



UNIVERSIDAD DE LA RIOJA

TRABAJO FIN DE ESTUDIOS

Título

Análisis mediante técnicas computacionales de interacciones
inhibidor-enzima

Autor/es

DAVID PÉREZ GOICOA

Director/es

RODRIGO MARTÍNEZ RUIZ

Facultad

Facultad de Ciencia y Tecnología

Titulación

Grado en Química

Departamento

QUÍMICA

Curso académico

2020-21



Análisis mediante técnicas computacionales de interacciones inhibidor-enzima
, de DAVID PÉREZ GOICOA

(publicada por la Universidad de La Rioja) se difunde bajo una Licencia Creative Commons Reconocimiento-NoComercial-SinObraDerivada 3.0 Unported.

Permisos que vayan más allá de lo cubierto por esta licencia pueden solicitarse a los titulares del copyright.

© El autor, 2021

© Universidad de La Rioja, 2021

publicaciones.unirioja.es

E-mail: publicaciones@unirioja.es



UNIVERSIDAD DE LA RIOJA

Facultad de Ciencia y Tecnología

TRABAJO FIN DE GRADO

Grado en Química

Análisis mediante técnicas computacionales de
interacciones inhibidor-enzima

Inhibitor-enzyme interaction analysis using computational
techniques

Realizado por:

David Pérez Goicoa

Tutelado por:

Rodrigo Martínez Ruiz

Logroño, Septiembre, 2021

ÍNDICE

RESUMEN/ABSTRACT	4
INTRODUCCIÓN	5
MECÁNICA MOLECULAR	11
DINAMICA MOLECULAR.....	17
PASOS DE LA SIMULACIÓN	19
PROGRAMAS Y HERRAMIENTAS UTILIZADAS.....	22
RESULTADOS.....	24
INTERPRETACIÓN DE LOS RESULTADOS Y COMPARACIÓN CON DATOS BIBLIOGRÁFICOS	32
BIBLIOGRAFÍA	36
RESUMEN Y CONCLUSIONES	35

RESUMEN/ABSTRACT

El objetivo de este Trabajo de Fin de Grado es el estudio mediante Dinámica Molecular de las interacciones que se producen entre la proteasa principal del SARS- CoV-2, M^{Pro}, y el inhibidor N3. Sabiendo cómo interacciona N3, un inhibidor muy efectivo de esta proteasa con la proteína se pueden diseñar fármacos dirigidos específicamente al sitio activo en el que se sitúa este inhibidor, por lo que estas interacciones proteína-ligando presentan gran interés.

Los cálculos se han realizado en el sistema operativo Linux(Ubuntu) con el software libre GROMACS. La estructura de rayos X de partida se obtuvo de un repositorio público de proteínas. Se parametrizaron por separado la proteína y el inhibidor, lanzando posteriormente varios cálculos de Mecánica Molecular. Se obtuvieron los ficheros de los colectivos NVT y NPT y, a partir de ellas, se lanzó la dinámica molecular para obtener las interacciones proteína-ligando. Se observó que los enlaces de hidrógeno son las interacciones principales entre ambos. Así, se analizó el número, las distancias donador-aceptor y el ángulo de los enlaces de hidrógeno presentes entre N3 y la proteína. También se identificaron los aminoácidos participantes en el anclaje de N3. Los resultados se compararon con la bibliografía para comprobar los mecanismos de unión proteína-ligando.

The objective of this Degree Thesis is the study, using Molecular Mechanics, of the interactions present between SARS-CoV-2's main protease, M^{Pro}, and N3, an inhibitor. By knowing the way N3, a very effective inhibitor for this particular protease, interacts with M^{Pro}, specifically designed drugs can be developed to bind in the active site, making these interactions particularly interesting.

Calculations were made using Linux (Ubuntu) operative system and free software GROMACS. M^{Pro}-N3's X-Ray structure file was obtained from a public protein databank. M^{Pro} and N3 were separately parametrized, launching several Molecular Dynamics calculations. NVT and NPT colectives were obtained, from which a molecular Dynamic was launched to obtain protein-ligand interactions. Hydrogen bonds were detected as the main interaction type between both. Thus, the number of interactions along the

time, the donor-acceptor distances and angle distributions in the hydrogen bonds were analyzed. The amino acids participating the N3 docking were identified and compared to several references to corroborate protein-ligand interaction mechanisms.

INTRODUCCIÓN

El objetivo de este Trabajo de Fin de Grado es el estudio del complejo proteína-ligando M^{Pro}-N3, siendo M^{Pro} la proteasa principal del SARS-CoV-2. Mediante esta enzima, el coronavirus causante de la pandemia ocasionada por la enfermedad COVID-19 cataliza la uno de los pasos principales en la replicación del material genético vírico que introduce en las células huésped.

En diciembre de 2019 se informó en Wuhan, China, de un número creciente de pacientes que presentaban síntomas de una neumonía desconocida hasta la fecha. Posteriores investigaciones revelaron que el causante de la enfermedad fue un nuevo coronavirus, que fue nombrado oficialmente por la Organización Mundial de la Salud como Severe Acute Respiratory Syndrome – Coronavirus 2 (SARS-CoV-2). La enfermedad que este virus provoca se designó como COVID-19, y ha ocasionado una gran cantidad de fallecimientos entre la población, así como graves trastornos respiratorios crónicos para los pacientes que lograron superar la enfermedad. Su gran capacidad de contagio, adaptabilidad y las similitudes de los síntomas que ocasiona (fiebre, tos seca, pérdida de olfato, fatiga,...)¹ con enfermedades comunes provocó una situación de pandemia mundial en muy poco tiempo, y que todavía se mantiene.

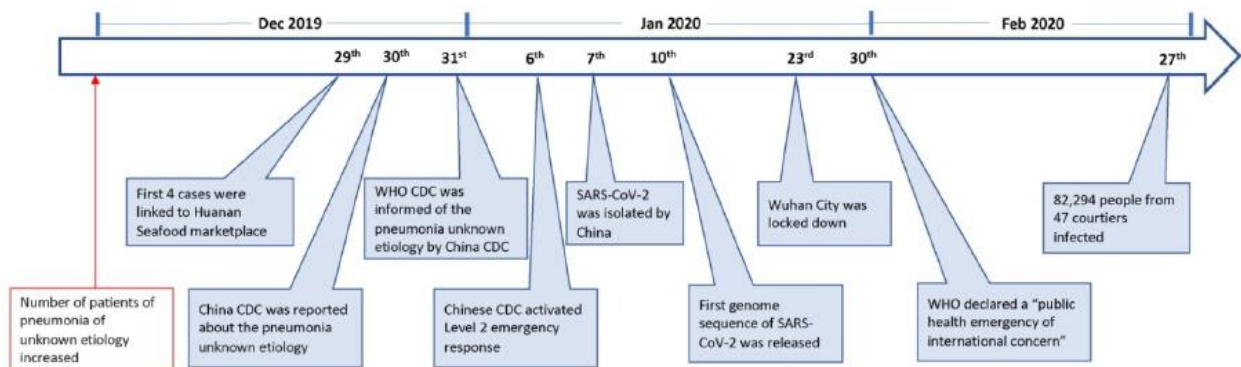


Figura 1. Línea de tiempo de la enfermedad COVID-19²

El SARS-CoV-2 es un betacoronavirus cuyo aspecto es el de una corona solar esférica de unos 60 a 140 nm de diámetro que posee varias espigas o "spikes" de unos 8 a 12 nm de longitud. Esta partícula vírica consiste en un nucleocápside, que protege al material genético del virus y una envoltura externa. El material genético está asociado a la nucleocápside a través de la proteína N, fosforilada e insertada en la bicapa lipídica de

la envoltura externa. En la envoltura externa se encuentran la proteína S de los spikes, la proteína membranal M, y la de envoltura E, además de varias proteínas accesorias. La proteína M mantiene la curvatura de la membrana, mientras que S facilita la unión del virión a la célula huésped y E participa en el ensamblaje y liberación del virus^{3,4}.

