configuración del sistema operativo:

```
$ adt (o bien, $ pmv)
```

Nota: es importante asegurarse de que ADT esté correctamente instalado y operativo en el entorno de trabajo. En caso contrario, puede consultarse la documentación

oficial: <https://autodock.scripps.edu/> y seguir las instrucciones del Apéndice I para su instalación.

Una vez dentro de la interfaz de ADT, se debe establecer la carpeta recién creada como directorio de inicio por defecto para futuras sesiones con **File>Preferences>Set... Startup Directory>** (navegue hasta el directorio de trabajo, por ejemplo, /DF/Práctica9). Hacer clic en **Set**, y posteriormente en **Close**.

Con el entorno preparado, cargar la estructura del receptor con **File>Read Molecule>c1_model0.pdb**.

Esto cargará la molécula representativa del clúster 1 generada previamente mediante análisis de la simulación de DM y *clustering* (ver Figura 8.1). A partir de esta estructura, se realizarán los pasos necesarios para su preparación como receptor en el proceso de acoplamiento molecular.

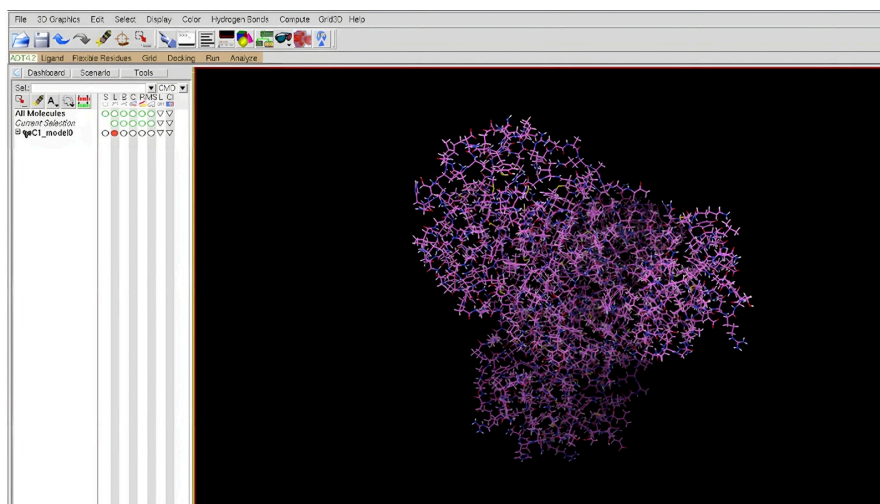


Figura 8.1. Representación en varillas de la SCV2-Mpro visualizada con el programa ADT. Los átomos de hidrógeno se destacan mediante puntos de color blanco, lo cual facilita su identificación y verificación durante la preparación del receptor para el acoplamiento molecular.

El siguiente paso consiste en eliminar los hidrógenos no polares, lo cual simplifica el modelo del receptor sin afectar las interacciones relevantes para el acoplamiento. Para ello, **Edit>Hydrogens>Merge non Polar**.

Al ejecutar esta acción, se observará que algunos átomos blancos desaparecen de la visualización, lo cual es normal y corresponde a la fusión de hidrógenos no polares con sus átomos pesados correspondientes.

El siguiente paso consiste en añadir cargas parciales de Kollman a los átomos del receptor. Esta operación es crítica para el cálculo energético durante el *docking*, ya que determina la distribución electrostática de la molécula receptora. Para realizar esta acción: **Edit>Charges>Add Kollman Charges**. Una vez ejecutado el comando ADT



mostrará un mensaje con la carga neta total de la proteína, confirmando que el procedimiento fue exitoso.

Se recomienda firmemente utilizar cargas Kollman en lugar de recurrir directamente a campos de fuerza genéricos, ya que las primeras están específicamente adaptadas al algoritmo de AutoDock Vina y ofrecen un mejor desempeño en la predicción de energías de unión.

En caso de que se esté utilizando una estructura descargada directamente del PDB, será necesario añadir hidrógenos polares manualmente mediante **Edit>Hydrogens>Add**.

Este paso es esencial para asegurar la protonación correcta de residuos ionizables, como lisina, glutamato, aspartato, histidina, entre otros, especialmente en condiciones fisiológicas.

Una vez completados los pasos preparación, se debe guardar la molécula del receptor en el formato requerido por AutoDock Vina, es decir, *.pdbqt*. Este formato incluye información sobre coordenadas atómicas, cargas parciales y grados de libertad rotacional (aunque para proteínas suelen fijarse rígidamente). Para guardar el archivo: **Grid>Macromolecule>Choose**.

En la ventana emergente, seleccionar la molécula cargada (por ejemplo, *c1_model0*) y proceder a guardarla.

El paso final en la preparación del receptor es definir una caja de búsqueda (*grid box*), la cual delimitará el espacio tridimensional donde AutoDock Vina explorará posibles posiciones de acoplamiento para los ligandos. Para ello ir al menú **Grid>Grid Box**

Se abrirá la ventana **Grid Options**, y se mostrará en la pantalla una caja tridimensional superpuesta a la estructura de la proteína (ver Figura 8.2). Desde esta interfaz se pueden realizar las siguientes acciones:

- I. Ajustar el tamaño de la caja modificando los valores numéricos en los campos **Number of points in x, y, z**, según la extensión del sitio de unión
- II. Visualizar los cambios en tiempo real utilizando los controles tipo dial o mediante manipulación directa con el ratón
- III. Centrar la caja sobre el sitio activo de interés, ya sea escribiendo manualmente las coordenadas en los campos **X center**, **Y center**, **Z center**, o arrastrando la caja directamente en la vista 3D

Debe verificarse que la caja de búsqueda cubra completamente el sitio de unión previsto, sin ser excesivamente grande. Un valor muy alto en **Current Total Grid Pts per map** incrementará innecesariamente el tiempo de cómputo durante el proceso de *docking*. Una vez configurada la caja de búsqueda, se debe guardar los parámetros con **File>Output Grid Dimensions File**.

Asignar un nombre descriptivo al archivo (por ejemplo, *grid_box_SCV2-Mpro.gpf*). Guardar el archivo en el directorio de trabajo actual. Es importante anotar las dimensiones (**Number of points in x, y, z**) y las coordenadas del centro (**X center**, **Y**

center, Z center) de la caja, ya que serán requeridas durante la configuración y ejecución de AutoDock Vina.

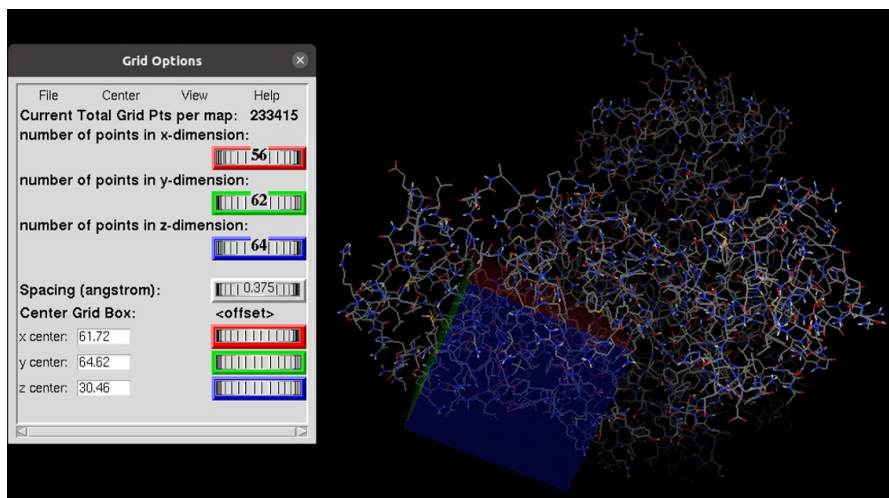


Figura 8.2. Ejemplo de diseño de la *grid box* en ADT, utilizada para delimitar el volumen espacial en el que se realizará el acoplamiento molecular o cribado virtual. La caja fue centrada sobre el sitio activo de la SCV2-Mpro y ajustada para cubrir completamente la región de interés, garantizando una exploración adecuada por parte del programa de *docking*.

Para realizar el cribado virtual, se empleará una quimioteca compuesta por 100 fármacos aprobados por la FDA (Food and Drug Administration). Este conjunto de compuestos se encuentra disponible en el archivo *FDA_Manual.sdf*, incluido como material complementario del presente manual. Descargar el archivo y guardarlo en una subcarpeta denominada *Ligandos* dentro de la carpeta de trabajo.

Antes de realizar el acoplamiento, es necesario optimizar la geometría de los ligandos mediante un proceso de minimización de la energía. Este procedimiento disminuye tensiones conformacionales y genera estructuras tridimensionales más realistas. Primero deberá instalarse Open Babel (Apéndice I), una herramienta de software libre ampliamente utilizada en química computacional y modelado molecular. Su utilidad principal radica en convertir, manipular y analizar archivos moleculares entre una gran variedad de formatos, lo que la convierte en un puente entre diferentes programas de modelado, *docking* o simulación de DM.

Además, permite:

- Generar estructuras 3D a partir de fórmulas químicas o archivos 2D
- Añadir o eliminar hidrógenos, calcular masas moleculares y propiedades básicas
- Optimizar geometrías mediante métodos de mecánica molecular

Preparar ligandos para estudios de acoplamiento o simulación. desde la terminal para ejecutar el siguiente comando en la carpeta *Ligandos*:



```
$ obminimize -o sdf -ff MMFF94 -h FDA_Manual.sdf >
FDA_minimizado.sdf
```

Este comando realiza las siguientes acciones:

- obminimize: ejecuta el módulo de minimización de Open Babel
- -o sdf: Especifica el formato de salida (SDF)
- -ff MMFF94: aplica el *campo de fuerzas* MMFF94, apropiado para compuestos orgánicos
- -h FDA_Manual.sdf: define el archivo de entrada, añadiendo hidrógenos implícitos
- > FDA_minimizado.sdf: redirecciona la salida al nuevo archivo *FDA_minimizado.sdf*

Nota: el proceso puede requerir varias horas, dependiendo del número de ligandos y la capacidad de procesamiento de su sistema.

A diferencia de los ligandos, la estructura de la SCV2-Mpro ya ha sido optimizada estructuralmente durante la simulación de DM, por lo que no requiere una etapa adicional de minimización de energía.

Una vez finalizada la minimización de los ligandos, se debe proceder a generar los archivos *.pdbqt*, formato requerido por AutoDock Vina para realizar el acoplamiento molecular. Para ello es necesario iniciar Open Babel en modo gráfico. En la sección **INPUT FORMAT**, cargar el archivo *FDA_minimizado.sdf* desde la carpeta *Ligandos* (ver Figura 8.3). En las opciones de configuración centrales:

- I. Marcar las casillas de **Add hydrogens from polar atoms** y **Delete hydrogens from non-polar atoms**
- II. En **Calculate partial charges by specified method**, seleccionar **Gasteiger**
- III. En **OUTPUT FORMAT**, seleccionar **pdbqt** y definir el nombre de salida como *ligand*.pdbqt*

Finalmente, hacer clic en **CONVERT**. Este proceso generará 100 archivos *.pdbqt*, uno por cada ligando procesado, dentro de la subcarpeta *Ligandos*. Para verificar que tanto el receptor como los ligandos contienen las cargas parciales correctamente asignadas, abrir cualquier archivo *.pdbqt* con un editor de texto, por ejemplo con *gedit*:

```
$ gedit C1_model0.pdbqt
```

Ubicar la penúltima columna del archivo, la cual debe contener el valor de la carga parcial asignada a cada átomo. Por ejemplo, se observará un valor como el siguiente, correspondiente al nitrógeno terminal del residuo SER1 (residuo inicial de la proteína):

```
EMARK 4 XXXX COMPLIES WITH FORMAT V. 2.0
TOM 1 N SERA 1 85.837 62.394 33.549 1.00 0.00 -0.055 N
```

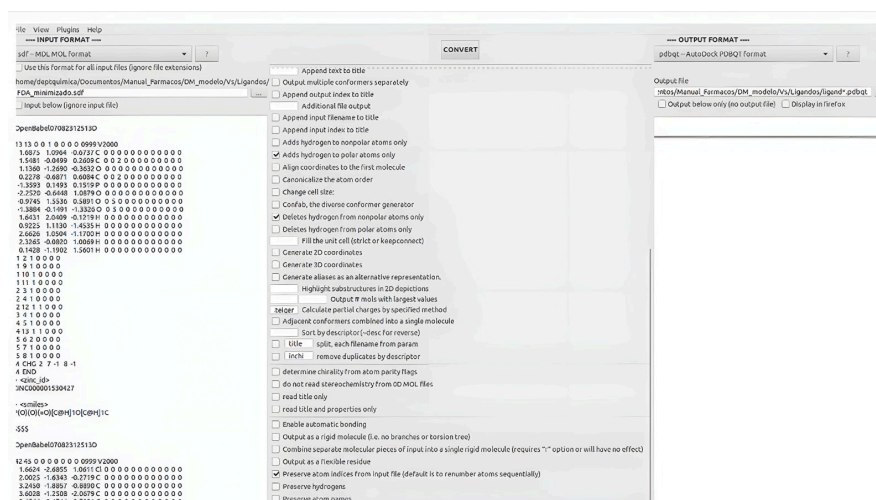


Figura 8.3. Interfaz gráfica de Open Babel utilizada para el procesamiento de ligandos o moléculas pequeñas previo al acoplamiento molecular. Se visualizan las opciones seleccionadas durante la preparación de los ligandos, incluyendo la adición de hidrógenos a átomos polares, eliminación de hidrógenos no polares, asignación de cargas parciales de Gasteiger y conversión al formato *.pdbqt*, requerido para su uso en acoplamiento molecular con AutoDock Vina.

Repetir esta verificación en algunos de archivos *ligand*.pdbqt* para verificar que el preprocesamiento de ligandos se ha realizado correctamente.

8.2 Cribado virtual: obtención y edición del archivo *batch*

Con el fin de ejecutar el cribado virtual de una quimioteca de manera automatizada (*batch*), se utilizará un conjunto de *scripts* en *shell* que permiten evaluar múltiples ligandos sin intervención manual. Se parte del supuesto de que AutoDock Vina ha sido instalado correctamente y se encuentra accesible en la carpeta *bin* del sistema (por ejemplo, */usr/local/bin/vina*), o que se conoce su ruta exacta.

Desde la carpeta de trabajo, descargar los *scripts* *config.txt*, *nameLoop.sh* y *vina_screen_local.sh* desde <https://github.com/skandasastry/VS-autodock-vina>.

Una vez descargados, abrir la terminal en la carpeta de trabajo y ejecutar:

```
$ gedit vina_screen_local.sh
```

Esto abrirá el *script* principal (ver Figura 8.4), el cual contiene las instrucciones que AutoDock Vina ejecutará para cada ligando contenido en la subcarpeta *Ligandos*.

Como resultado del proceso de acoplamiento, se generarán los siguientes archivos para cada ligando:

- *out.pdbqt*: estructura del ligando correspondiente a la mejor pose de acoplamiento predicha
- *log.txt*: archivo de registro que incluye la energía de acoplamiento estimada para cada interacción proteína–ligando



```
1 #! /bin/bash
2
3 for f in Sample_Ligands/Ligand_*.pdbqt; #iterates through a folder with ligand files
4 do
5     b='basename $f .pdbqt'
6     echo Processing ligand $b
7     mkgtr -p $b
8     #vina command for a config file "config.txt", outputs an out file with docking scores and a log file
9     vina --config config.txt --ligand $f --out $(b)/out.pdbqt --log $(b)/log.txt
10 done
```

Figura 8.4. Imagen digital del archivo *vina_screen_local.sh* utilizado como *script batch* para la ejecución automatizada del cribado virtual mediante AutoDock Vina. El archivo recorre los archivos de todos los ligandos en formato *.pdbqt* contenidos en la carpeta especificada, realiza el acoplamiento molecular con el receptor definido en *config.txt*, y genera archivos de salida que incluyen las mejores poses (**_out.pdbqt*) y registros de la energía de interacción (**_log.txt*) para cada ligando evaluado.

Se debe editar el archivo *vina_screen_local.sh* según las necesidades de cada usuario. Es conveniente verificar que las rutas y nombres de los archivos de entrada –receptor, ligandos y archivo de configuración (*config.txt*)– coincidan con la estructura y ubicación de archivos del entorno de trabajo. La versión final del *script*, tras su edición, deberá presentar un formato similar al mostrado en la Figura 8.5.



```
1 #! /bin/bash
2
3 for f in Ligandos/Ligand*.pdbqt; #iterates through a folder with ligand files
4 do
5     b='basename $f .pdbqt'
6     echo Processing ligand $b
7     mkgtr -p $b
8     #vina command for a config file "config.txt", outputs an out file with docking scores and a log file
9     vina --config config.txt --ligand $f --out $(b)/out.pdbqt --log $(b)/log.txt
10 done
```

Figura 8.5. Ejemplo de archivo *vina_screen_local.sh* modificado para ejecutar el cribado virtual con AutoDock Vina. En esta versión adaptada se especifica el receptor, se eliminan referencias explícitas al nombre del ligando (para permitir el procesamiento automatizado por lotes) y se asegura la ejecución secuencial del acoplamiento de todos los ligandos presentes en la carpeta correspondiente, utilizando los parámetros definidos en el archivo *config.txt*.

El fragmento modificado del *script* *vina_screen_local.sh* recorre todos los archivos *.pdbqt* contenidos en la carpeta *Ligandos*, ejecuta AutoDock Vina para cada uno, y guarda los resultados con sufijos únicos que identifican cada ligando procesado.

El archivo *config.txt* contiene los parámetros esenciales para la ejecución del acoplamiento molecular. En particular, define:

- Coordenadas del centro de la caja de acoplamiento: **X center, Y center, Z center**
- Dimensiones de la caja: **Number of points in x, y, z,**
- Ruta del archivo del receptor: *receptor*
- Parámetros de exhaustividad: *exhaustiveness*, que controla la profundidad de búsqueda del algoritmo

Es importante eliminar la línea que define el ligando (*ligand*) dentro del archivo *config.txt*, ya que los ligandos serán definidos dinámicamente por el script por lotes.

A continuación, se muestra un ejemplo del archivo *config.txt* ajustado:

```
1 receptor = C1_model0.pdbqt2
2 center_x = 61.7193
3 center_y = 64.6144
4 center_z = 30.4615
5 size_x = 566
6 size_y = 627
7 size_z = 668
8 exhaustiveness = 89
9 num_modes = 10
```

Nota: sustituir los valores de **center_*** y **size_*** por los obtenidos previamente al definir la caja de acoplamiento en ADT. El parámetro *exhaustiveness* puede aumentarse para mejorar la exploración conformacional del ligando, aunque a costa de mayor tiempo de cálculo. Para esta práctica, 20 es un valor apropiado.

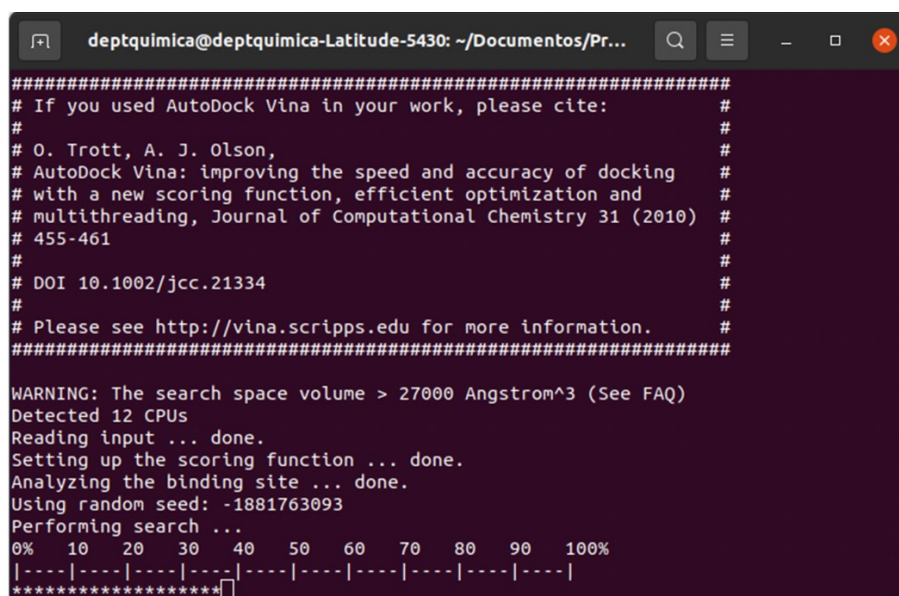
8.3 Ejecutando el cribado virtual

Una vez configurado el entorno de trabajo y verificada la preparación del receptor, los ligandos (archivo *FDA_Manual.sdf*) y los archivos de control, puede ejecutarse el cribado virtual entre la SCV2-Mpro modelada y la quimioteca de fármacos optimizados. Para ello, abrir una nueva terminal dentro de la subcarpeta de trabajo (la misma que contiene los scripts y la subcarpeta Ligandos) y ejecutar los siguientes comandos para iniciar el proceso:

```
$ chmod +x vina_screen_local.sh
$ ./vina_screen_local.sh
```

El primer comando otorga permisos de ejecución al script, mientras que el segundo lo ejecuta.

Durante la ejecución, se visualizará un mensaje similar al mostrado en la Figura 8.6, indicando el procesamiento de cada uno de los ligandos contenidos en la carpeta. Para cada ligando, el script creará un subdirectorío con su nombre, dentro del cual se almacenan dos archivos claves: *out.pdbqt* (estructura del complejo proteína–ligando en su mejor pose) y *log.txt* (registro de la energía estimada de acoplamiento).



```
deptquimica@deptquimica-Latitude-5430: ~/Documentos/Pr...
#####
# If you used AutoDock Vina in your work, please cite:      #
#                                                           #
# O. Trott, A. J. Olson,                                     #
# AutoDock Vina: improving the speed and accuracy of docking #
# with a new scoring function, efficient optimization and    #
# multithreading, Journal of Computational Chemistry 31 (2010) #
# 455-461                                                      #
#                                                           #
# DOI 10.1002/jcc.21334                                       #
#                                                           #
# Please see http://vina.scripps.edu for more information.   #
#####

WARNING: The search space volume > 27000 Angstrom^3 (See FAQ)
Detected 12 CPUs
Reading input ... done.
Setting up the scoring function ... done.
Analyzing the binding site ... done.
Using random seed: -1881763093
Performing search ...
0% 10 20 30 40 50 60 70 80 90 100%
|----|----|----|----|----|----|----|----|----|
*****
```

Figura 8.6. Imagen digital de la terminal de comandos durante la ejecución del cribado virtual con AutoDock Vina. La parte superior de la ventana muestra la referencia bibliográfica recomendada para el uso de AutoDock Vina, mientras que la sección inferior proporciona información en tiempo real sobre el progreso de las simulaciones de acoplamiento, incluyendo el nombre de cada ligando procesado y el valor de energía de acoplamiento calculado para su mejor pose.

Una vez completado el cribado virtual, es necesario reunir y organizar los resultados en un solo archivo para facilitar su interpretación. Para evitar tener que revisar manualmente cada *log.txt*, se utilizarán dos scripts auxiliares que automatizan la extracción de los resultados y permiten construir un resumen global de la afinidad calculada para cada ligando.

El primer *script* para utilizar es *nameLoop.sh* (ver Figura 8.7), el cual ya fue descargado previamente. Este archivo se debe copiar los subdirectorios generados tras la ejecución del cribado virtual. Antes de ejecutar el *script*, se debe crear una carpeta denominada *.log* dentro del mismo directorio, que servirá para almacenar los archivos recolectados:

```
$ mkdir log
```

Ahora se debe abrir el *script* con el siguiente comando:

```
$ gedit nameLoop.sh
```

Este *script* busca cada archivo *log.txt* en los subdirectorios generados por el *docking*, y copia su contenido en archivos individuales dentro de una nueva carpeta llamada *log*. Durante este proceso, renombra automáticamente cada archivo con base en el nombre del ligando correspondiente, lo que facilita su posterior identificación.

Script original

```
1#!/bin/bash
2
3#4 files in sample, edit the second number for the amount of files you have docked
4for i in {1..4};
5do
6    str1="log"
7    str2=".txt"
8    #make sure to replace Sample_Ligands with your folder of ligands as needed
9    cp Ligand_$i/log.txt Sample_Ligands/$str1$i$str2
10done
11
```

Script modificado

```
1#!/bin/bash
2
3#4 files in sample, edit the second number for the amount of files you have docked
4for i in {1..100};
5do
6    str1="log"
7    str2=".txt"
8    #make sure to replace Sample_Ligands with your folder of ligands as needed
9    cp /home/deptquimica/Documentos/Manual_Farmacos/DM_modelo/Vs/Resultados/-
    ligand$i/log.txt /home/deptquimica/Documentos/Manual_Farmacos/DM_modelo/Vs/-
    Resultados/log/$str1$i$str2
10done
11
```

Figura 8.7. Imagen digital de un archivo tipo *script nameLoop.sh*, utilizado para automatizar la extracción de los resultados de cribado virtual. Se muestra el contenido del archivo antes y después de ser modificado, destacando la inclusión de la ruta de destino donde se almacenarán los archivos *log.txt* generados por AutoDock Vina en cada subcarpeta correspondiente a los ligandos analizados. Esta modificación permite organizar eficientemente los datos en una carpeta centralizada.

Ahora se deben ejecutar los siguientes comandos en la terminal para otorgar permisos de ejecución al script y ponerlo en marcha:

```
$ chmod +x nameLoop.sh
$ ./nameLoop.sh
```

Tras la ejecución, se obtendrán 100 archivos *log.txt*, uno por cada ligando analizado, ahora ubicados dentro de la carpeta *log*.

Para ordenar los resultados en función de la energía de acoplamiento más favorable (es decir, el valor más negativo) obtenida para cada ligando en su mejor pose, se debe descargar el *script vs_analysis.py* desde <https://github.com/MunibaFaiza/VS-Analysis>. Se debe guardar el archivo dentro de la carpeta *log* y ejecutar el siguiente comando:

```
$ python3 vs_analysis.py
```

Este script analizará cada archivo *log.txt* y extraerá el valor de la energía de acoplamiento más favorable para cada fármaco. Al finalizar, generará un archivo *output.txt*, en el cual se listan todos los ligandos junto con sus respectivas energías, ordenados de forma ascendente (de mayor afinidad a menor). Este archivo puede



abrirse con *LibreOffice Calc* o *Microsoft Excel* u otro programa de hojas de cálculo para su inspección o presentación gráfica.

Consideraciones adicionales

Una de las limitaciones del *docking* es el tratamiento del receptor como una estructura rígida. Aunque existen algoritmos capaces de incorporar cierto grado de flexibilidad –ya sea en residuos específicos del sitio activo o mediante enfoques más complejos como el *docking* inducido por ajuste conformacional–, su aplicación resulta computacionalmente costosa y poca práctica al manejar bibliotecas extensas de compuestos.

Una alternativa viable es el *docking* por ensamble, estrategia que permite abordar la flexibilidad del receptor mediante el uso de múltiples conformaciones representativas de la misma proteína. Estas conformaciones pueden derivarse de estructuras cristalográficas disponibles o, como en este caso, de simulaciones de DM. En la presente práctica fue identificada la conformación más representativa del clúster más poblado de la SCV2–Mpro; sin embargo, el análisis puede extenderse realizando el cribado sobre las estructuras promedio de otros clústeres obtenidos.

Actividad 8.1. Realice un cribado virtual adicional empleando como receptor la estructura representativa (centroide) de al menos otro clúster identificado en la Práctica 7 (Generación de conformaciones proteicas mediante dinámica molecular). Una vez obtenido el ranking de afinidades de unión, compárelo con el ranking del cribado realizado previamente en esta práctica y analice las diferencias observadas considerando los siguientes aspectos:

- Afinidad de unión: compare las puntuaciones de *docking* de los ligandos mejor clasificados (por ejemplo, top 10 o top 20).
- Poses de unión: evalúe si los ligandos conservan su orientación y los contactos clave con el sitio de unión, o si se observan cambios en el modo de interacción.
- Ligandos compartidos: determine cuántos ligandos aparecen en ambos rankings dentro del conjunto de mejores candidatos, e interprete este solapamiento en términos de la influencia de la conformación del receptor.

Con base en este análisis, discuta brevemente cómo la variabilidad conformacional de la proteína, capturada mediante dinámica molecular, impacta los resultados del cribado virtual.

Bibliografía

Fan, J., Fu, A., & Zhang, L. (2019). *Progress in molecular docking*. Quantitative Biology 7: 83-89.

Jaghoori, M. M., Bleijlevens, B., & Olabarriaga, S. D. (2016). *1001 Ways to run AutoDock Vina for virtual screening*. Journal of computer-aided molecular design 30: 237-249.



O'Boyle, N. M., M. Banck., C. A. James., C. Morley., T. Vandermeersch., G. R. Hutchison. (2011). Open Babel: An open chemical toolbox. *Journal of Cheminformatics* (7)3: 1-14.

Schneider, G., & H. J. Böhm (2002). Virtual screening and fast automated docking methods. *Drug discovery today* 7: 64-70.



Práctica 9. Diseño de inhibidores para la proteasa Mpro del SARS-CoV-2 utilizando un enfoque basado en fragmentos

Objetivos

- Comprender los fundamentos estructurales y funcionales de la Mpro del SARS-CoV-2 como blanco terapéutico.
- Analizar estructuras cristalográficas de la Mpro del SARS-CoV-2 en estado libre y en complejo con fragmentos moleculares para identificar el sitio activo, su flexibilidad estructural y su proximidad con la interfaz dimérica.
- Explorar e interpretar interacciones moleculares clave entre los fragmentos co-cristalizados y la proteasas utilizando UCSF Chimera.
- Diseñar racionalmente al menos dos inhibidores teóricos a partir de fragmentos combinados, justificando su potencial farmacológico con base en principios de química medicinal, criterios estructurales y las reglas de Lipinski.

Software especializado requerido

- UCSF Chimera versión 1.18 o posterior

Durante las últimas décadas, se han desarrollado múltiples esfuerzos científicos para identificar fármacos efectivos contra enfermedades causadas por coronavirus. Inicialmente, dichos esfuerzos se enfocaron en el SARS-CoV-1 (2002), que infectó a cerca de 800 mil personas; posteriormente, en el MERS-CoV (2012), caracterizado por una mayor tasa de mortalidad, pero con un menor número de casos; y más recientemente, en el SARS-CoV-2, agente etiológico de la pandemia de COVID-19, responsable del mayor número de infecciones y muertes.

A pesar del vasto conocimiento acumulado, aún persisten numerosos aspectos por esclarecer, particularmente en lo referente al Diseño Racional de Fármacos dirigidos a blancos virales específicos.

Uno de los principales blancos terapéuticos en el SARS-CoV-2 es su proteasa principal, Mpro (también conocida como 3CLpro). Esta enzima ha captado la atención de la comunidad científica debido a su papel esencial en el procesamiento de las poliproteínas virales, un paso indispensable para la replicación del virus. Su inhibición específica representa una estrategia prometedora, ya que interrumpe una función clave en el ciclo de replicación del patógeno sin afectar enzimas humanas homólogas, reduciendo así el riesgo de efectos secundarios.

En el caso del SARS-CoV-2, las poliproteínas virales *pp1a* y *pp1ab* deben ser procesadas por dos proteasas codificadas por el propio virus: la SCV2-Mpro y la PLpro. En particular, la SCV2-Mpro reconoce y escinde secuencias específicas en los precursores virales, con preferencia por enlaces peptídicos entre residuos de glutamina y aminoácidos pequeños como glicina, alanina o serina. Esta especificidad de reconocimiento no es compartida por ninguna proteasa humana, lo que hace de



la SCV2–Mpro en un blanco terapéutico atractivo y selectivo, con bajo riesgo de provocar efectos *off-target*.

El desarrollo de inhibidores capaces de unirse al sitio activo de la SCV2–Mpro con alta afinidad y especificidad permitiría bloquear el ciclo de replicación viral de forma eficaz. Este tipo de compuestos explota la complementariedad estructural y electrónica con el sitio catalítico de la enzima, optimizando interacciones moleculares como enlaces de hidrógeno, interacciones iónicas, dipolo–dipolo, apilamiento π , así como fuerzas de *van der Waals*.

Estrategias similares han demostrado ser exitosas en el caso de otras enfermedades virales, como la infección por VIH, donde el diseño de inhibidores dirigidos a las proteasas virales ha permitido el desarrollo de antivirales altamente efectivos. Por tanto, aplicar esta lógica molecular al diseño de inhibidores para la SCV2–Mpro representa un enfoque racional y prometedor en la búsqueda de nuevos tratamientos contra el COVID-19.

En esta práctica se trabajará con estructuras tridimensionales de la SCV2–Mpro en complejo con fragmentos moleculares, obtenidas mediante cristalografía de rayos X en el centro de investigación Diamond Light Source (Reino Unido). Estos fragmentos corresponden a pequeñas moléculas orgánicas (habitualmente con masa molecular inferior a 200 Da) que interactúan con diferentes regiones de la superficie de la proteína.