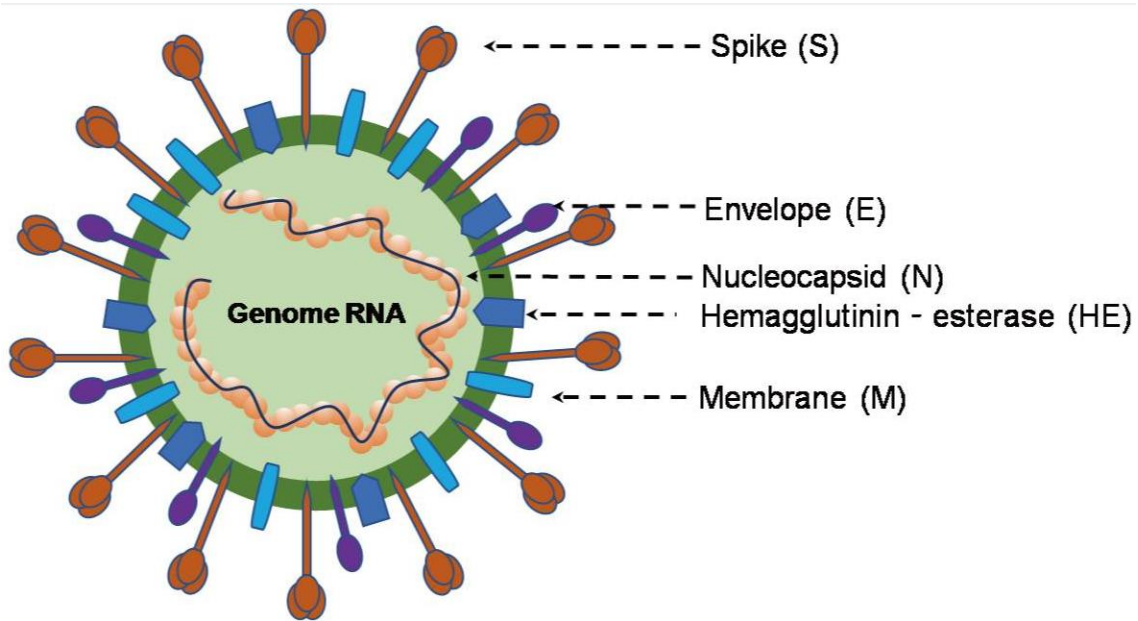


Figura 2. Esquema del virión de SARS-CoV-2⁵

El genoma del SARS-CoV-2 consta de una cadena de RNA monocatenario de polaridad positiva de entre 26 y 32 kilobases de longitud. Presenta un capuchón metilado en el extremo 5' y una cola poliadenilada en el extremo 3', haciéndolo similar al mRNA de una célula eucariota huésped. El genoma se puede dividir en tercios, siendo los dos primeros los marcos abiertos de lectura ORF 1a y ORF 1b, situados en el extremo 5'. El último tercio codifica los genes de las proteínas estructurales del virus, S, M, E y N, así como los de las proteínas accesorias. Los ORF 1^a y 1^b, por su parte, codifican dos proteínas de gran tamaño, pp1a y pp1ab, que se dividen posteriormente^{3,4}.

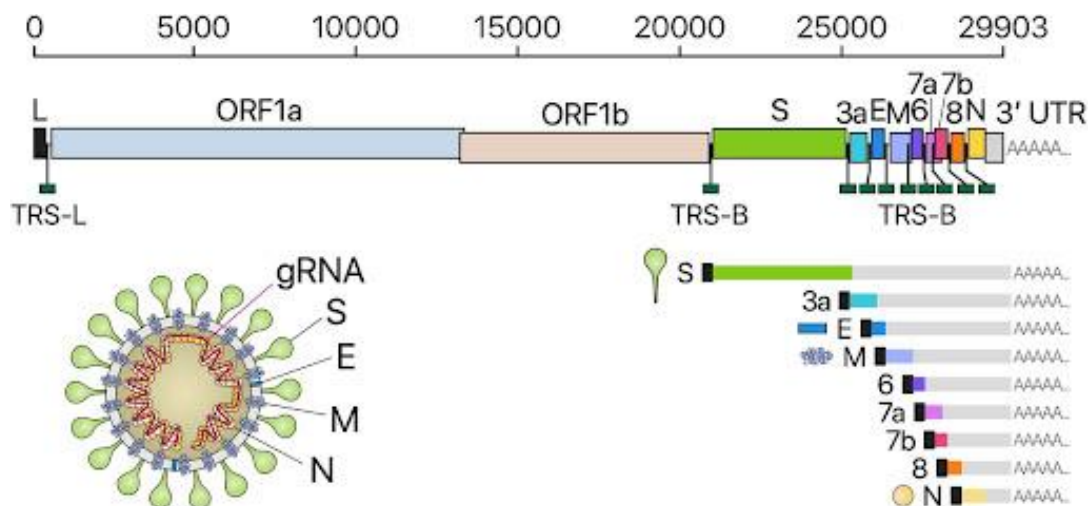


Figura 3. Esquema del genoma de SARS-CoV-2⁴

El proceso de infección del virus comienza con la unión del virión y la célula huésped a través de la proteína S y el receptor de la enzima convertidora de la angiotensina 2 ACE-2. ACE-2 regula la presión arterial y sus receptores se sitúan en el tracto respiratorio bajo, riñones, corazón, estómago, vejiga esófago e intestino. La proteína S se une al receptor a través de su subunidad S1, mientras que la S2 se encarga de la fusión de la membrana del virus con la célula huésped. La subunidad S2 es cortada por la enzima proteasa TMPRSS 2, fusionando las membranas y provocando la entrada del material genético en la célula mediante endocitosis. Entonces, la nucleocápside del virus queda libre en la célula infectada, provocando que el RNA viral actúe como mRNA. El gen de la replicasa viral, a través de ORF1a y ORF1b, se traduce en dos poliproteínas, que son procesadas por enzimas como la M^{pro}, quimotripsina y otras similares a la papaína. Estas poliproteínas, pp1ab y pp1b, generan dieciséis proteínas no estructurales, que son clave para formar el llamado complejo RTC y se dedican principalmente a la replicación y transcripción del virus mediante actividades polimerasa, proteasa, helicasa, Para la replicación del material genético viral, la cadena de polaridad positiva de partida sirve como molde para una cadena de polaridad negativa, -ssRNA, a partir de la cual se forman pp1a y pp1ab. Al procesar estas poliproteínas y formar el complejo RTC, éste crea nuevas copias del genoma +ssRNA a partir del molde negativo. El material viral se asocia en el

aparato de Golgi con las proteínas estructurales formadas a procesar el tercer tercio del material que entró a la célula, generando nuevos viriones^{3,4}.