A partir del análisis estructural de dichos complejos, se propondrá el diseño de un nuevo inhibidor tipo *drug-like*, mediante la unión racional de dos o más fragmentos complementarios, con el objetivo de mejorar tanto la afinidad de unión como la especificidad hacia el sitio activo de la SCV2–Mpro. Este enfoque se fundamenta en principios del Diseño Basado en Fragmentos (Fragment–Based Drug Design, FBDD), ampliamente utilizado en la generación de nuevos candidatos terapéuticos con propiedades moleculares optimizadas.

Actividad 9.1. Realizar una búsqueda y elaborar una lista con el nombre de las vacunas y fármacos disponibles a la fecha para el tratamiento del COVID-19.

Para el caso de los fármacos, investigar su mecanismo de acción.

Ir a la página principal del sincrotrón Diamond Light Source (<https://www.diamond.ac.uk/covid-19/for-scientists/Main-protease-structure-and-XChem.html>), y elaborar una breve reseña sobre la información presentada.

Para observar correctamente la estructura dimérica de la SCV2–Mpro en ausencia de ligandos, es necesario realizar algunos ajustes visuales en UCSF Chimera antes de cargar el ensamblaje biológico. Los pasos por seguir son:

- I. Iniciar sesión en UCSF Chimera

- 
- II. Cargar la estructura cristalográfica de la SCV2-Mpro sin ligando (PDB ID 6YB7)
 - III. Cambiar el fondo de la ventana principal a color blanco para mejorar la visibilidad, seleccionando: **Presets>Publication 3** (depth-cued, rounded ribbon)
 - IV. Para visualizar el ensamblaje biológico dimérico ir a **Favorites>Model Panel** y seleccionar el modelo cargado (6YB7). Hacer clic en **biological unit** y presionar **Close** para aplicar los cambios

Este procedimiento expandirá la estructura monomérica a su forma dimerica funcional, como se encuentra en condiciones fisiológicas.

Para facilitar la identificación de cada monómero, se recomienda aplicar colores distintos mediante **Tools>Depiction>Rainbow**, seleccionando la opción **Change color every: model** y confirmando con **Ok**.

Con la visualización actual, se puede evaluar si el sitio activo de la SCV2-Mpro se encuentra próximo a la interfaz entre ambos monómeros. Esta información es crítica, ya que, en caso afirmativo, será necesario incluir ambos monómeros en los estudios de FBDD, y en las simulaciones de acoplamiento molecular, a fin de representar adecuadamente el entorno catalítico y las posibles interacciones intermoleculares.

Para facilitar la identificación de posibles sitios catalíticos o cavidades relevantes, activar la representación de isosuperficie para cada monómero con la opción **Actions>Surface Show**.

Girar la estructura tridimensional para observar hendiduras o cavidades en ambos monómeros. En caso de que la inspección visual no sea concluyente, realizar un alineamiento estructural con una proteína co-cristalizada con un inhibidor para transferir la información espacial del sitio de unión.

En la **Práctica 3. Examinando interacciones covalentes y no covalentes en un complejo proteína-ligando relacionado con el COVID-19**, se examinó un complejo proteína-inhibidor de la SCV2-Mpro unida al inhibidor de naturaleza peptídica N3 (PDB ID 6LU7). Para cargar dicha estructura, seleccione **File>Fetch by ID>PDB** e ingresar **6LU7** y hacer clic en **Ok**.

Una vez cargada, la comparación entre estructuras cristalográficas de la misma proteína resulta fundamental para evaluar la conservación estructural, la flexibilidad del sitio activo y las diferencias inducidas por la unión de ligandos. UCSF Chimera permite realizar un alineamiento estructural entre modelos de forma sencilla, útil para superponer estructuras de la SCV2-Mpro en estado *apo* y *holo*. A continuación, se describe el procedimiento para alinear 6LU7 (complejo con inhibidor) con 6YB7 (estructura sin ligando):

- I. Ir a **Favorites>Model Panel**
- II. Desmarcar la casilla correspondiente a **MSMS main surface of 6YB7** en la columna **S (Shown)** para ocultar la superficie

- 
- III. Abrir la herramienta de alineamiento estructural desde **Tools>Structure Comparison>MatchMaker**
 - IV. En la ventana emergente, seleccionar **6YB7 (#0)** como **Reference structure** y **6LU7 (#2)** como **Structure(s) to match**
 - V. Dejar los parámetros predeterminados sin modificar y hacer clic en **OK**

Como resultado, se podrá identificar con mayor precisión la ubicación del sitio activo, su accesibilidad, y evaluar si se encuentra próximo a la interfaz del dímero. Esto servirá de guía para el análisis de fragmentos, el diseño de inhibidores más potentes y la posterior definición de cajas de *docking* con mayor criterio estructural (ver Figura 9.1).

Actividad 9.2. Discutir el resultado del alineamiento estructural mostrado en el *Reply Log* de *Favorites*, comentado si las estructuras alineadas son muy similares o presentan diferencias significativas, tanto en términos del valor de RMSD como de la superposición visual de sus dominios.

Comparar la estructura del inhibidor Ne con la de CY6, destacando semejanzas y diferencias en su tamaño, grupos funcionales y tipo de interacciones que cada uno establece con la SCV2–Mpro.

Para completar el análisis estructural del complejo SCV2–Mpro:N3 (PDB ID 6LU7), y observar de forma precisa la interacción del inhibidor con la proteína, es necesario realizar una serie de ajustes en UCSF Chimera. Esto incluye habilitar la visualización del dímero completo, ocultar los residuos estándar y resaltar los no estándar (incluyendo ligandos y moléculas co-cristalizadas):

- I. Utilizar el botón **Biological Unit** en el **Model Panel** para añadir el segundo monómero correspondiente a la estructura **6LU7**. Verificar que la fila etiquetada como 6LU7 esté seleccionada (resaltada) para que la acción se aplique correctamente
- II. Comprobar que ambos homodímeros (**6YB7** y **6LU7**) estén alineados en el espacio tridimensional
- III. Seleccionar todos los residuos estándar mediante el menú **Select>Residue>standard amino acids** y ocultarlos con **Actions>Atoms/Bonds>Hide**
- IV. Seleccionar los residuos no estándar mediante **Select>Residue>all non standard** y mostrarlos con **Actions>Atoms/Bonds>Show**

Con ello se visualizarán todas las moléculas co-cristalizadas con la enzima, incluyendo agua, etilenglicol (PEG), dimetilsulfóxido (DMS), así el inhibidor peptídico N3. Dado que N3 está compuesto por residuos tanto estándar como no estándar, no se puede seleccionar directamente desde la etiqueta **N3**; por lo tanto, para visualizarlo completamente, se deberá seleccionar por partes.

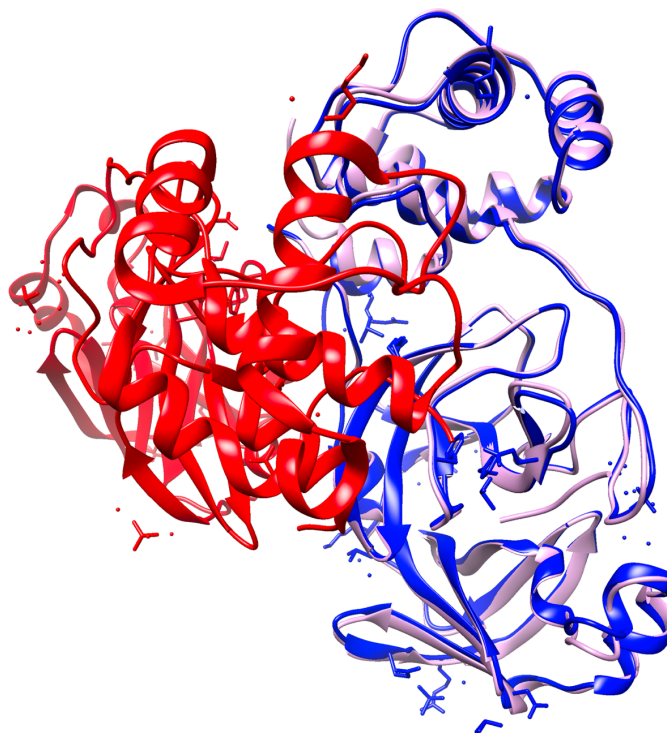


Figura 9.1. Alineamiento estructural de los monómeros de la SCV2-Mpro, en estado *apo* (PDB ID 6YB7) y en complejo con el inhibidor peptídico N3 (PDB ID 6LU7), generado con UCSF Chimera version 1.18 mediante el módulo **MatchMaker**. La superposición permite comparar la conformación del sitio activo en ambos estados estructurales y visualizar posibles cambios inducidos por la unión del inhibidor.

Actividad 9.3. Investigar y discutir:

¿Bajo qué condiciones experimentales pueden resolverse moléculas de agua en estructuras cristalográficas de proteínas? Considerar factores como la resolución, la calidad del cristal y el modelado del mapa de densidad electrónica.

¿Qué tipo de información puede inferirse a partir de la localización de moléculas de agua en sitios específicos sobre la superficie o en el sitio activo de la enzima? Comentar su posible papel en estabilidad estructural, puentes de hidrógeno, reconocimiento de sustratos o mecanismos catalíticos.

Para continuar, ocultar nuevamente las moléculas de agua, PEG y DMS tal como se hizo con los residuos estándar. En pantalla deberán permanecer visibles únicamente los residuos no estándar del inhibidor N3.



Ahora, seleccionar todos los átomos del ligando y aquellos residuos de la proteasa ubicados a una distancia ≤ 5.0 Å respecto de cualquier átomo visible del ligando. Para ello, utilizar el siguiente procedimiento:

- I. Seleccionar un átomo del primer fragmento de N3 con **Ctrl** + clic
- II. Agregar a la selección otro átomo del ligando con **Shift** + **Ctrl** + clic
- III. Expandir la selección completa del ligando utilizando la tecla flecha hacia arriba o la tecla **double up arrow**
- IV. Ir a **Select>Zone...** y, en la ventana **Select Zone Parameters**, verificar que esté activada la opción **Select all atoms/bonds of any residue in selection zone**
- V. Hacer clic en **Ok**

Una vez hecha la selección, mostrar los átomos y enlaces con **Actions>Atoms/Bonds>Show**.

Este procedimiento desplegará el ligando N3 completo, junto con las cadenas laterales de los residuos de la proteasa que se encuentran dentro de un radio de 5.0 Å, tanto en la subunidad A como en la B (si están presentes). Finalmente, asegurarse de ocultar nuevamente las moléculas de agua, como se realizó anteriormente.

Por último, ocultar el modelo de la proteasa sin ligando desde el **Model Panel**. En este punto, debería visualizarse claramente al inhibidor N3 unido al residuo Cys145 de la SCV2–Mpro a través de un enlace covalente. Esta interacción es característica de los inhibidores covalentes, como lo es también el compuesto CY6.

Finalmente, para centrar la visualización en uno de los ligandos, seleccionar cualquier átomo de este (**Ctrl** + clic), expandir la selección a toda la molécula mediante la flecha hacia arriba del teclado y, posteriormente, ir a **Actions>Focus** para centrar la vista en el ligando seleccionado.

Para visualizar la isosuperficie del homodímero de la SCV2–Mpro, activar las casillas correspondientes a las superficies en el **Model Panel**. Ajustar la orientación del sistema para obtener una perspectiva adecuada. Si la superficie aparece seccionada o incompleta, modificar los planos de recorte mediante **Tools>Viewing Controls>Side View**, y deslizar las líneas verticales amarillas hacia la izquierda o derecha según sea necesario para restaurar la vista completa.

Ahora bien, una situación frecuente en el Diseño Racional de Fármacos es ¿qué ocurriría si no se contara con una estructura cristalográfica de la proteasa en complejo con un inhibidor? ¿Sería posible identificar el sitio activo utilizando otras estrategias? Una pista clave es que se sabe que una cisteína catalítica participa directamente en la unión del inhibidor. A partir de esta información, es posible inferir la localización del sitio activo mediante la visualización selectiva de residuos CYS en la proteína sin ligando. Proceder como se indica a continuación:

- I. Apagar (deseleccionar) las representaciones de isosuperficie y cinta de ambos monómeros de 6LU7 desde el **Model Panel**

- 
- II. Activar la representación en cinta del homodímero correspondiente a 6YB7
 - III. Ocultar todos los residuos mediante **Select>Select All** y luego **Actions>Atoms/Bonds>Hide**
 - IV. Visualizar solamente los residuos de cisteína con **Select>Residue>CYS**, seguido de **Actions>Atoms/Bonds>Show**

Actividad 9.4. Evaluar si el sitio activo se encuentra expuesto al solvente o si se localiza en la interfaz entre los monómeros.

¿Es indispensable la presencia del homodímero completo para realizar una simulación de acoplamiento molecular significativa? Justificar la respuesta con base en la organización estructural y el entorno del sitio catalítico.

Posteriormente, elaborar una representación visual del homodímero destacando ambos sitios catalíticos libres. Para ello, se pueden utilizar recursos como colorear por cadena, generar superficies parciales, aplicar transparencia selectiva y etiquetar residuos clave, a fin de obtener una visualización clara y científicamente informativa. Finalmente, capturar una imagen de la representación final y agregarla al informe.

A continuación, se deben explorar los fragmentos disponibles para el diseño racional de un inhibidor dirigido al sitio activo de la SCV2-Mpro. El enfoque de FBDD, consiste en identificar compuestos de baja masa molecular (≤ 200 Da) que se unen con afinidad moderada pero específica hacia subregiones del sitio activo de la proteína blanco. Estos fragmentos, al unirse a zonas complementarias del sitio catalítico, proporcionan una base estructural para el diseño racional. Posteriormente, mediante estrategias de química medicinal, es posible enlazar dos o más fragmentos –ya sea mediante uniones covalentes, enlaces flexibles, o mediante la adición de grupos conector– para generar un compuesto final que presente mayor afinidad y especificidad por el sitio activo, sea sintéticamente accesible, y cumpla con criterios de *drug-likeness*, como los propuestos por Lipinski. Estas reglas, conocidas como la “regla de las cinco” (Rule of Five), establecen parámetros empíricos que ayudan a predecir si una molécula posee propiedades adecuadas para ser absorbida por vía oral. En general, un compuesto con potencial farmacológico debe presentar un peso molecular menor a 500 Da, no más de 5 donadores y 10 aceptores de puentes de hidrógeno, y un valor de logP inferior a 5, lo que refleja un equilibrio adecuado entre lipofilia e hidrofilia.

Para comenzar el análisis de fragmentos moleculares obtenidos mediante cristalografía de rayos X, acceder al repositorio de estructuras del Diamond Light Source en <https://www.diamond.ac.uk/covid-19/for-scientists/Main-protease-structure-and-XChem/Downloads.html> (ver Figura 9.2). Luego:

- I. Acceder a la sección titulada **Non covalent fragment hits in the active site** y descargar el archivo *Mpro_XChem_screen_active_site_11 May 2020.zip*

- II. Descomprimir el archivo *.zip* en la carpeta de elección (contiene 23 archivos *.pdb*, cada uno con una estructura cristalográfica de SCV2–Mpro en complejo con un fragmento distinto)
- III. Abrir UCSF Chimera e iniciar una nueva sesión
- IV. Ir al menú **File>Open** y navegar hasta la carpeta donde se encuentran los archivos *.pdb*. Seleccionarlos (por ejemplo, haciendo **Shift** + clic) y hacer clic en **Open**

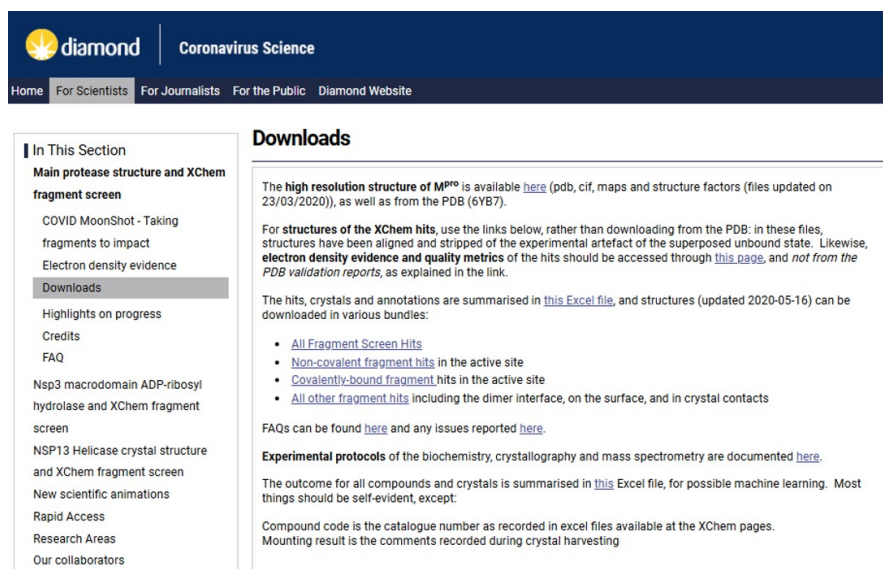


Figura 9.2. Captura de pantalla de la página principal del servidor Diamond Light Source. Se muestra la sección dedicada a la iniciativa internacional para el diseño de inhibidores contra la SCV2–Mpro, destacando el acceso a estructuras cristalográficas de complejos proteína–fragmento obtenidas mediante el experimento de alto rendimiento XChem. Esta base de datos proporciona información estructural esencial para el FBDD.