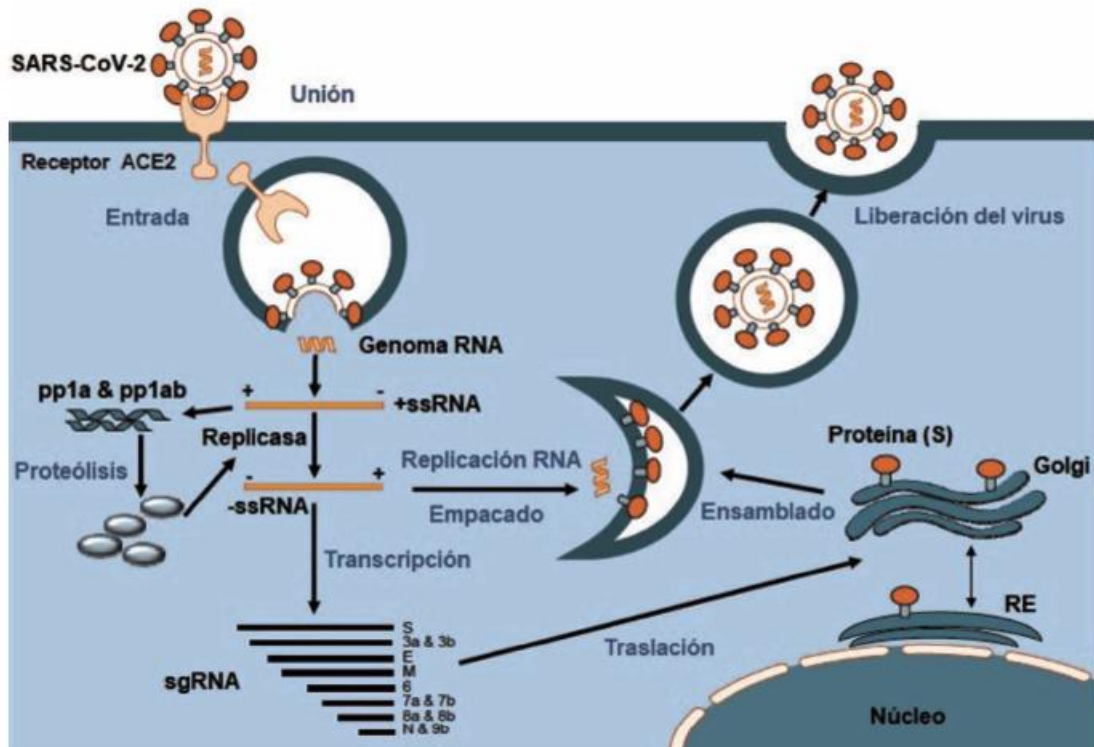


Figura 4. Esquema del mecanismo de infección de SARS-CoV-2³

M^{Pro} tiene, pues, un papel fundamental en la propagación de la infección del organismo huésped, ya que participa activamente en el procesamiento de las poliproteínas pp1a y pp1ab.

El inhibidor N3, propuesto mediante diseño de fármacos asistido por ordenador⁶, es un aceptor de Michael que inhibe eficientemente la M^{Pro} de varios de los coronavirus conocidos.

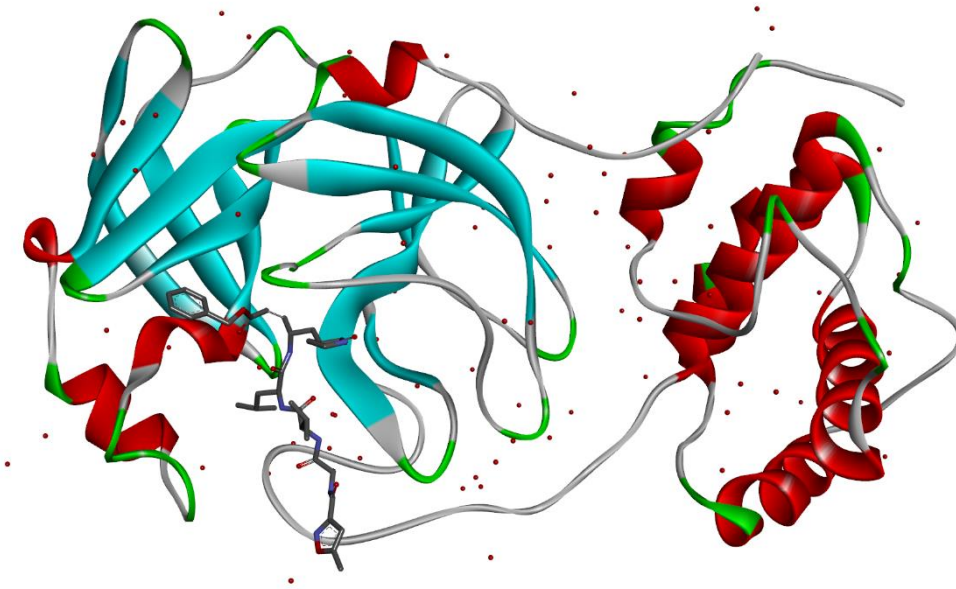


Figura 5. Visualización del sistema M^{Pro} -N3 (PDB:6lu7)

Analizando las interacciones que existen entre N3 y M^{Pro} , se puede conseguir que esta enzima se desactive, de manera que sea incapaz de procesar las poliproteínas formadas en la transcripción del RNA vírico y, por tanto, se paralice la infección por SARS-CoV-2.

La estructura cristalina de la M^{Pro} -N3 (PDB 6lu7⁷) muestra una proteína que consta de 306 residuos, ordenados en tres dominios. Los dominios I, II, III están constituidos por los residuos 8-101, 102-184 y 201-303, respectivamente. Los dominios I y II tienen una estructura de barril beta antiparalelo, mientras que el dominio III lo constituyen cinco alfa-hélices que forman un clúster globular antiparalelo, conectado al dominio II mediante un bucle largo formado por los residuos 185-200. El sitio activo donde se acopla el inhibidor N3 se sitúa en una hendidura entre los dominios I y II⁸.