Una vez cargadas, se observará que las cadenas principales de las proteínas están perfectamente alineadas, como se muestra en la Figura 9.3A. Esto indica que las diferencias estructurales entre los archivos se deben principalmente a los distintos fragmentos no covalentes unidos al sitio activo de la enzima.

Para mejorar la visualización de los fragmentos unidos al sitio activo, es recomendable generar la isosuperficie de una sola estructura de la SCV2–Mpro, excluyendo la superficie correspondiente al ligando (ver Figura 9.3B). Esto permite observar con mayor claridad la forma de la cavidad del sitio activo y cómo los fragmentos se acomodan dentro de ella. Para ello, es necesario habilitar la línea de comandos desde el menú **Favorites>Command Line** e introducir los siguientes comandos:

surf #0 (crea la representación de isosuperficie para el modelo #0)
~surf ligand (elimina la isosuperficie correspondiente al ligando, permitiendo visualizar únicamente la envoltura proteica del sitio activo)

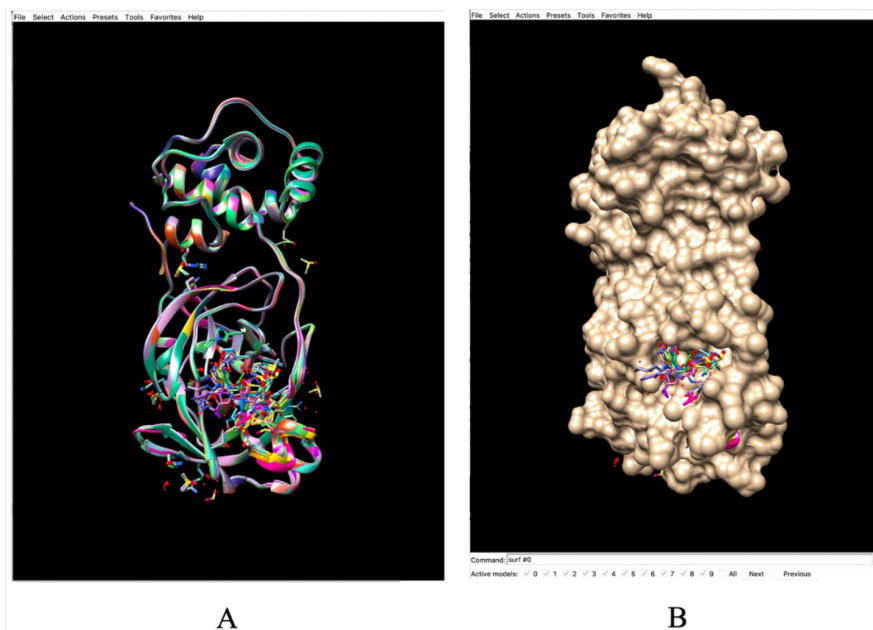


Figura 9.3. (A) Alineamiento estructural de múltiples complejos cristalográficos de la SCV2–Mpro, cada uno unido a un fragmento distinto en el sitio activo. Las estructuras fueron alineadas con UCSF Chimera version 1.18, evidenciando la conservación del plegamiento global. (B) Representación de la isosuperficie de una única SCV2–Mpro, generada con UCSF Chimera version 1.18, excluyendo la superficie del ligando. Se observa la cavidad catalítica donde se alojan los fragmentos, resaltando la región accesible para el diseño racional de inhibidores.

Es recomendable girar las moléculas en pantalla para explorar la topología del sitio activo. El **Model Panel** ofrece una forma eficiente de manipular los modelos: las columnas etiquetadas como **A (Active)** y **S (Show)** controlan la activación para análisis y la visibilidad, respectivamente (ver Figura 9.4).

Se recomienda verificar que todas las estructuras estén activas (columna **A** marcada), para conservarlas como parte del sistema, y utilizar la columna **S** para alternar su visibilidad según sea necesario.

Para enfocar el análisis en un fragmento específico, se deben seleccionar todas las estructuras excepto la de interés mediante **Shift + clic** en el **Model Panel** (excluyendo las isosuperficies), y utilizar la opción **hide** para ocultarlas de manera temporal. La isosuperficie del primer modelo no se ocultará automáticamente al desactivar la visibilidad del modelo asociado. Si se requiere visualizarla nuevamente, marcar manualmente la casilla **S** correspondiente en el **Model Panel**.

Ahora se deben explorar los fragmentos activando y desactivando sus estructuras individualmente. Observar cuidadosamente los grupos funcionales de cada

fragmento y las regiones de interacción dentro del sitio activo. Algunos complejos pueden contener moléculas de agua cristalográficas o cosolventes (por ejemplo, PEG o DMS). Estos pueden ocultarse fácilmente mediante **Select>Residue>HOH** seguido de **Actions>Atoms/Bonds>Hide**.

Es importante destacar que los fragmentos disponibles no forman enlaces covalentes con el residuo CYS145 ni con otros residuos del sitio activo. Para analizar el potencial de combinación entre fragmentos, se debe utilizar la herramienta **Distances** en **Tools>Structure Analysis**. Esto permitirá medir la distancia entre fragmentos contiguos y evaluar la factibilidad de unirlos mediante enlaces químicos. Con base en esta información, será posible proponer el diseño de un inhibidor tipo *drug-like* ensamblado racionalmente a partir de múltiples fragmentos.

El tamaño y características del inhibidor final dependerá de las distancias entre fragmentos, su complementariedad espacial dentro del sitio activo, y de las restricciones sintéticas consideradas en el diseño.

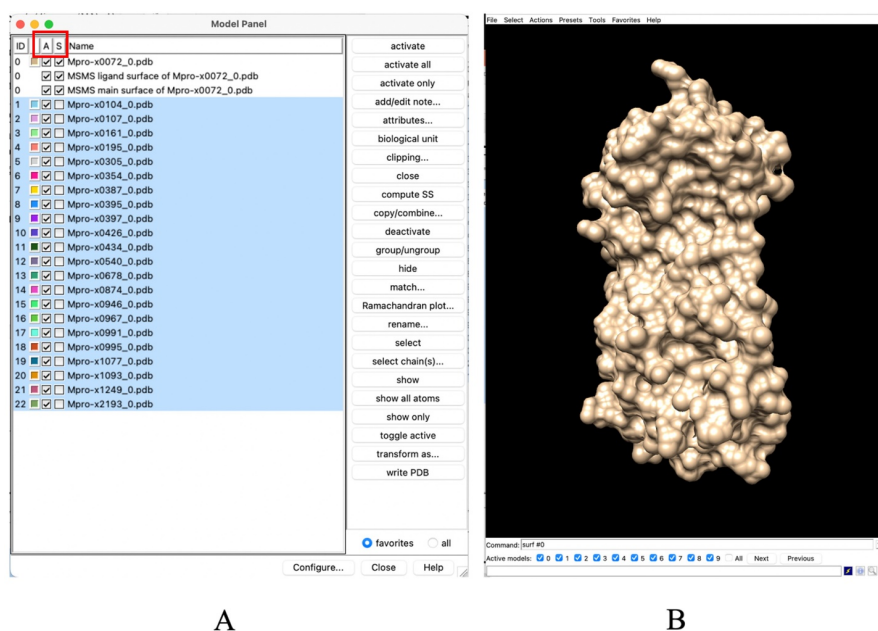


Figura 9.4. (A) Interfaz del **Model Panel** de UCSF Chimera version 1.18, mostrando las columnas **A (Active)** y **S (Shown)**, que permiten activar y visualizar selectivamente cada modelo cargado. Las casillas están remarcadas en rojo. (B) Representación de la isosuperficie generada para la SCV2–Mpro utilizando distintos complejos estructurales, evidenciando la cavidad del sitio activo donde se alojan los fragmentos co-cristalizados.

Además, se puede utilizar la herramienta **Tools>Structure Analysis>FindHBonds** de UCSF Chimera para identificar enlaces de hidrógeno que se establecen entre cada fragmento y los residuos del sitio activo de la SCV2–Mpro. Esta información estructural permite evaluar las interacciones intermoleculares clave responsables del reconocimiento molecular y la afinidad de unión. En el contexto del Diseño Racional

de Fármacos, es importante integrar criterios de química orgánica tales como la flexibilidad conformacional, estabilidad química, reactividad funcional de los grupos presentes y la posibilidad de establecer interacciones específicas (enlaces de hidrógeno, interacciones π - π , de *van der Waals*, entre otras).

Una propuesta de inhibidor diseñado surge de la cercanía entre los fragmentos #15 y #16, mostrados en la Figura 9.5. En dicha propuesta, el oxígeno del hidroxilo de la bencensulfonamida y el carbono alfa del grupo N-hidroxipropilo del fragmento #16, que se encuentran a una distancia de 1.3 Å, deben enlazarse covalentemente.

Una posible ruta de síntesis del derivado de bencensulfonamida se presenta en la Figura 9.6.

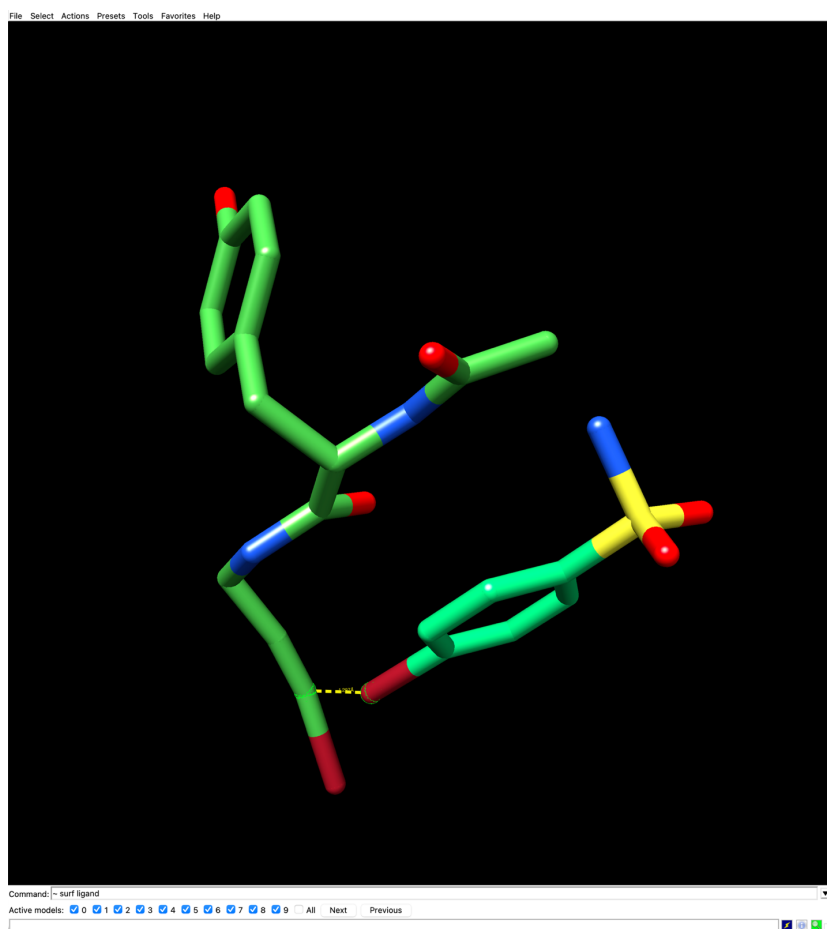


Figura 9.5. Los fragmentos #15 y #16 están dispuestos cercanamente en el sitio activo de la SCV2-Mpro. El oxígeno del hidroxilo de la bencensulfonamida y el carbono alfa del grupo N-hidroxipropilo del fragmento #16 se encuentran a una distancia de 1.3 Å.

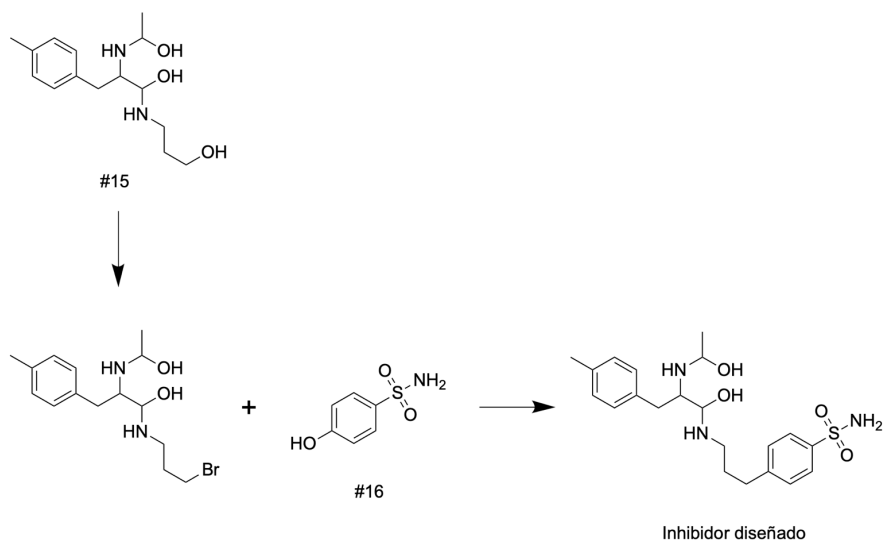


Figura 9.6. Propuesta de síntesis de un inhibidor diseñado a partir de los fragmentos #15 y #16. Para que se forme un enlace O–C entre el oxígeno del hidroxilo de la bencensulfonamida y el carbono alfa del grupo N-hidroxipropilo, debe activarse químicamente el grupo hidroxipropilo, por ejemplo, convirtiéndolo en un derivado como el N-bromopropilo. Luego, a través de una sustitución nucleofílica (S_N2) podría obtenerse el inhibidor diseñado.

Actividad 9.5. Con base en los fragmentos analizados, diseñar dos ligandos distintos uniendo covalentemente fragmentos que ocupen bolsillos diferentes del sitio activo. Dibujar sus estructuras químicas utilizando algún editor especializado, como *ChemDraw* o *ChemSketch*, y justificar el diseño de cada ligando considerando aspectos como:

- Compatibilidad espacial y orientación dentro del sitio activo
- Potencial para establecer múltiples interacciones favorables (enlaces de hidrógeno, contactos hidrofóbicos, interacciones iónicas, entre otros)
- Preservación de grupos funcionales clave de los fragmentos originales
- Factibilidad sintética del compuesto propuesto (tamaño, complejidad y tipo de enlaces formados)

Estas propuestas constituyen modelos teóricos de inhibidores dirigidos contra la SCV2–Mpro, susceptibles de ser evaluados mediante acoplamiento molecular para estimar su afinidad y especificidad.



Bibliografia

Brandt G. S. & W. R. P. Novak. (2021). SARS-CoV-2 virtual biochemistry labs on bioinformatics and drug design. *Biochemistry and Molecular Biology Education*, 49:26-28.

Bekker, H., H. J. C. Berendsen., E. J. Dijkstra., S. Achterop., R. van Drunen., D. van der Spoel., A. Sijbers., H. Keegstra., *et al.*, (1993). GROMACS: A parallel computer for molecular dynamics simulations; pp. 252–256 in *Physics computing* 92. Edited by R.A. de Groot and J. Nadrchal. World Scientific, Singapore.

Humphrey, W., Dalke, A. and Schulten, K.(1996). *VMD - Visual Molecular Dynamics*. *Journal of Molecular Graphics and Modelling* 14: 33-38.

O'Boyle, N. M., M. Banck., C. A. James., C. Morley., T. Vandermeersch., G. R. Hutchison, (2011). Open Babel: An open chemical toolbox. *Journal of Cheminformatics* (7)3: 1-14.