Esta gran capacidad de inhibición de N3 sobre M^{Pro} permite el diseño de fármacos orientados a inhibir la actividad de la proteasa mediante la modificación del inhibidor en su unión con uno o varios de los subsitios activos de la M^{Pro}

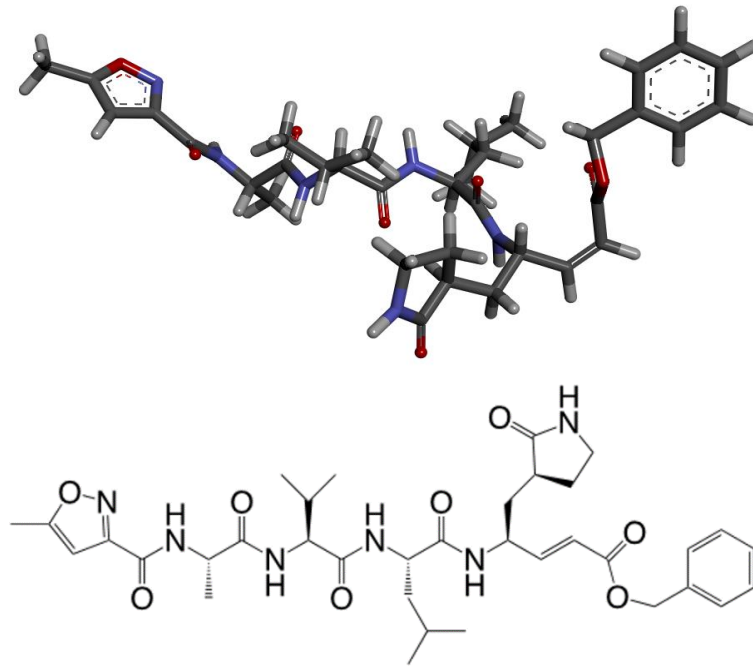


Figura 7: Estructuras 3D (arriba) y 2D (abajo) del inhibidor N3

El mecanismo de inhibición del N3 es conocido y responde a una adición de Michael entre el carbono beta del doble enlace de N3 y la cisteína 145 de M^{Pro} , proponiéndose dos mecanismos de reacción, y siendo el mecanismo zwitteriónico ($E^{(+/-):I}$) el más plausible^{9,10}.

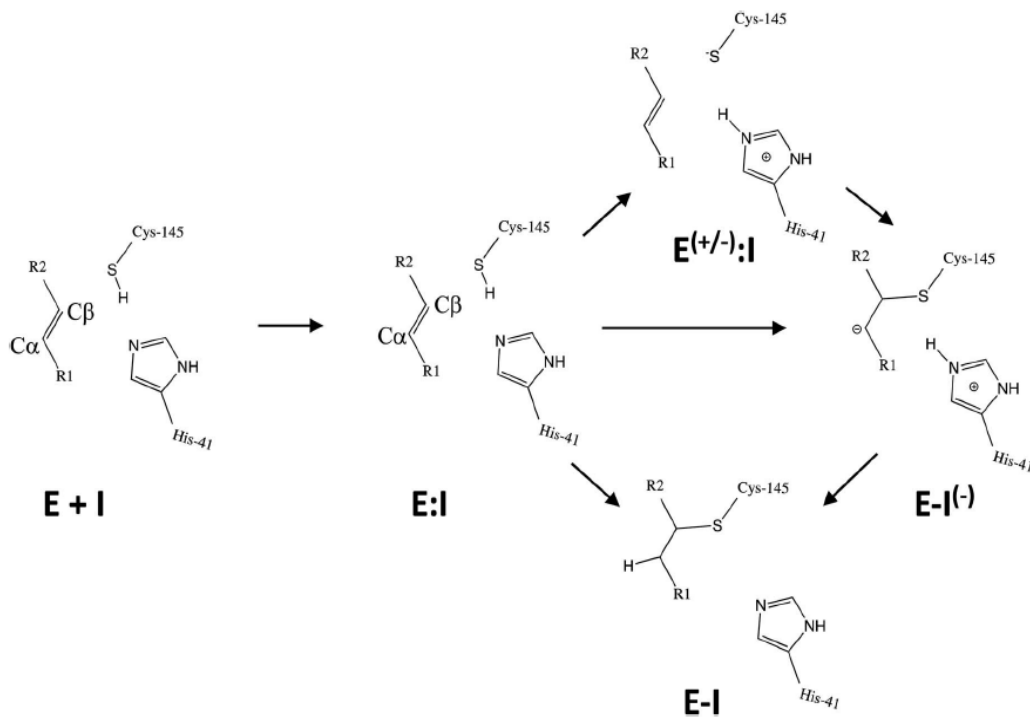


Figura 6: Esquema del mecanismo de la Adición de Michael entre M^{Pro} y N3¹⁰

En el mecanismo, la cisteína 145 de la proteína cede su protón, atacando posteriormente al carbono beta del doble enlace que contiene el ligando. La ruta por la que sucede este mecanismo es la explicada en la *Figura 7*, e implica la presencia de una especie zwitteriónica con la carga negativa sobre azufre la cisteína 145 (CYS 145) y la carga positiva sobre la histidina 41 (HIS 41) de la proteína. Estos dos aminoácidos forman una díada catalítica en la que el grupo imidazol de HIS 41 polariza y activa el grupo -SH de la CYS 145, con lo que se forma un par iónico altamente nucleófilo que reacciona con el carbono beta del inhibidor N3. El motivo de qué la ruta zwitteriónica sea la que se considera más estable reside en las energías libres de los intermedios E:I y E-I. Se ha observado que la energía libre de la forma neutra E:I es menor que la de la forma E-I, lo que dirige el mecanismo hacia la formación del intermedio $E^{(+/-):I}$. A partir de este intermedio se forma un enlace entre la CYS 145 y el carbono beta del inhibidor, formando el intermedio $E:I^{(-)}$. Para finalizar, se produce la transferencia del protón de la HIS 41 al carbono alfa del inhibidor se produce prácticamente sin energía de activación en el paso siguiente del mecanismo, formando la especie covalente final $E-I^{10}$.

Analizando las interacciones entre el inhibidor y la proteasa en el complejo de Michaelis (E:I) se puede obtener gran cantidad de información importante en el desarrollo de posibles medicamentos contra la infección del SARS-CoV-2.

La Dinámica Molecular es una técnica computacional que, entre otras características, permite analizar el movimiento, a lo largo del tiempo y a nivel atómico, de moléculas de interés biológico en disolución acuosa. Este movimiento está influenciado por todos los átomos que componen el sistema, por lo que conocer cómo interaccionan a lo largo del tiempo los átomos de los fragmentos que componen el inhibidor y la proteasa cuando forman el complejo de Michaelis es una información de partida interesante en la propuesta de nuevos inhibidores de la M^{Pro} . Se debe tener en cuenta que estos posibles inhibidores deberán imitar los puntos de anclaje del N3 sobre la M^{Pro} .

En este trabajo, se analizará mediante Dinámica Molecular el comportamiento a lo largo de 100 ns de las interacciones no enlazantes entre la N3 y M^{Pro} en el complejo E:I.

MECÁNICA MOLECULAR

Para estudiar este sistema, se ha recurrido a la Dinámica Molecular, que resulta útil para conocer las interacciones que presenta el inhibidor con la proteína a la que inhibe.

La Dinámica Molecular se engloba dentro de la Mecánica molecular, que utiliza la física clásica para estudiar la superficie de energía potencial (SEP). La Dinámica Molecular estudia el movimiento de las moléculas dentro de esa SEP. Así, utilizando ecuaciones más sencillas que las que utiliza la Mecánica Cuántica, se pueden llevar a cabo computacionalmente cálculos que ésta no permitiría por el elevado coste computacional que suponen. Esto significa que los cálculos aplicados a moléculas grandes, como la proteasa de estudio, pueden realizarse mediante Mecánica Molecular. Por el contrario, la Mecánica Molecular no permite conocer el comportamiento de los electrones, por lo que no es útil para proporcionar información sobre reacciones químicas¹¹.