Krojer, T., Talon, R., Pearce, N., Collins, P., Douangamath, A., Brandao-Neto, J., Dias, A., Marsden B and von Delft, F. (2017). *The XChem Explorer graphical workflow tool for routine or large-scale protein-ligand structure determination*. *Acta Crystallographica Section D* 73: 267-278.



Práctica 10. Predicción de biodisponibilidad y propiedades farmacológicas de ligandos

Objetivos

- Analizar y comparar propiedades fisicoquímicas, farmacocinéticas y toxicológicas entre un compuesto candidato y un fármaco aprobado, empleando herramientas *in silico*.
- Interpretar con bases científicas los parámetros clave que afectan la biodisponibilidad, la permeabilidad, el metabolismo hepático y la toxicidad de moléculas bioactivas.
- Establecer criterios técnicos fundamentados para priorizar de candidatos a fármaco, mediante la integración de modelos predictivos computacionales y perfiles ADMET.

Para que una molécula actúe eficazmente como fármaco, debe ser capaz de alcanzar su diana terapéutica en concentraciones adecuadas, mantener su actividad biológica durante un periodo de tiempo suficiente y minimizar efectos adversos, incluyendo la toxicidad. En este contexto, la evaluación de las propiedades de absorción, distribución, metabolismo, excreción y toxicidad (ADMET) resulta esencial en las etapas de descubrimiento, optimización y validación de candidatos a fármaco. Se estima que hasta un 40 % de los compuestos candidatos son descartados durante fases avanzadas del desarrollo debido a propiedades ADMET desfavorables.

Aunque tradicionalmente estas propiedades se han determinado mediante ensayos *in vitro* e *in vivo*, dichos enfoques presentan limitaciones, tales como: tiempos prolongados de evaluación (superiores a los 10 años en algunos casos), elevados costos económicos y restricciones éticas derivadas del uso de modelos animales. Estas limitaciones han impulsado el desarrollo de herramientas *in silico*, que permiten predecir propiedades ADMET con mayor rapidez, menor costo y sin necesidad de realizar ensayos experimentales preliminares.

En las últimas décadas, el crecimiento de las bases de datos experimentales y los avances en inteligencia artificial (IA) han facilitado la construcción de modelos predictivos robustos, ubicados en diversas plataformas web especializadas. Algunas de las más utilizadas son:

- *SwissADME* (www.swissadme.ch): predicción de propiedades fisicoquímicas, parámetros farmacocinéticos y evaluación de cumplimiento de reglas de *drug-likeness*
- *admetSAR 3.0* (<http://lmmd.ecust.edu.cn/admetSar3/>): predicción de perfiles ADMET a partir de descriptores moleculares y aprendizaje automático
- *ADMETlab 3.0* (<https://admetlab3.scbdd.com/>): plataforma integral para la evaluación de perfiles ADMET y toxicidad
- *ADMETboost* (<https://ai-druglab.smu.edu/admet/>): modelo avanzado basado en algoritmos de *boosting* para predicción de propiedades ADMET multiclase

- *InterpretableADMET* (<http://cadd.pharmacy.nankai.edu.cn/interpretableadmet/>): herramienta enfocada en interpretación de propiedades ADMET a través de modelos explicables
- ProTox-3.0 (<https://tox.charite.de/protox3>): predicción de toxicidad aguda, hepática, carcinogenicidad, mutagenicidad y otros riesgos toxicológicos

Estas plataformas permiten estimar de forma automatizada características clave como: la solubilidad, permeabilidad intestinal, capacidad de atravesar barreras biológicas (como la barrera hematoencefálica), metabolismo hepático (predicción de isoenzimas del citocromo P450), vida media plasmática, potencial de inhibición de transportadores y riesgos de toxicidad a distintos niveles.

El uso estratégico de estas herramientas durante el proceso de Diseño Racional de Fármacos facilita la identificación de los candidatos con las mejores perspectivas de éxito terapéutico, y reduce significativamente los costos y tiempos asociados al desarrollo de nuevos fármacos.

El presente taller tiene como finalidad evaluar comparativamente el perfil ADMET de un compuesto candidato frente a un fármaco aprobado, con el objetivo de identificar las propiedades que influyen en su viabilidad como agente terapéutico. Para ello, se utilizarán herramientas computacionales de acceso libre ampliamente validadas, como *SwissADME*, *ADMETlab 3.0* y *ProTox-3.0*, reconocidas por su precisión y fiabilidad en la predicción de parámetros farmacocinéticos y toxicológicos. Asimismo, el ejercicio busca fortalecer la comprensión de los criterios esenciales que deben considerarse durante el Diseño Racional de Fármacos y en la priorización de candidatos para ensayos experimentales posteriores.

10.1 Predicción de características con *SwissADME*

Para realizar el análisis de propiedades fisicoquímicas y farmacológicas y de *drug-likeness* de un compuesto o fármaco de interés, es necesario contar con su estructura química en formato *SMILES* (Simplified Molecular Input Line Entry System) o en representación bidimensional. A modo de ejemplo, se empleará N-(o-nitrofenil)-maleimida, y se empleará el servidor *PubChem* para obtener dicha información.

Como ejemplo, se utilizará el compuesto N-(o-nitrofenil)maleimida, y se empleará el servidor *PubChem* para obtener su estructura. Para obtener el código *SMILES* debes:

- I. Acceder al sitio web de **PubChem** (<https://pubchem.ncbi.nlm.nih.gov>)
- II. En el campo de búsqueda, escribir **Maleimide, N-(o-nitrophenyl)** y hacer clic en **Enter** o en el ícono de la lupa
- III. Esperar a que aparezca la lista de resultados relacionados con el término ingresado (ver Figura 10.1)
- IV. Hacer clic en el resultado con el nombre exacto **Maleimide, N-(o-Nitrophenyl)** (**CID: 18097**) para abrir su ficha completa

- V. Dentro de la página del compuesto, localizar la sección **Structure** o **Canonical SMILES**
- VI. Copiar el código *SMILES* mostrado (por ejemplo: O=[N+](O-)]C1=CC=CC=C1NC2=CC=CO2), que luego se utilizará como entrada en el servidor **SwissADME** para predicciones moleculares

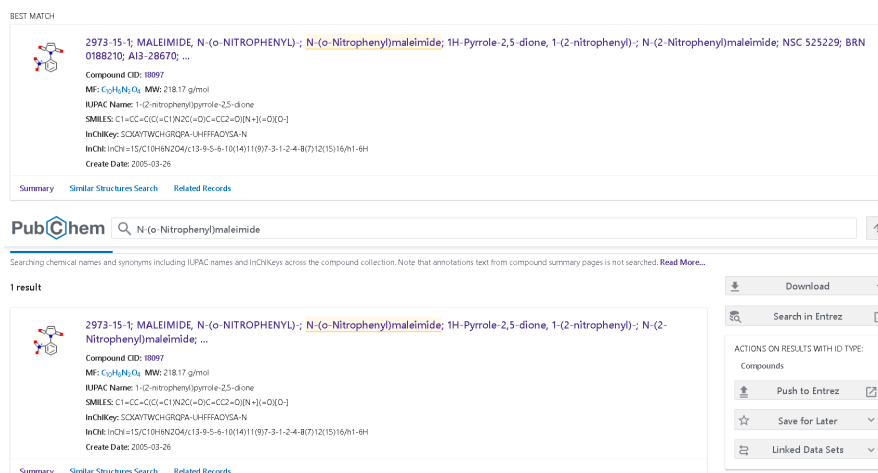


Figura 10.1. Captura de pantalla de la página de resultados obtenida tras buscar el término *Maleimide, N-(o-nitrophenyl)* en el servidor *PubChem*. Se observa la entrada correspondiente al N-(o-nitrofenil) maleimida (CID: 18097), desde la cual puede accederse al código *SMILES* y a la información estructural y fisicoquímica de la molécula.

10.2 Predicción de propiedades fisicoquímicas y farmacocinéticas

Acceder al sitio web de *SwissADME* (<http://www.swissadme.ch/>). Al hacerlo, se desplegará una ventana similar a la mostrada en la Figura 10.2, que permite ingresar la estructura del compuesto o fármaco de interés mediante dibujo o utilizando su representación *SMILES*. Para este ejercicio, el código *SMILES* de la molécula debe colocarse en el recuadro titulado **Enter a list of SMILES here**, localizado en la parte derecha de la interfaz.

Una de las ventajas más relevantes de esta plataforma es su capacidad para procesar múltiples entradas simultáneamente. Es posible introducir varios códigos *SMILES* en el mismo campo de texto, lo cual resulta especialmente útil para comparar perfiles farmacocinéticos entre diferentes candidatos a fármaco.

Una vez ingresadas las estructuras deseadas, se debe hacer clic en **Run**, ubicado en la parte inferior del recuadro, para iniciar el proceso de predicción de propiedades.

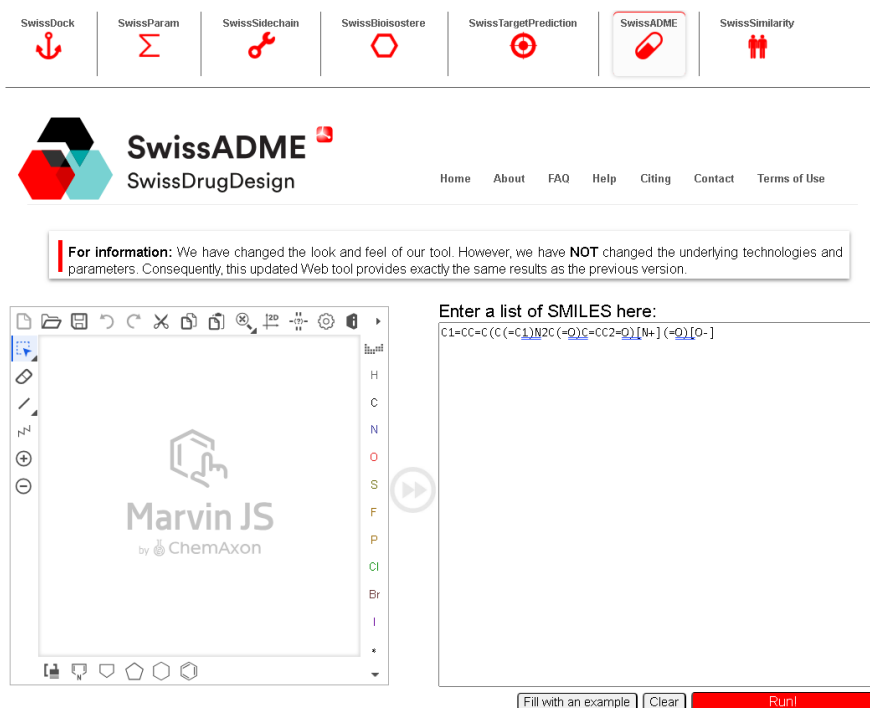


Figura 10.2. Captura de pantalla de la página de inicio del servidor *SwissADME* (<http://www.swissadme.ch/>). En el recuadro izquierdo se encuentra la herramienta para dibujar la estructura bidimensional del compuesto o fármaco de interés, mientras que en el recuadro derecho se pueden ingresar uno o varios códigos *SMILES* para su análisis.

10.3 Interpretación de resultados

Una vez ejecutado el análisis, el servidor desplegará una nueva ventana que mostrará la estructura química del compuesto junto con un conjunto detallado de propiedades moleculares, clasificadas en diferentes secciones (ver Figura 10.3):

- I. *Physicochemical Properties*: peso molecular, número de átomos de hidrógeno aceptores y donadores, número de enlaces rotables, área de superficie polar topológica (TPSA), entre otros
- II. *Lipophilicity*: predicciones de logP mediante distintos algoritmos (iLOGP, XLOGP3, WLOGP, MLOGP y SILICOS-IT)
- III. *Pharmacokinetics*: permeabilidad intestinal, penetración en el sistema nervioso central (SNC), interacción con glicoproteína P (P-gp), y posibles rutas metabólicas por enzimas del citocromo P450
- IV. Otras secciones como Water Solubility, Druglikeness, y Medicinal Chemistry

Esta información también se resume visualmente mediante dos herramientas gráficas de gran utilidad para la visualización de propiedades ADME:

Hide BOILED-Egg

Retrieve data:  Powered by Chemaxon

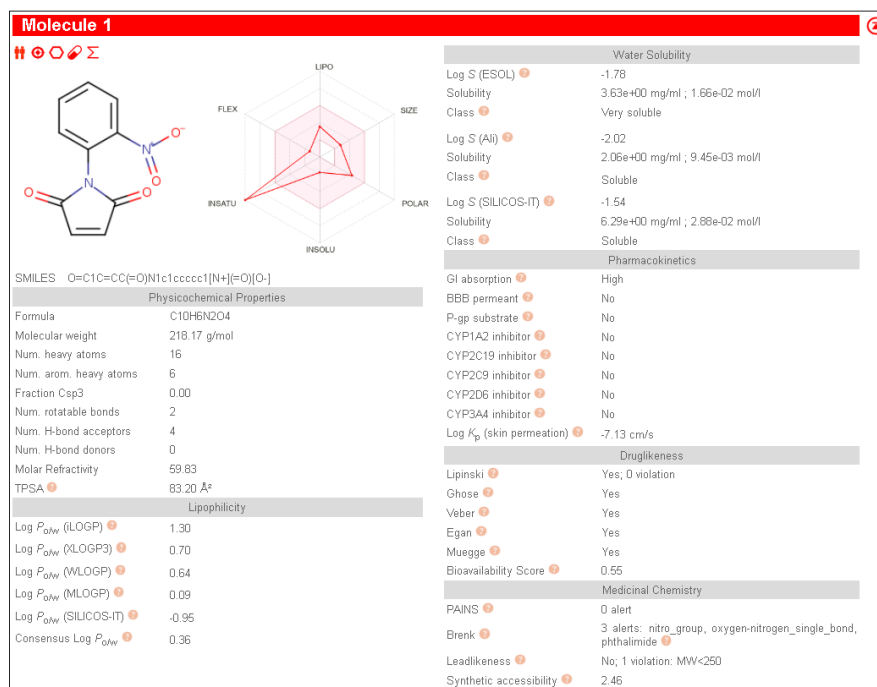
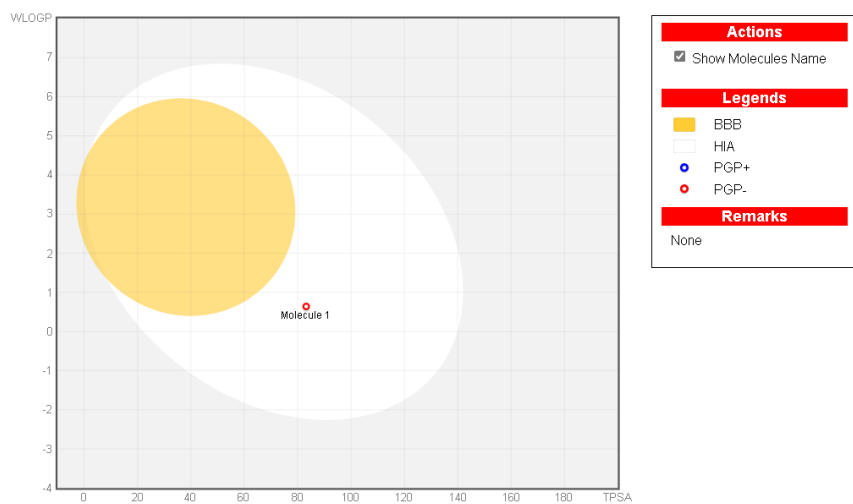


Figura 10.3. Captura de pantalla de los resultados obtenidos en el servidor *SwissADME* (<http://www.swissadme.ch/>) para el análisis de propiedades fisicoquímicas y farmacocinéticas. En la parte superior se muestra el esquema *BOILED-Egg*, que predice la HIA y la permeabilidad a través de la BBB según la relación entre la lipofiliadad (wLogP) y el área de superficie polar topológica (TPSA) de cada molécula. La zona blanca representa el espacio químico asociado a alta absorción intestinal, mientras que la zona amarilla indica probabilidad de penetración al sistema nervioso central. En la parte inferior, se presentan los valores calculados de propiedades fisicoquímicas, lipofiliadad, parámetros farmacocinéticos, reglas de semejanza con fármacos y química medicinal, organizados en diferentes secciones para cada compuesto analizado.