Los campos de fuerzas son las herramientas de la mecánica molecular. Son una suma de los términos enlazantes y no enlazantes del potencial que describen las características del sistema a estudiar. Los términos enlazantes son las tensiones de enlace, ángulos de enlace y diedros (propios e impropios) y términos cruzados (acoplamientos entre movimientos de tensión, torsión y flexión de los enlaces). Los términos no enlazantes son las interacciones coulombianas y las fuerzas de Van der Waals, que se engloban dentro del llamado potencial de Lennard-Jones. La energía potencial del sistema, pues, se obtiene de la suma de estos términos¹¹.

$$\begin{aligned} V_{total} = & \sum_{Enlaces} k_{ij}(k_{ij} - k_{ij})^2 + \sum_{Ángulos} k_{ijk}(\theta_{ijk} - \theta_0)^2 \\ & + \sum_{Diedros} k_{ijkl} (1 + \cos(n \cdot \varphi_{ijkl} - \varphi_0))^2 \\ & + \sum_{Impropios} k'_{ijkl}(\varphi'_{ijkl} - \varphi'_0)^2 + \sum_{Cruzados} k_{ik}(r_{ik} - r_{ik0})^2 \\ & + \sum_{Coulomb} \frac{1}{4\pi\epsilon_0\epsilon} \cdot \frac{q_i \cdot q_j}{r_{ij}} + \sum_{vdW} 4\epsilon_{ij} \cdot \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right) \end{aligned} \quad (1)$$

Todos los parámetros de un sistema deben estar definidos en el campo de fuerzas para poder estudiarlo, por lo que es necesario parametrizar estos átomos, enlaces, ángulos, etc, si no lo están previamente. Esta parametrización consiste en reproducir la estructura energética y la información sobre los momentos dipolares del sistema. Los distintos campos de fuerzas se diferencian precisamente en la manera en la que operan para obtener estos parámetros del campo. Además, así como en espectroscopía las propiedades de un tipo de enlace (como la zona de aparición en el espectro IR de un grupo carbonilo) no dependen de la molécula en la que estén incluidos, los campos de fuerzas poseen la misma transferibilidad de las propiedades, aunque se encuentren en moléculas distintas, por lo que los campos de fuerzas poseen una generalidad que los hace tremendamente útiles. Así pues, no es de extrañar que los primeros campos de aplicación de la mecánica molecular fueran las biomoléculas, puesto que los tipos de átomos y enlaces se repiten entre los aminoácidos naturales¹¹.

Empirical Potential Energy Function

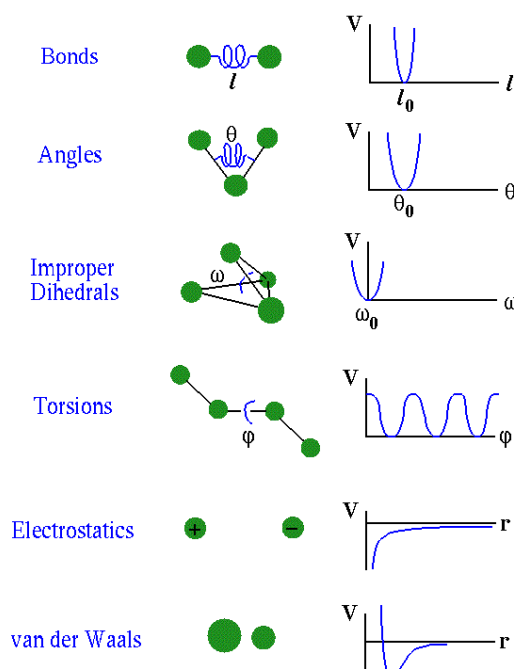


Figura 8: Visualización de los parámetros del campo de fuerzas¹²

Los campos de fuerzas realizan la parametrización mediante cálculos *ab initio* aplicados a fragmentos de moléculas, que se van ajustando para que las propiedades de estructura, energía y momentos dipolares del sistema estén bien definidas en el campo de fuerzas. Estas operaciones se repiten hasta que el campo de fuerzas aplicado a ese sistema está completo y lo describe correctamente¹¹.

Una vez definido el campo de fuerzas a utilizar, y sabiendo qué tipo de átomos se tienen en el sistema de estudio, es necesario conocer la topología del mismo. La topología de las moléculas define la conectividad entre átomos, los ángulos diedros, tensiones, etc¹¹. En el caso de M^{Pro}, se parte de una estructura cristalina de rayos X obtenida de un repositorio público⁷, que se parametriza para obtener un fichero que contiene los aminoácidos a simular, su conectividad y el tipo de átomos que los componen.

Para el cálculo de las interacciones no enlazantes, el coste computacional del cálculo aumenta exponencialmente con el número de átomos a estudiar, por lo que se utilizan los llamados cut-off. Consiste en marcar una distancia a partir de la cual el programa no calcula las fuerzas interatómicas por situarse estos átomos en puntos muy alejados entre sí. Sin embargo, se debe ser cuidadoso en este punto, puesto que, para algunos

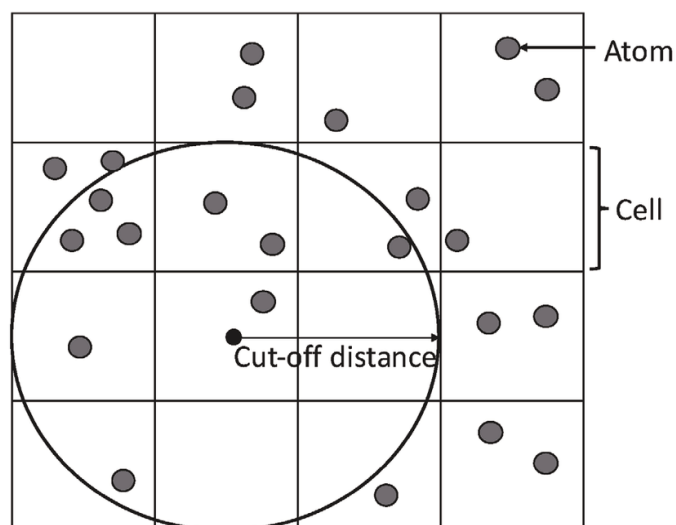


Figura 9. Esquema de un cut-off¹³

sistemas, como metal-proteína, las fuerzas colombianas a larga distancia son importantes y no puede usarse un cut-off¹¹.

El disolvente que rodea la molécula de estudio es imprescindible a la hora de estudiar las biomoléculas, que en su mayoría se encuentran en medio acuoso. En este trabajo, se ha incluido el disolvente en forma explícita, es decir, añadiendo un número determinado de moléculas de agua. Prácticamente todos los campos de fuerzas incluyen el agua como disolvente en la simulación, aunque se puede elegir entre varios modelos de moléculas de agua según la complejidad del sistema en su interacción molécula-disolvente. Estos modelos se distinguen en el número de puntos que se utilizan en la definición de cada

molécula de agua. Para este estudio, se ha utilizado el modelo de tres puntos, TIP3P, puesto que es el que menor coste computacional tiene. Otros modelos añaden átomos “dummy”, que representan la carga negativa de los pares solitarios del oxígeno, mejorando la distribución electrostática de la molécula. No obstante, el disolvente y sus interacciones con M^{Pro} no son el objeto de estudio de este trabajo, por lo que se ha buscado el menor coste computacional¹¹.