- 
- I. Radar de bioactividad: representa seis dimensiones relacionadas con la idoneidad de la molécula como fármaco: tamaño, lipofilidad, polaridad, solubilidad, flexibilidad y saturación. Cuanto más se acerque el perfil de la molécula al área rosada central, mayor será su probabilidad de éxito como candidato a fármaco
 - II. BOILED-Egg (Brain Or IntestinaL EstimateD permeation method): esta gráfica representa la posición del compuesto en función de dos parámetros clave: el coeficiente de partición octanol-agua estimado (wLogP) y el área de superficie polar topológica (TPSA). Es una herramienta predictiva útil para estimar la absorción intestinal y la capacidad de atravesar la barrera hematoencefálica (BBB). Su interpretación es la siguiente:
 - Zona blanca: alta probabilidad de absorción gastrointestinal (HIA)
 - Zona amarilla: alta probabilidad de atravesar la BBB
 - Puntos azules: moléculas con nula afinidad por la glicoproteína P (P-gp)
 - Puntos rojos: moléculas con alta afinidad por la P-gp, lo que puede reducir su biodisponibilidad al ser expulsadas activamente

Finalmente, en la parte inferior del recuadro de entrada de *SMILES*, se encuentra una opción para descargar los resultados en formato CSV (Comma-Separated Values). Este archivo incluye todos los parámetros predichos para cada compuesto analizado y es compatible con programas de hojas de cálculo como *Microsoft Excel* o *LibreOffice Calc*.

Se recomienda guardar el archivo en la carpeta de trabajo asignada al proyecto para facilitar su análisis posterior, permitir comparaciones entre múltiples compuestos y documentar los resultados obtenidos durante la sesión.

Actividad 10.1. Responder las siguientes preguntas con base en los resultados obtenidos:

¿Qué propiedades estimadas o predichas se observan en los resultados de la herramienta empleada?

¿Cuáles propiedades fisicoquímicas o farmacocinéticas son las más importantes para que un compuesto pueda ser aprobado como medicamento?

En el esquema de radar, investigar el significado del área coloreada en rosa y explicar qué representa en términos de *drug-likeness*.

En el esquema BOILED-Egg, ¿qué significan las áreas de color amarillo y blanco?

¿Cuál es la importancia de la relación wLogP y TPSA, y qué información aporta sobre la absorción o permeabilidad del compuesto?

10.4 Predicción de características con *ADMETlab 3.0*

La predicción de las propiedades ADMET de un compuesto o fármaco, se realizará mediante la herramienta *ADMETlab 3.0*, la cual permite evaluar eficientemente

parámetros clave como absorción, distribución, metabolismo, excreción y toxicidad. El procedimiento es el siguiente:

- I. Ingresar al servidor *ADMETlab 3.0* y localizar el botón **GET STARTED** en la página principal. Hacer clic en él
- II. El servidor desplegará una nueva ventana (ver Figura 10.4) en la que se debe pegar el código *SMILES* del compuesto de interés en el recuadro correspondiente
- III. Una vez ingresado el código, presionar **Submit** situado en la parte inferior

El servidor desplegará una nueva ventana que mostrará los resultados del análisis, organizados en secciones como *Absorción*, *Distribución*, *Metabolismo*, *Toxicidad* y otras propiedades relevantes (ver Figura 10.5).

De forma similar a *SwissADME*, *ADMETlab 3.0* también genera un diagrama tipo radar que permite visualizar gráficamente las propiedades fisicoquímicas del compuesto (ver Figura 10.6). Para acceder a esta gráfica, hacer clic en el botón **Radar View**.

En dicha gráfica, la línea amarilla representa el perfil de propiedades del compuesto evaluado, mientras que el área azul corresponde a los valores promedio obtenidos experimentalmente para fármacos ya aprobados. Esta comparación facilita la evaluación visual del desempeño del compuesto frente a estándares farmacológicos conocidos.

Finalmente, para conservar los resultados, hacer clic en la opción **Download as CSV**, ubicada junto al botón **Radar View**, y guardar el archivo en la carpeta de trabajo para análisis posterior.

Figura 10.4. Captura de pantalla de la interfaz principal del servidor *ADMETlab 3.0* (<https://admetlab3.scbdd.com/>). En el recuadro izquierdo **Enter a single SMILES string**, el usuario puede ingresar directamente el código *SMILES* del compuesto de interés. En el recuadro derecho **Draw a molecule**, se ofrece la opción de dibujar la estructura bidimensional de la molécula para su análisis computacional.

Actividad 10.2. Comparar los valores obtenidos en *ADMETlab 3.0* con los resultados proporcionados por *SwissADME*, y analizar las coincidencias o discrepancias observadas entre ambas herramientas.

O=C1C=CC(=O)N1c2ccccc2[N+](=O)[O-]

[Radar view](#) [Download as CSV](#) [Download as PDF](#)

STRUCTURE		ABSORPTION			TOXICITY		
		Caco-2 Permeability	-4.571	●	HERG Blockers	0.049	●
		MDCK Permeability	0.0	●	HERG Blockers (10um)	0.666	●
		PAMPA	+++	●	DILI	0.912	●
		Pgp inhibitor	---	●	AMES Toxicity	0.839	●
		Pgp substrate	---	●	Rat Oral Acute Toxicity	0.746	●
PHYSICOCHEMICAL PROPERTY		HIA	---	●	FDAMDD	0.891	●
Molecular Weight (MW)	218.03	F20%	---	●	Skin Sensitization	0.996	●
Volume	203.103	F30%	---	●	Carcinogenicity	0.852	●
Density	1.073	F50%	---	●	Eye Corrosion	0.999	●
nHA	6.0	DISTRIBUTION			Eye Irritation	1.0	●
nHD	0.0	PPB	86.1%	●	Respiratory	0.885	●
nRot	2.0	VDss	0.607	●	Human Hepatotoxicity	0.635	●
nRing	2.0	BBB	++	●	Drug-induced Nephrotoxicity	0.026	●
MaxRing	6.0	Fu	9.8%	●	Drug-induced Neurotoxicity	0.649	●
nHet	6.0	OATPIB1 inhibitor	--	●	Ototoxicity	0.063	●
fChar	0.0	OATPIB3 inhibitor	++	●	Hematotoxicity	0.796	●
nRig	14.0	BCRP inhibitor	---	●	Genotoxicity	1.0	●
Flexibility	0.143	MRPI inhibitor	+++	●	RPMT-8226 Immunotoxicity	0.015	●
Stereo Centers	0.0	BSEP inhibitor	+++	●	A549 Cytotoxicity	0.005	●
TPSA	80.52	METABOLISM			Hek293 Cytotoxicity	0.774	●
logS	-2.695	CYP1A2 inhibitor	++	●	BCF	1.154	●
logP	1.035	CYP1A2 substrate	+	●	IGC50	4.206	●
logD7.4	1.319	CYP2C19 inhibitor	+	●	LC50DM	5.031	●
pKa (Acid)	7.126	CYP2C19 substrate	---	●	LC50FM	4.703	●
pKa (Base)	0.558	CYP2C9 inhibitor	---	●	TOX21 PATHWAY		
Melting point	110.928	CYP2C9 substrate	---	●	NR-Ahr	---	●
Boiling point	312.849	CYP2D6 inhibitor	-	●	NR-AR	---	●
MEDICINAL CHEMISTRY		CYP2D6 substrate	---	●	NR-AR-LBD	---	●
QED	0.421	CYP3A4 inhibitor	-	●	NR-Aromatase	---	●
SAscore	Easy	CYP3A4 substrate	---	●	NR-ER	---	●

Figura 10.5. Captura de pantalla de los resultados obtenidos en el servidor *ADMETlab 3.0* (<https://admetlab3.scbdd.com/server/evaluationCal>). La ventana presenta la estructura bidimensional del compuesto analizado, junto con un resumen tabular de parámetros relacionados con absorción, distribución, metabolismo, excreción y toxicidad, además de otras propiedades relevantes para el Diseño Racional de Fármacos.

10.5 Predicción de toxicidad con *ProTox-3*

La evaluación de la toxicidad de un compuesto es un paso fundamental en el desarrollo de fármacos, ya que se busca minimizar los efectos adversos en la salud del paciente. Este proceso implica considerar factores bioquímicos, celulares y genéticos, lo que lo convierte en una tarea compleja. Sin embargo, gracias a la información disponible de fármacos ya conocidos, es posible realizar predicciones

utilizando modelos de interpolación que comparan las características de la molécula en estudio con aquellas de compuestos previamente estudiados. En este contexto, *ProTox-3* constituye una herramienta útil para realizar estas predicciones.

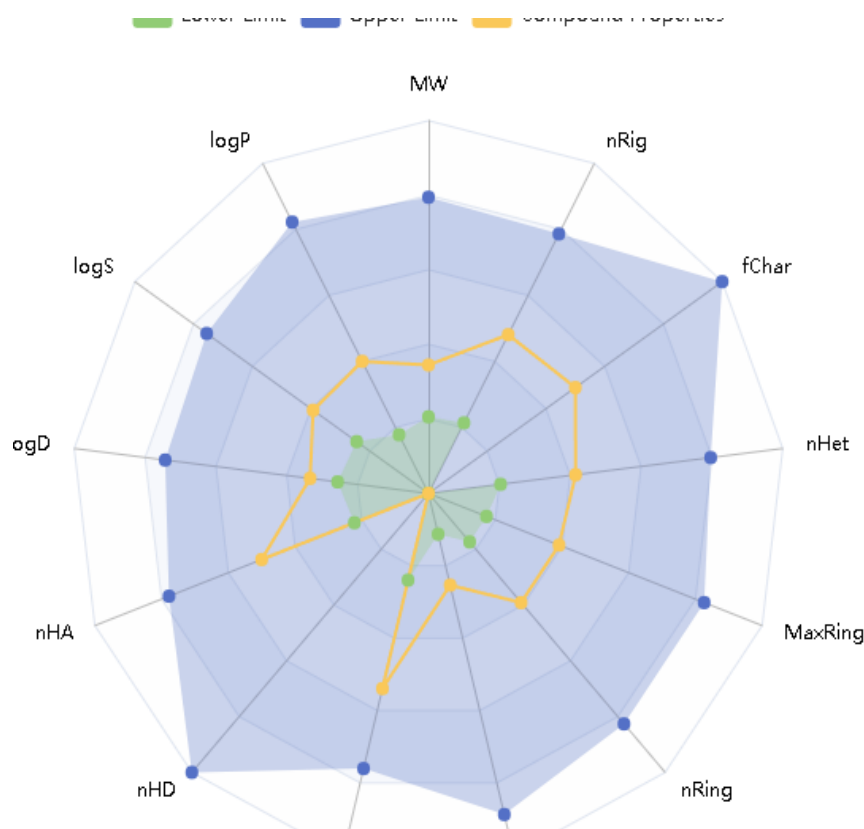


Figura 10.6. Captura de pantalla del diagrama tipo radar generado por *ADMETlab 3.0* (<https://admetlab3.scbdd.com/server/evaluationCal>), que resume las propiedades fisicoquímicas del compuesto N-(o-nitrofenil) maleimida. Se incluyen parámetros clave como el número de aceptores y donadores de enlaces de hidrógeno (nHA y nHD), número de enlaces rotables (nRot), área de superficie polar topológica (TPSA), número de anillos (nRing), heteroátomos (nHet), peso molecular (MW), así como medidas de lipofilicidad (logP), solubilidad (logS) y disociación (logD). La línea amarilla representa los valores calculados para el compuesto, mientras que el área azul delimita los rangos típicos de fármacos con propiedades aceptables.

- I. Acceder a *ProTox III* en <https://tox.charite.de/protox3/>
- II. En la sección **Tox Prediction**, introducir el código *SMILES* en el recuadro **Canonical SMILES** (ver Figura 10.7)
- III. Hacer clic en *SMILES* una vez ingresado el código
- IV. Activar únicamente las opciones **Organ Toxicity** y **Toxicity End Points** (recomendado para fines prácticos)
- V. Hacer clic en **Start Tox Prediction**
- VI. Activar otras opciones según sea necesario para tu caso específico (opcional)

- VII. En la ventana de resultados, revisar las dos secciones principales: **Oral Toxicity Prediction Results for Input Compound** y **Toxicity Model Report** (ver Figura 10.8)
- VIII. Debajo de estas secciones, usar las opciones disponibles para guardar los resultados en el formato deseado (Excel, CSV o PDF)

Tox-Prediction

Here you can input a compound via pubchem search, smiles string or drawing:

Pubchem-Name: search e.g. Tamoxifen Tolcazone Vorinostat Troglitazone Aspirin

Canonical Smiles: smiles e.g. CC(C)=C(C1=CC=CC=C1)C2=CC=C(C(=C2)OCN(C(C)C)C3=CC=CC=C3)

Selected molecule : O=C1C=CC(=O)N1c1cccc1[N+](=O)[O-]

ChemDoodle®

To load a molecule click . The deletes clicked atoms and deletes everything on the canvas.

Please select any additional models to predict:

Acute Toxicity and binding to 16 toxicity targets is always computed, further models can take ~10s time each

Organ Toxicity ☒ Hepatotoxicity ☒ Neurotoxicity ☒ Nephrotoxicity ☒ Respiratory toxicity ☐ Cardiotoxicity Toxicity end points ☒ Carcinogenicity ☒ Immunotoxicity ☒ Mutagenicity ☒ Cytotoxicity ☐ BBB-barrier ☐ Ecotoxicity ☐ Clinical toxicity ☐ Nutritional toxicity Tox21 Nuclear receptor signalling pathways ☐ Aryl hydrocarbon Receptor (AhR) ☐ Androgen Receptor (AR) ☐ Androgen Receptor Ligand Binding Domain (AR-LBD) ☐ Aromatase ☐ Estrogen Receptor Alpha (ER) ☐ Estrogen Receptor Ligand Binding Domain (ER-LBD) ☐ Peroxisome Proliferator Activated Receptor Gamma (PPAR-Gamma) Tox21 Stress response pathways ☐ Nuclear factor (erythroid-derived 2)-like 2/antioxidant responsive element (nr2/ARE) ☐ Heat shock factor response element (HSE) ☐ Mitochondrial Membrane Potential (MMP) ☐ Phosphoprotein (Tumor Suppressor) p53 ☐ ATPase family AAA domain containing protein 5 (ATAD5)	Molecular Initiating Events ☐ Thyroid hormone receptor alpha (THRo) ☐ Thyroid hormone receptor beta (THRB) ☐ Transthyretin (TTR) ☐ Ryanodine receptor (RYR) ☐ GABA receptor (GABAR) ☐ Glutamate N-methyl-D-aspartate receptor (NMDAR) ☐ alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionate receptor (AMPA) ☐ Kainate receptor (KAR) ☐ Acetylcholinesterase (AChE) ☐ Constitutive androstane receptor (CAR) ☐ Pregnane X receptor (PXR) ☐ NADH-quinone oxidoreductase (NADHox) ☐ Voltage gated sodium channel (VGSC) ☐ Na+/K+ symporter (NIS) Metabolism ☐ Cytochrome CYP1A2 ☐ Cytochrome CYP2C19 ☐ Cytochrome CYP2C9 ☐ Cytochrome CYP2D6 ☐ Cytochrome CYP3A4 ☐ Cytochrome CYP2E1 All None
--	--

Figura 10.7. Captura de pantalla de la sección Tox-Prediction del servidor web *ProTox-3.0* (https://tox.charite.de/protox3/index.php?site=compound_input) para la predicción de efectos toxicológicos. La interfaz permite evaluar distintos tipos de toxicidad, incluyendo toxicidad en órganos como hígado, riñones y sistema nervioso, así como propiedades mutagénicas, carcinogénicas e inmunotóxicas. También ofrece análisis de interacciones con receptores biológicos relevantes en farmacología.

Para este caso, descargar los resultados en un archivo Excel haciendo clic en la opción correspondiente. Además, se mostrará una gráfica de radar que resume visualmente la toxicidad (ver Figura 10.9), indicando la confianza de los resultados positivos de



toxicidad en comparación con el promedio de su clase. Esta gráfica es útil para interpretar rápidamente los datos obtenidos.

Hasta este punto, se han identificado características generales relacionadas con los parámetros ADMET. No obstante, en el contexto de una investigación exhaustiva, es fundamental profundizar en cada uno de estos aspectos, ya que determinan directamente el comportamiento de un compuesto dentro del organismo.

La absorción de un fármaco comprende su paso desde el sitio de administración (por ejemplo, por vía oral) hasta su ingreso al sistema circulatorio. Posteriormente, la distribución implica su transporte hacia los distintos compartimentos del cuerpo. Luego, el compuesto es sometido a procesos de metabolismo, principalmente hepáticos, que transforman su estructura para facilitar su excreción. Debido a la complejidad de estas etapas, resulta esencial realizar cálculos detallados para predecir con precisión el destino y comportamiento del fármaco en el organismo.