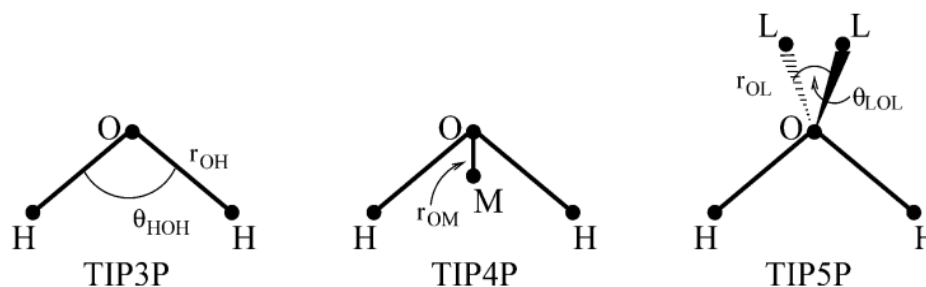


Figura 10: Tipos de representación del agua en campos de fuerzas ¹⁴

Tras definir el tipo de moléculas de agua, se debe abordar el tamaño del sistema. El espacio debe ser limitado para reducir en la medida de lo posible el coste computacional, por lo que se establecen unos límites espaciales para el sistema. Tradicionalmente se utilizaba el modelo de gota, que aplica un potencial esférico al sistema. En la actualidad, se utilizan los modelos de cajas, siendo el más eficiente en el llenado del espacio tridimensional la caja con forma de dodecaedro truncado. Se llama Periodic Boundary Conditions (PBC) a la extensión de esta simulación a imágenes adyacentes del sistema situadas en las tres direcciones del espacio. Así, puesto que las interacciones no enlazantes se aplican a las imágenes periódicas adyacentes, se simula un sistema continuo. Para evitar que un átomo pueda ver su imagen en la celda adyacente, se aplican cut-off, que deben ser menores que el tamaño de la caja para evitar errores de cálculo¹¹.

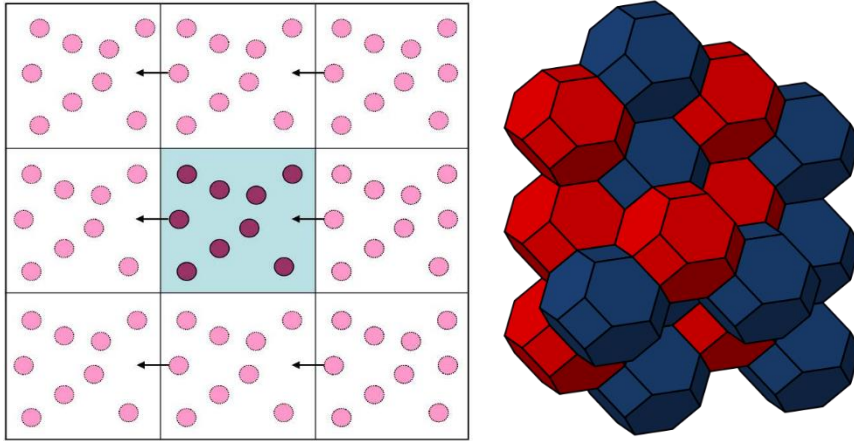


Figura 11: Visualización de Periodic Boundary Conditions (izq.)¹⁵ y el clúster de celdas dodecaédricas(drcha.)¹⁶.

Una vez parametrizado y limitado el sistema, se pueden buscar mínimos de energía potencial a través de la expresión del potencial que calcula el campo de fuerzas (**Ecuación 1**). Para ello, el programa debe realizar una minimización energética de la estructura que la calculado en primera instancia. Es necesario conocer el sistema con el que se trabaja porque es posible, y en la mayoría de los casos es la realidad, que el mínimo energético que se obtiene con el programa no sea el mínimo absoluto, sino uno

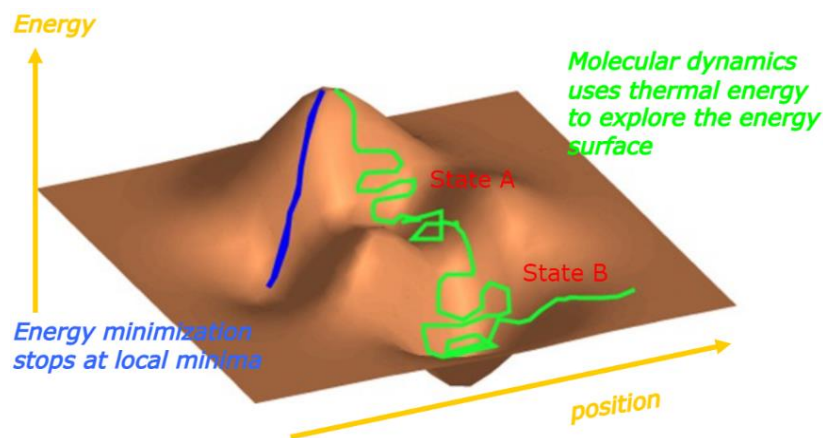


Figura 12: Visualización del paso de minimización energética, con varios mínimos locales¹⁷

de los posibles mínimos relativos. De cualquier manera, se sigue la dirección del gradiente de energía del sistema para encontrar los mínimos¹¹.

El gradiente de la energía potencial se define como la derivada del potencial respecto a las coordenadas del sistema. Así pues, para minimizar la energía, se ha usado el método matemático steepest descent. Se varía una coordenada en la dirección contraria a la que marca el gradiente con un paso de minimización ϵ , que marca cuánto debe moverse esa coordenada para ir hacia el mínimo, y se define el valor mínimo del gradiente, en el que termina la optimización. Como la mayoría de los sistemas operan a 300K, se debe tener en cuenta la energía cinética de las moléculas debida a esa temperatura, por lo que, tras realizar la minimización energética, se aplican los métodos de la Dinámica Molecular¹¹.

$$\nabla V = \left(\frac{\partial V}{\partial r_1}, \frac{\partial V}{\partial r_2}, \frac{\partial V}{\partial \theta_1}, \dots \right) \quad (2)$$

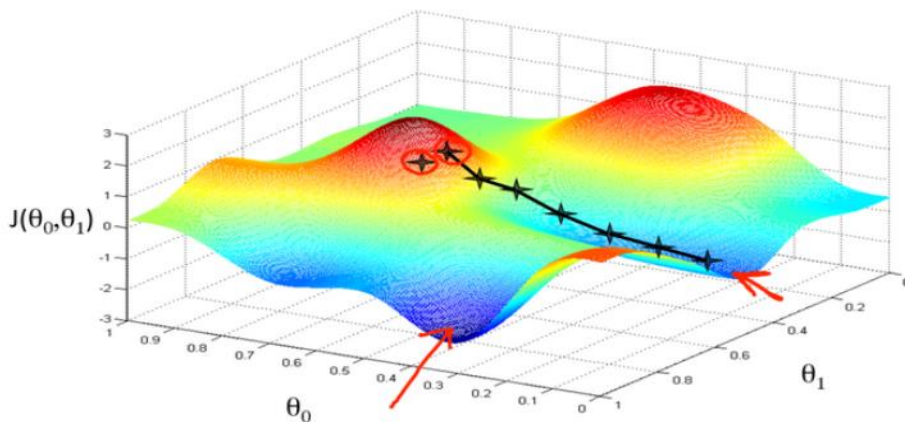


Figura 13: Visualización del método steepest descent de minimización energética¹⁸.

DINAMICA MOLECULAR

La Mecánica Molecular limita su estudio a la energía del sistema a 0K, lo cual no representa adecuadamente la realidad. Los sistemas, en su mayoría, se estudian a 300 K, lo que implica que las moléculas se encuentran en movimiento y poseen cierta energía cinética y una determinada velocidad. La ley de distribución de Maxwell relaciona la velocidad de las partículas del sistema con la temperatura del mismo¹¹.

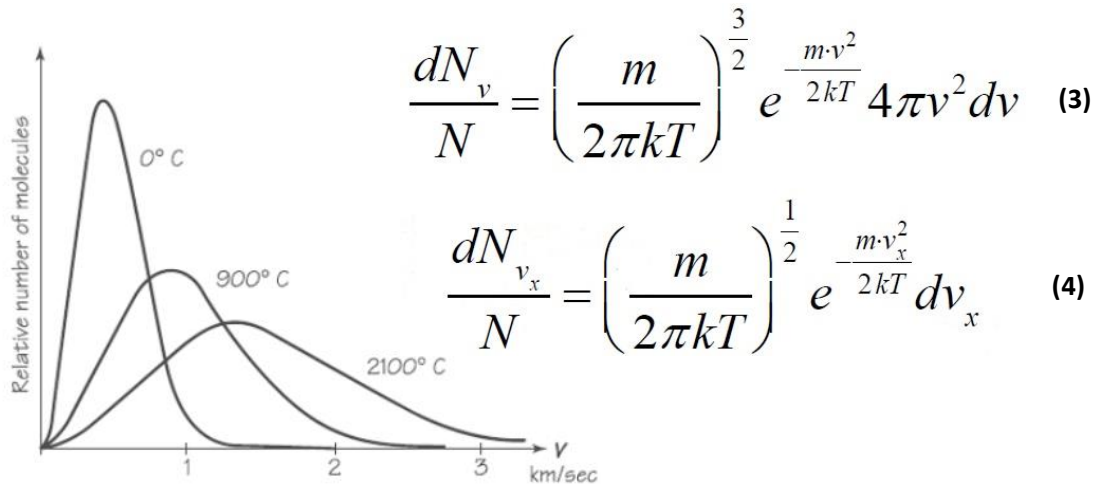


Figura 14: Diagrama de la distribución de velocidades de Maxwell a distintas temperaturas¹⁹

A los distintos átomos de la estructura minimizada mediante MM se les asigna una velocidad de acuerdo con la distribución de velocidades de Maxwell, y su movimiento se ajusta tanto a esta distribución como al potencial del sistema. Así, la estructura del sistema se va moviendo a lo largo de la SEP, llegando habitualmente al mínimo más estable, ya que se suele trabajar a 300 K. Puesto que la única fuerza que actúa sobre el sistema es el campo de fuerzas, conociendo posición y velocidad de las partículas se puede conocer el siguiente punto en el que se encontrará la estructura. El conjunto de estos puntos forma la trayectoria del sistema. Además, se define como espacio-fase al conjunto de todas las posiciones y momentos lineales de las partículas del sistema¹¹.

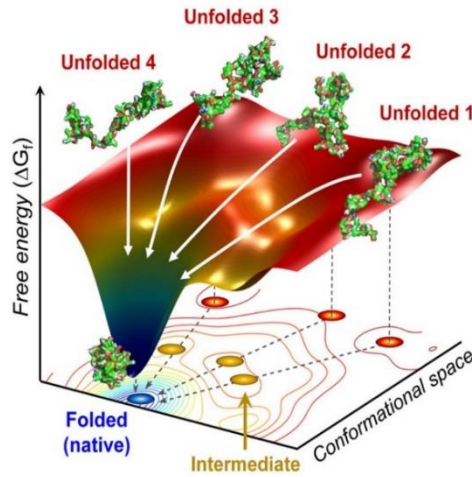


Figura 15: Visualización de la relación entre las distintas conformaciones de una proteína y sus mínimos de energía potencial²⁰

Para asegurar que la temperatura se mantiene durante la simulación, se introducen termostatos: ecuaciones que parten de la ley de distribución de Maxwell y que garantizan una temperatura constante durante toda la trayectoria¹¹.

$$\langle v^2 \rangle = \frac{3 \cdot R \cdot T}{M} \quad ; \quad T(t) = \frac{1}{3 \cdot R} \cdot \frac{\prod_{i=1}^N M_i v_i^2(t)}{N} \quad (5;6)$$

Por otra parte, las propiedades del sistema que se desean conocer se calculan mediante promedios a lo largo de la simulación en cada estructura que se ha definido en la trayectoria. En ocasiones, las propiedades a estudiar no tienen solución analítica, por lo que se utilizan propagadores: ecuaciones a las que se proporciona un valor inicial que se modifica con cada variación de la propiedad a lo largo del tiempo¹¹.

$$\vec{r}(t + \Delta t) = \vec{r}(t) + \vec{v}\left(t + \frac{1}{2} \Delta t\right) \cdot \Delta t \quad (7)$$

$$\vec{v}\left(t + \frac{1}{2} \Delta t\right) = \vec{v}\left(t - \frac{1}{2} \Delta t\right) + \vec{a}(t) \cdot \Delta t \quad (8)$$

Tras alcanzar la temperatura deseada, se deja evolucionar al sistema con el tiempo, comprobando así que es estable, no se disocia y mantiene la temperatura alrededor del valor deseado. Una vez equilibrado el sistema, se procede a la fase más larga de la simulación: la fase de producción. A partir de la trayectoria del sistema (estructura y velocidades de cada partícula), se calculan las propiedades¹¹.

Las simulaciones reproducen unas condiciones fisicoquímicas que se relacionan con ciertos colectivos de la termodinámica estadística. El colectivo canónico NVT corresponde a la simulación en la que se mantienen constantes el número de partículas, el volumen del sistema y la temperatura, y sirve para simular sistemas endotérmicos y exotérmicos a través de un termostato. Por otra parte, el colectivo isobárico-isotermo NPT mantiene el número de partículas, la presión y la temperatura, y se utiliza para simular sistemas a presión atmosférica y temperatura ambiente constante¹¹.

PASOS DE LA SIMULACIÓN

Una vez introducidos los conceptos de la mecánica y dinámica moleculares, se van a detallar los pasos seguidos para la obtención de los datos que se presentarán. El siguiente diagrama de flujo resume el orden de los pasos seguidos para la obtención de la información sobre los enlaces de hidrógeno y los aminoácidos que participan en ellos.