El uso de múltiples herramientas computacionales para estimar estos parámetros ofrece un enfoque robusto y complementario. Comparar los resultados generados por distintos programas permite establecer un consenso más confiable y tomar decisiones informadas a través de un análisis crítico. Esto cobra especial relevancia en la evaluación de la toxicidad, una de las etapas más sensibles del desarrollo farmacológico, ya que muchos compuestos no son completamente específicos y pueden interactuar con blancos biológicos no deseados, generando efectos adversos. Estos efectos pueden ir desde daños localizados en órganos específicos hasta complicaciones sistémicas severas.

Por tanto, resulta indispensable estudiar detalladamente cada etapa del metabolismo de fármacos, también conocido como metabolismo de xenobióticos, a nivel celular y molecular. Esta aproximación permite evaluar el grado aceptable de toxicidad de un compuesto, considerando que existen distintos tipos de toxicidad (aguda, subcrónica, crónica, genotóxica, inmunotóxica, entre otras), cada una con implicaciones particulares en la seguridad del fármaco.

Actividad 10.3. Identificar los tipos de toxicidad reportada para el compuesto de interés.

¿Cuál es la importancia bioquímica de los datos presentados en el radar de toxicidad?

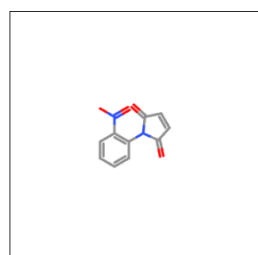
Enunciar la Regla de 5 de Lipinski.

¿Qué relación tienen estas reglas con las características de los fármacos aprobados?

¿Por qué algunos fármacos necesitan ser más hidrofóbicos que otros para ejercer su efecto terapéutico?

¿Qué es el CYP450 y qué relación tiene con las propiedades ADMET de un fármaco (metabolismo, interacciones medicamentosas, eliminación, etc.)?

Oral toxicity prediction results for input compound



Predicted LD50: 250mg/kg

Predicted Toxicity Class: 3

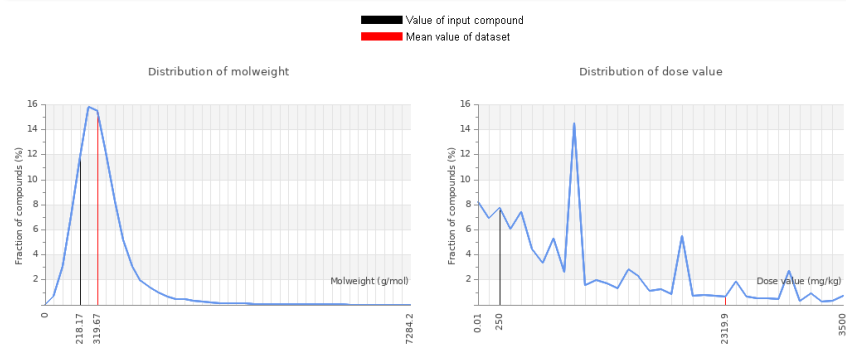
Average similarity: 62.65%

Prediction accuracy: 68.07%

Print Toxicity Report

Name	O=C1C=CC(=O)N1c1ccc(=O)[O-]
Molweight	218.17
Number of hydrogen bond acceptors	4
Number of hydrogen bond donors	0
Number of atoms	16
Number of bonds	17
Number of rotatable bonds	2
Molecular refractivity	59.83
Topological Polar Surface Area	83.2
octanol/water partition coefficient(logP)	1.61

Comparison of input compound with dataset compounds



Similar compounds

Toxicity Model Report

Classification	Target	Shorthand	Prediction	Probability
Organ toxicity	Hepatotoxicity	dili	Active	0.53
Organ toxicity	Neurotoxicity	neuro	Inactive	0.71
Organ toxicity	Nephrotoxicity	nephro	Inactive	0.64
Organ toxicity	Respiratory toxicity	respi	Active	0.67
Toxicity end points	Carcinogenicity	carcino	Active	0.67
Toxicity end points	Immunotoxicity	immuno	Inactive	0.96
Toxicity end points	Mutagenicity	mutagen	Active	0.94
Toxicity end points	Cytotoxicity	cyto	Inactive	0.64

Figura 10.8. Captura de pantalla de los resultados obtenidos en el servidor *ProTox-3* (https://tox.charite.de/protox3/index.php?site=compound_search_similarity), específicamente en las secciones **Organ Toxicity**, **Comparison of Input Compound** y **Toxicity Model Report**. La predicción de los efectos toxicológicos se presenta mediante un código de colores: rosa (baja probabilidad de toxicidad), verde (probabilidad intermedia) y rojo (alta probabilidad), lo que permite una interpretación visual rápida de los riesgos asociados al compuesto analizado.

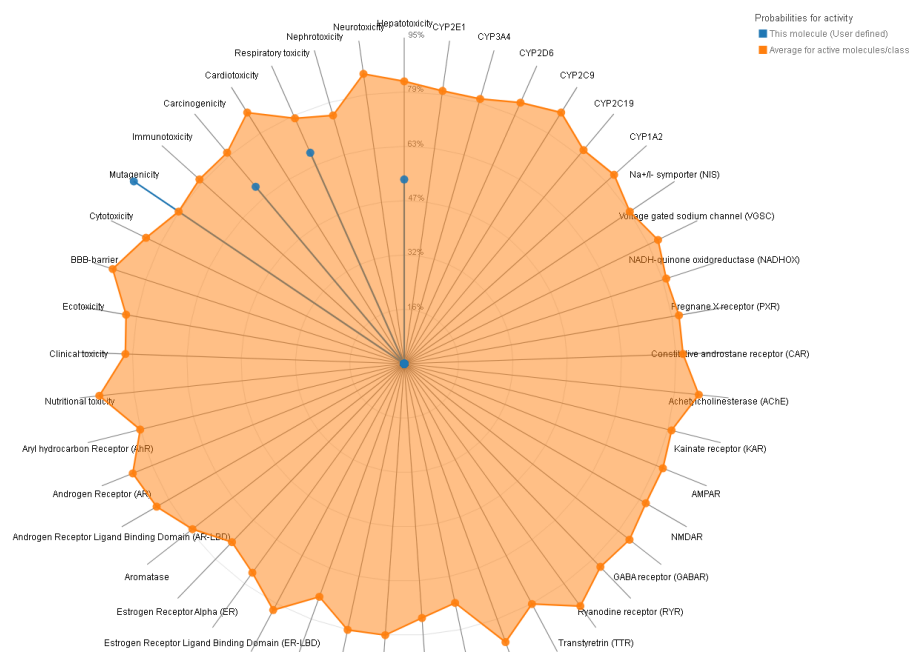


Figura 10.9. Captura de pantalla de la gráfica de radar de toxicidad generado por el servidor *ProTox-3* (https://tox.charite.de/protox3/index.php?site=compound_search_similarity). Se comparan múltiples parámetros toxicológicos del compuesto *N*-(*o*-nitrofenil) maleimida con datos de compuestos similares. Los puntos en azul representan las predicciones de toxicidad del compuesto analizado, incluyendo mutagenicidad, inmunotoxicidad, carcinogenicidad, y toxicidad hepática y respiratoria. El área sombreada en color naranja indica los valores promedio de estos parámetros en fármacos con características estructurales semejantes, lo que permite visualizar de forma comparativa el perfil toxicológico del compuesto.

Referencias

- Banerjee, P.,** Kemmler, E., Dunkel, M., & Preissner, R. (2024). ProTox 3.0: a webserver for the prediction of toxicity of chemicals. *Nucleic acids research*, 52(W1), W513–W520. <https://doi.org/10.1093/nar/gkae303>
- Daina, A.,** Michielin, O., & Zoete, V. (2017). SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific reports*, 7, 42717. <https://doi.org/10.1038/srep42717>.
- Fu, L.,** Shi, S., Yi, J., Wang, N., He, Y., Wu, Z., Peng, J., Deng, Y., Wang, W., Wu, C., Lyu, A., Zeng, X., Zhao, W., Hou, T., & Cao, D. (2024). ADMETlab 3.0: an updated comprehensive online ADMET prediction platform enhanced with broader coverage, improved performance, API functionality and decision support. *Nucleic acids research*, 52(W1), W422–W431. <https://doi.org/10.1093/nar/gkae236>



Apéndice A. Guía para la instalación de software para la realización del Taller de Bioquímica Computacional para el Análisis de Proteínas y Compuestos Bioactivos

Este apéndice proporciona una guía paso a paso para instalar el software necesario en sistemas Ubuntu. Se recomienda contar con conexión a internet estable y privilegios de administrador para ejecutar los comandos. Las instrucciones han sido verificadas para versiones recientes de los sistemas operativos Ubuntu 22.04.

1. UCSF Chimera version 1.18 o posterior

UCSF Chimera es una herramienta ampliamente utilizada en el ámbito científico para la manipulación y análisis estructural de biomoléculas. Su versatilidad y potencia la convierten en una opción preferida por investigadores y profesionales. Para garantizar un entorno óptimo de funcionamiento, uno de los sistemas operativos más utilizados es Ubuntu, una distribución de GNU/Linux que ofrece estabilidad, seguridad y compatibilidad con herramientas científicas.

En esta sección se detallará el procedimiento para instalar UCSF Chimera en Ubuntu, proporcionando una guía clara y precisa que facilite su implementación en entornos de trabajo o investigación.

Primero, es necesario ingresar al sitio oficial: <https://www.cgl.ucsf.edu/chimera/download.html>. Después seleccionar la versión para **Linux 64-bit** (por ejemplo: *chimera-1.18-linux_x86_64.bin*). Aparecerá otra ventana **Download UCSF Chimera**. Nos dirigimos a parte final del texto y aceptamos los términos dando clic en el botón **Accept**. Se descargará el archivo .bin (*chimera-1.18-linux_x86_64.bin*).

Navegar a la carpeta donde se descargó el archivo .bin, hacer clic derecho con el ratón sobre un espacio vacío y seleccionar **Abrir en una terminal**.

Antes de ejecutar el archivo es necesario dar permisos de ejecución. Para esto se usa el comando `chmod`:

```
$ chmod +x chimera-1.18-linux_x86_64.bin
```

Una vez hecho esto se puede proceder a ejecutar el instalador con:

```
$ sudo ./chimera-1.18-linux_x86_64.bin
```

El comando anterior le pedirá su contraseña, la cual deberá ingresar y dar **Enter**. No se preocupe si parece que no se escribe: esto es normal, ya que las contraseñas no se muestran en pantalla por motivos de seguridad.

No preguntará donde deseamos instalar el programa, Dejar la ruta por default: **opt/UCSF/Cmirea64-1-18**. Dar **Enter**.



El sistema preguntará si deseamos crear un icono (**install desktop menú?**). Escribir **yes** y dar **Enter**.

Nuevamente, el sistema preguntará si se desea crear un link para ejecutar el programa mediante comandos, y nos mostrará una lista con las rutas enumeradas como se muestra a continuación:

```
install symbolic link to chimera executable for command line use
in which directory?
0 -- no link
1 -- /usr/local/sbin/
2 -- /usr/local/bin
```

En este caso la mayoría de los programas son almacenados en **/usr/local/bin/**. Para elegir esta ruta escribir el número **2** y dar **Enter**.

Verificar la instalación escribiendo en terminal la palabra **chimera**. También podemos verificar mediante la búsqueda del icono creado. Para ello apretamos la tecla Super (la de Windows) y navegamos entre las aplicaciones.

2. AutoDock Vina (versión 1_1_2)

AutoDock Vina (Vina) es un programa de código abierto utilizado para el acoplamiento molecular, ampliamente reconocido por su velocidad y precisión en la predicción de interacciones entre moléculas (Proteína–Ligando). Vina se puede descargar desde <https://vina.scripps.edu/downloads/>. Se recomienda utilizar Google Chrome.

Para Ubuntu se debe seleccionar la versión para Linux 64-bit (autodock_vina_1_1_2_linux_x86.tgz). Después, ingresar a la carpeta donde fue descargado el paquete del programa. Para descomprimir el paquete se debe descomprimir el archivo ZIP con el siguiente comando:

```
$ unzip autodock_vina_1_1_2_linux_x86_64.zip
```

También se puede extraer el archivo haciendo clic derecho sobre el archivo, y buscar la opción como **Extraer aquí** o **Extraer en...**

Para facilitar el uso de Vina desde cualquier ubicación del sistema, es recomendable mover o copiar el archivo ejecutable **Vina** a una carpeta accesible, como **/usr/local/bin/**. Para esto podemos usar la instrucción mv (mover) o cp (copiar), según se prefiera. Por ejemplo:

```
$ sudo mv vina /usr/local/bin/
```

o

```
$ sudo cp vina /usr/local/bin/
```



El comando anterior solicitará su contraseña, la cual deberá ingresar y dar **Enter**. No preocuparse si parece que no se escribe: esto es normal, ya que las contraseñas no se muestran en pantalla por motivos de seguridad.

Esto permitirá ejecutar **Vina** directamente desde la terminal sin necesidad de especificar su ruta completa. Para hacerlo, simplemente se escribe el comando `vina` en la terminal y presionar **Enter**.

3. AutoDock Tools

AutoDock Tools (ADT), también conocido como MGLTools, es un conjunto de utilidades que facilita la preparación de archivos de entrada para programas de docking molecular como AutoDock4 y Vina. ADT permite:

- Preparar ligandos y receptores en formato PDBQT
- Definir visualmente el espacio de búsqueda (grid box)
- Visualizar y analizar los resultados del acoplamiento molecular

El paquete de instalación está disponible en el sitio oficial: <https://ccsb.scripps.edu/mgltools/downloads/>. Se encuentra disponible para sistemas operativos Windows y Linux. Para usuarios de Ubuntu, se debe descargar el paquete correspondiente a Linux: `mgltools_Linux-x86_64_1.5.7.tar.gz`.

Después de guardar el archivo en una carpeta de elección se procede a descomprimir el paquete. Abrir una terminal y navegar a la carpeta donde se descargó el archivo para ejecutar el siguiente comando:

```
$ tar -xvzf mgltools_Linux-x86_64_1.5.7.tar.gz
```

También se puede extraer el archivo haciendo clic derecho sobre el archivo y buscar la opción **Extraer aquí** o **Extraer en...**

Para iniciar la instalación se debe entrar a la carpeta para ejecutar el script de instalación con el siguiente comando:

```
$ ./install.sh
```

Si el comando no funciona, intentar ejecutarlo con privilegios de superusuario anteponiendo `sudo`:

```
$ sudo ./install.sh
```

Esto instalará ADT, Python Molecular Viewer (PMV) y otros componentes necesarios. Después de la instalación, se puede iniciar ADT con:

```
$ ./adt
```

o

```
$ adt
```



Si el comando no funciona, verificar que sea la carpeta correcta donde se encuentra el archivo **adt**. Si el problema persiste, añadir la ruta del archivo **adt** a la variable de entorno **PATH** para poder ejecutarlo desde cualquier ubicación.

Por ejemplo:

```
$ export PATH=$PATH:/ruta/a/la/carpeta/adt
```

Nota. Requisitos adicionales

- Asegurarse de tener Python 2.5–2.7 instalado (ADT no es compatible con Python 3).
- Instalar dependencias como libX11, libGL, libGLU, y Tkinter si faltan.

4. VMD

VMD (Visual Molecular Dynamics) al igual que Pymol, UCSF Chimera y ADT es un programa de visualización y análisis molecular diseñado para sistemas biológicos como proteínas, ácidos nucleicos, ensamblajes de bicapas lipídicas, entre otros. Sin embargo, VMD se distingue por:

- Su capacidad para operar eficientemente con trayectorias de dinámica molecular de varios gigabytes
- Su interoperabilidad con una amplia variedad de paquetes de simulación molecular
- Capacidad para crear películas en VMD

VMD es compatible con las principales plataformas de cómputo (Windows, macOS y Linux) y puede obtenerse desde su página oficial: <https://www.ks.uiuc.edu/Research/vmd/>

Hacer clic en **Download VMD** para acceder a la sección de descargas. Es obligatorio registrarse para poder descargar el paquete del programa.

Elegir la versión para Linux (**vmd-1.9.4a51.bin.LINUX.opengl.tar.gz**).

Abrir una terminal y navegar a la carpeta donde se descargó el archivo. Luego, ejecutar el siguiente comando para descomprimirlo:

```
$ tar -xvzf vmd-1.9.4a51.bin.LINUX.opengl.tar.gz
```

Antes de proceder con la instalación, es necesario realizar una configuración previa. Para ello, acceder a la carpeta descomprimida (**vmd-1.9.4a51**) utilizando el siguiente comando:

```
$ cd vmd-1.9.4a51
```

Una vez dentro de la carpeta descomprimida, ejecutar el script de configuración con el siguiente comando:

```
$ ./configure LINUX
```

Después, ingresar a la carpeta **src** con el comando:



```
$ cd src
```

También se puede ingresar a la carpeta de manera gráfica con el ratón. Una vez dentro, ejecutar la instalación con:

```
$ sudo make install
```

Después de la instalación, se puede iniciar VMD escribiendo la palabra *vmd* en la terminal:

```
$ vmd
```

Si no se reconoce el comando, añadir la ruta de instalación a tu variable PATH. Por ejemplo:

```
$ export PATH=$PATH:/home/usuario/vmd-1.9.4a51/bin
```

Se puede agregar esta línea al final del archivo *~/.bashrc* para iniciar automáticamente cada vez que inicies sesión.

5. Open Babel

Open Babel es una herramienta de software libre que permite convertir estructuras químicas entre más de 110 formatos, preparar moléculas agregando hidrógenos o generando conformaciones 3D, analizar propiedades moleculares, y automatizar tareas mediante su interfaz de línea de comandos o API en C++ y Python. Es ampliamente utilizada en química computacional y diseño de fármacos, y se integra fácilmente con programas como AutoDock Vina, GROMACS, PyMOL y VMD.

Existen al menos tres formas de instalar Open Babel: la instalación “general” o “estándar”, la instalación de versiones más recientes, y la instalación gráfica (sin el uso de comandos). La elección dependerá del nivel de exigencia del usuario.

Opción 1: Instalación desde repositorios (Ubuntu/Debian)

Primero se deben actualizar paquetes del sistema con el comando:

```
$ sudo apt update
```

Se procede con la instalación con el comando:

```
$ sudo apt install openbabel
```

para verificar la instalación ejecutamos:

```
$ obabel -V
```

Esto debería mostrar la versión instalada.

Opción 2: Instalación desde código fuente (versión más reciente)

Para instalar dependencias, copiar y ejecutar los siguientes comandos siguiendo el mismo orden en cada paso:



```
$ sudo apt install build-essential cmake git \  
$ libcairo2-dev libeigen3-dev libffi-dev libfreetype6-dev \  
$ libpng-dev libpython3-dev libxml2-dev python3 python3-pip \  
$ python3-pyqt5 python3-pyqt5.qtsvg python3-setuptools qtbase5-  
dev zlib1g-dev
```

Antes de proceder con la instalación, es necesario clonar el repositorio de Open Babel y crear una carpeta de construcción. Para ello, ejecutar los siguientes comandos:

```
$ git clone https://github.com/openbabel/openbabel.git  
$ cd openbabel  
$ mkdir build  
$ cd build
```

Para proceder con el proceso, compilar e instalar el programa utilizando los siguientes comandos:

```
$ cmake ..  
$ make -j$(nproc)  
$ sudo make install  
$ vsudo ldconfig
```

Finalmente podemos ejecutar el programa:

```
$ obabel -V
```

Opción 3: Instalación en modo gráfico

En la mayoría de las distribuciones de GNU/Linux se incluye una tienda oficial de aplicaciones que permite instalar diversos programas. En el caso de Ubuntu con entorno gráfico GNOME, esta tienda se llama *Ubuntu Software*. Funciona como una interfaz visual para buscar, instalar, actualizar y eliminar software en el sistema, sin necesidad de utilizar comandos en la terminal.

Para este proceso, utilizar *Ubuntu Software*. Hacer clic en el ícono de *Ubuntu Software* en la barra lateral (parece una bolsa de compras, naranja). Si no se logra ver, presionar la tecla Super (Windows) y escribir **Ubuntu Software**.

En la parte superior, hacer clic en el ícono de lupa. Escribir el nombre del programa que se desea instalar (en este caso Open Babel).

Hacer clic en el resultado de interés y revisar la descripción, capturas de pantalla, calificaciones y permisos.

Hacer clic en el botón *Instalar* e ingresa su contraseña si el sistema lo solicita y esperar a que finalice la instalación.

Una vez instalado, hacer clic en **Abrir** desde Ubuntu Software. También se puede buscar en el menú de aplicaciones, donde aparecerá su ícono para iniciarlo fácilmente.



6. XMGrace

XMGrace es una herramienta para crear gráficos bidimensionales de datos científicos, disponible tanto en modo gráfico (GUI) como mediante comandos o scripts automatizados. Permite importar datos desde diversos formatos de archivo, incluyendo múltiples columnas de datos. Gracias a esta versatilidad, es uno de los programas preferidos para procesar y visualizar datos provenientes de simulaciones de dinámica molecular.

La opción más común (y rápida) para instalarlo es mediante el gestor de paquetes en sistemas Ubuntu/Debian, utilizando tres comandos sencillos.

El primer paso es abrir una terminal, ya sea presionando Ctrl + Alt + T o buscándola en el menú de aplicaciones. Luego, actualiza los repositorios y procede con la instalación ejecutando los siguientes comandos:

```
$ sudo apt update
$ sudo apt upgrade
$ sudo apt install grace
```

Verificar la instalación ejecutando el siguiente comando:

```
$ xmgrace
```

Debería abrirse la interfaz gráfica de *XMGrace*.

Notas finales

Antes de instalar cualquier programa, es buena práctica asegurarse de que el sistema esté actualizado. En este sentido, el comando `sudo apt update` actualiza la lista de paquetes disponibles desde los repositorios configurados en tu sistema. Este comando no instala ni modifica nada, simplemente se asegura de que el sistema “sepa” qué versiones están disponibles, como si se consultara un catálogo antes de comprar.

Este paso es necesario antes de instalar o actualizar cualquier paquete. Por otro lado, el comando `sudo apt upgrade` instala las versiones más recientes de los paquetes que ya tiene el sistema operativo, siempre que haya actualizaciones disponibles. Esta práctica suele resolver varios inconvenientes cuando se presentan errores al intentar instalar nuevos programas.

7. TTclust

TTclust es un programa diseñado para agrupar trayectorias de simulaciones de dinámica molecular. Permite identificar las distintas conformaciones que adopta una molécula durante la simulación, facilitando el análisis de grandes cantidades de datos. Requiere un archivo de trayectoria y un archivo de topología, compatibles con paquetes como AMBER, GROMACS, CHARMM, NAMD o trayectorias en formato PDB.



TTclust genera gráficas como dendrogramas para visualizar y analizar los clústeres obtenidos.

Existen al menos dos formas de instalar TTclust:

- Instalación con Conda

Primero se debe obtener Conda desde el sitio: <https://docs.conda.io/en/latest/miniconda.html> (usa Python 3). Después, se puede instalar TTclust con el comando:

```
$ conda install -c tubiana -c conda-forge ttclust
```

- Instalación con PIP

TTclust requiere numpy, una biblioteca de Python que permite trabajar de forma eficiente con matrices, vectores y operaciones numéricas. Esto la hace especialmente útil para manipular colecciones de coordenadas atómicas a lo largo del tiempo.

Primero se debe instalar numpy con pip:

```
$ pip install cython numpy
```

Posteriormente se procede a la instalación con:

```
$ pip install ttclust
```

Si no funciona, intenta ejecutar el comando con privilegios de superusuario usando:

```
$ sudo pip install ttclust
```

Nota. Se recomienda usar Conda, ya que con pip se compilará mdtraj, lo que aumenta el riesgo de fallos en la instalación (por ejemplo, por problemas con la biblioteca de C++ de Microsoft o cgg).

- Instalación desde el código fuente

Este proceso requiere clonar el repositorio de TTclust. Para ello, se usa la herramienta git con el siguiente comando:

```
$ git clone https://github.com/tubiana/ttclust
```

También es necesario instalar las dependencias:

```
$ sudo pip install -r requirements.txt
```

A continuación, se procede a crear un nuevo entorno virtual con Conda:

```
$ conda env create -f environment.yml
```

Finalmente, activamos el entorno creado con el siguiente comando:

```
$ conda activate ttclust
```

Los scripts de TTclust se pueden ejecutar con:

```
$ python {RUTA DE TRABAJO}/ttclust/ttclust.py
```



```
O
$ python {RUTA}/ttclust/ttclustGUI.py
```

Nota: en ocasiones, la instalación de mdtraj puede presentar dificultades. Si se utiliza pip, primero instalar manualmente cython con el siguiente comando:

```
$ sudo pip install cython
```

Si surgen problemas al instalar mdtraj, se puede hacer con Conda usando el siguiente comando:

```
$ conda install mdtraj
```

8. GROMACS

Como se mencionó anteriormente en la **Práctica 6. Generación de conformaciones de una proteína mediante simulación de dinámica molecular**, GROMACS es un software de código abierto diseñado para realizar simulaciones de dinámica molecular con alta precisión y rendimiento.

Para instalar GROMACS en Ubuntu, se puede optar por: 1) utilizar el gestor de paquetes del sistema para una instalación sencilla, o 2) compilarlo desde el código fuente si necesitas una versión personalizada con optimizaciones específicas. La elección dependerá de los requerimientos del usuario.

- Instalación utilizando el gestor de paquetes del sistema

Esta es la forma más rápida y sencilla de instalar GROMACS. El método utiliza una versión pre-compilada disponible en los repositorios de Ubuntu. Aunque la versión instalada puede no ser la más reciente, es estable y fácil de configurar. Sin embargo, resulta adecuada para uso básico y pruebas iniciales en la práctica. Por lo tanto, es suficiente para los objetivos de la actividad, en la que los tiempos de simulación son muy cortos.

Primero, es necesario actualizar la lista de paquetes con los siguientes comandos:

```
$ sudo apt update
$ sudo apt upgrade
```

Finalmente, para completar la instalación, se debe ejecutar el siguiente comando:

```
$ sudo apt install gromacs
```

Para conocer la versión instalada de GROMACS, ejecuta el siguiente comando en la terminal:

```
$ gmx --version
```

- Instalación desde el código fuente

Esta opción está dirigida al usuario que desea una versión más reciente o personalizada, como la compilación desde el código fuente.



Antes de proceder, es necesario instalar varios paquetes esenciales para compilar el software. Para ello, debe ejecutarse el siguiente comando:

```
$ sudo apt install build-essential cmake git wget
```

Proceder a descargar la versión más reciente del código fuente desde el sitio web oficial de GROMACS. Por ejemplo, para la versión 2025.3, sería:

```
$ wget https://ftp.gromacs.org/gromacs/gromacs-2025.3.tar.gz
```

Nos dirigimos a la carpeta donde se descargó el paquete y procedemos a extraer el archivo:

```
$ tar -xzf gromacs-2025.3.tar.gz
```

Para que GROMACS funcione correctamente, es necesario configurar su compilación y descargar los archivos requeridos para preparar el entorno. Para ello, se debe crear un directorio de compilación y configurar la construcción:

```
$ cd gromacs-2025.3
$ mkdir build
$ cd build
$ cmake .. -DGMX_BUILD_OWN_FFTW=ON -DREGRESSIONTEST_DOWNLOAD=ON
```

Finalmente, procedemos con la compilación e instalación:

```
$ make
$ sudo make install
```

En general, el tiempo de instalación puede variar desde unos minutos hasta más de media hora, dependiendo principalmente de las características del hardware de la computadora.

Para activar la instalación, se ejecuta el script de entorno para agregar los comandos de GROMACS a la variable PATH:

```
$ source /usr/local/gromacs/bin/GMXRC
```

Para comprobar que la instalación se ha realizado correctamente, ejecutar el siguiente comando:

```
$ gmx --version
```

Nota. Esta forma de instalación está configurada por defecto para utilizar el procesador, lo cual resulta ideal para fines prácticos. Sin embargo, cuando se trabajan trayectorias de tiempo relativamente largas (a partir de 10 nanosegundos) y repetitivas, lo más eficiente es aprovechar los núcleos CUDA de la GPU (Unidad de Procesamiento Gráfico) de la tarjeta gráfica. Usar GPU en GROMACS acelera los cálculos intensivos, mientras que la CPU se encarga de tareas generales. Lo óptimo es combinar ambos recursos para obtener el mejor rendimiento. Por ello, se recomienda utilizar una tarjeta gráfica con una buena cantidad de núcleos CUDA, como mínimo una GeForce RTX 3050.



Para activar el uso de GPU en GROMACS, es necesario instalar el toolkit CUDA y verificar su correcta instalación:

```
$ sudo apt install nvidia-cuda-toolkit
$ nvidia-smi
```

Compilar GROMACS desde el código fuente con soporte GPU:

```
$ cmake.. -DGMX_GPU=CUDA -DGMX_BUILD_OWN_FFTW=ON -
DREGRESSIONTEST_DOWNLOAD=ON
$ make -j4
$ sudo make install
```

Verificar que GROMACS detecta la GPU:

```
$ gmx mdrun -version
```

Identificamos la línea que indica soporte para CUDA o GPU. Si no aparece o se encuentra desactivada, verificar que se tiene instalados los controladores adecuados para el modelo de su tarjeta gráfica. Si tiene dudas o desea más información sobre la configuración, instalación o uso de GROMACS, visitar su sitio oficial: <https://www.gromacs.org>.



Apéndice B. Metodología para la selección del parámetro -ng en TTClust

En versiones anteriores, TTClust incluía un procedimiento automático para seleccionar el número óptimo de clústeres mediante el análisis de codo (elbow analysis). Sin embargo, a partir de la versión 4.10.1, esta función dejó de estar disponible, por lo que el usuario debe especificar manualmente el número de grupos mediante la opción -ng (por ejemplo, -ng 3, -ng 4, -ng 5).

Para solventar esta limitación, se construirá un diagrama de codo, que relacione el número de clústeres con el cutoff obtenido para cada agrupamiento. En el contexto de *clustering* estructural, el cutoff corresponde al RMSD umbral que define cuánta similitud deben compartir las conformaciones para pertenecer de un mismo grupo.

La metodología general consiste en:

- I. Ejecutar varios *clusterings* con diferente número de grupos
- II. Registrar el cutoff reportado en los archivos .log
- III. Construir una gráfica (número de clústeres vs. cutoff)
- IV. Identificar el punto a partir del cual agregar más clústeres deja de disminuir significativamente el cutoff; ese punto de inflexión se interpreta como el número óptimo de clústeres.

Esta aproximación es equivalente al método tradicional del diagrama de codo empleado en algoritmos como K-means.

TTClust requiere un archivo *.pdb* que describa la topología de los átomos incluidos en la trayectoria centrada. Si no se dispone de él, puede generarse con:

```
$ echo 1 1 | gmx trjconv -f traj_center.xtc -s md.tpr -o md.pdb -dump 0
```

Este comando extrae el frame inicial (-dump 0) seleccionando únicamente la proteína. Se recomienda utilizar PDB en lugar de GRO, ya que el formato PDB mantiene identificadores de cadenas y facilita etapas posteriores como *docking* o análisis estructural.

Con el fin de simplificar este proceso, se incluye un script bash (ttclust.sh) que:

- I. Ejecuta automáticamente los *clusterings* para distintos valores de -ng
- II. Extrae los valores de cutoff
- III. Genera un archivo resultados_ttclust.csv listo para graficarse en Excel

Después de otorgar permisos de ejecución:

```
$ chmod +x ttclust.sh  
$ ./ttclust.sh
```

El programa solicitará:

- la trayectoria *.xtc*
- el archivo *.pdb* generado
- el número máximo de clústeres a evaluar



El archivo final contiene columnas con:

- número de clústeres
- cutoff
- RMSD promedio
- tamaños de los grupos

lo que permite reconstruir el diagrama de codo y justificar la elección final.

Una vez identificado el número óptimo de grupos, se procede al *clustering* definitivo.

Por ejemplo, si se seleccionan 5 clústeres, el comando sería:

```
$ ttclust -f traj_center.xtc -t md.pdb -s 2 -l log.log -sr backbone  
-m centroid -ng 5
```

Este comando:

- analiza únicamente el backbone (-sr backbone)
- selecciona el centroide como estructura representativa de cada clúster (-m centroid)
- usa una frecuencia de muestreo fina (-s 2)

TTClust generará una carpeta clustering con las estructuras representativas (cluster_#.pdb), matrices de RMSD y reportes estadísticos. Estas estructuras se utilizarán como conformaciones alternativas en el proceso de cribado virtual.

De este modo, la elección del número de clústeres deja de ser arbitraria: se fundamenta en un procedimiento reproducible, basado en *clusterings* exploratorios y en un análisis gráfico tipo método del codo. Esto permite seleccionar valores como -ng 3, -ng 4 o -ng 5 con evidencia cuantitativa, asegurando que las conformaciones representativas describan adecuadamente la dinámica estructural del sistema bajo estudio.