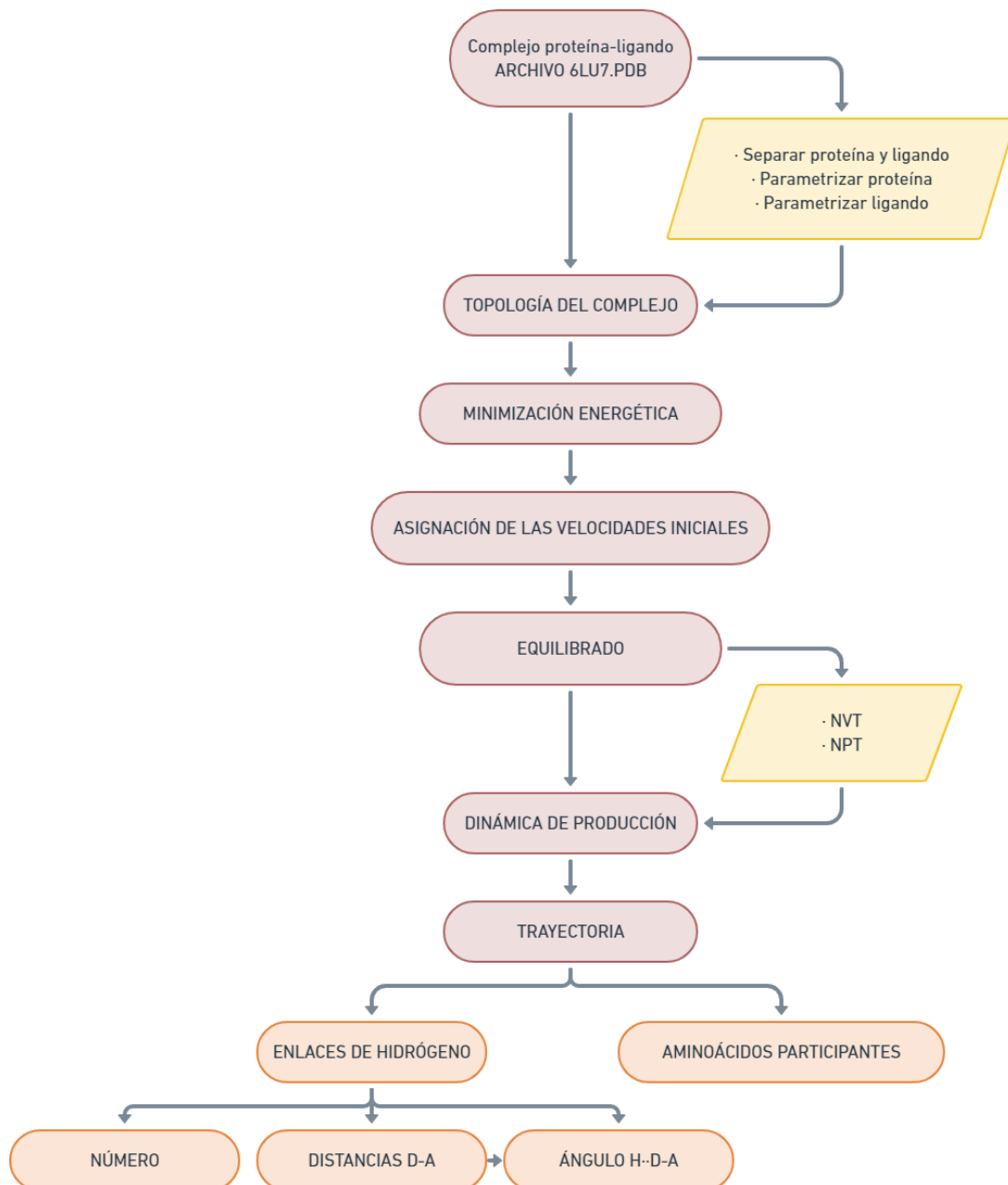


Figura 16. Diagrama de flujo de los pasos de la simulación

El fichero de partida es un archivo PDB obtenido del sitio web RCSB.org⁷, un banco de proteínas que contiene gran cantidad de información sobre biomoléculas, los estudios a los que pertenecen y los artículos en los que se han utilizado. El código del archivo de estudio es 6lu7, y contiene la estructura de rayos X del complejo M^{Pro}-N3.

ATOM	2341	N	THR	A	304	-40.002	6.614	34.069	1.00	78.33	N
ATOM	2342	CA	THR	A	304	-41.037	5.589	34.051	1.00	83.04	C
ATOM	2343	C	THR	A	304	-40.429	4.204	33.814	1.00	78.65	C
ATOM	2344	O	THR	A	304	-40.511	3.349	34.700	1.00	79.54	O
ATOM	2345	CB	THR	A	304	-42.140	5.955	33.037	1.00	83.63	C
ATOM	2346	OG1	THR	A	304	-43.200	4.988	33.088	1.00	85.88	O
ATOM	2347	CG2	THR	A	304	-41.611	6.143	31.607	1.00	79.08	C
ATOM	2348	N	PHE	A	305	-39.779	3.983	32.668	1.00	77.77	N
ATOM	2349	CA	PHE	A	305	-39.123	2.709	32.349	1.00	80.32	C
ATOM	2350	C	PHE	A	305	-40.037	1.518	32.626	1.00	79.61	C
ATOM	2351	O	PHE	A	305	-39.623	0.503	33.191	1.00	76.29	O
ATOM	2352	CB	PHE	A	305	-37.799	2.567	33.104	1.00	75.88	C
ATOM	2353	CG	PHE	A	305	-36.716	1.894	32.306	1.00	75.68	C
ATOM	2354	CD1	PHE	A	305	-36.485	2.260	30.994	1.00	79.28	C
ATOM	2355	CD2	PHE	A	305	-35.930	0.902	32.865	1.00	71.73	C
ATOM	2356	CE1	PHE	A	305	-35.492	1.649	30.250	1.00	79.40	C
ATOM	2357	CE2	PHE	A	305	-34.931	0.285	32.126	1.00	73.41	C
ATOM	2358	CZ	PHE	A	305	-34.712	0.661	30.818	1.00	75.84	C
ATOM	2359	N	GLN	A	306	-41.298	1.653	32.232	1.00	80.67	N
ATOM	2360	CA	GLN	A	306	-42.296	0.612	32.439	1.00	77.22	C
ATOM	2361	C	GLN	A	306	-42.032	-0.599	31.550	1.00	70.87	C
ATOM	2362	O	GLN	A	306	-42.774	-1.581	31.583	1.00	68.39	O
ATOM	2363	CB	GLN	A	306	-43.702	1.159	32.175	1.00	77.17	C
ATOM	2364	CG	GLN	A	306	-44.077	1.275	30.699	1.00	78.18	C
ATOM	2365	CD	GLN	A	306	-43.097	2.109	29.893	1.00	78.00	C
ATOM	2366	OE1	GLN	A	306	-42.557	3.100	30.385	1.00	83.17	O
ATOM	2367	NE2	GLN	A	306	-42.864	1.710	28.648	1.00	71.18	N
TER	2368		GLN	A	306						

Figura 17: Fragmento del fichero 6lu7.pdb

Habitualmente, antes de trabajar con estos archivos, se suele “limpiar” la estructura de moléculas de agua de cristalización y otros grupos a priori no deseados en el estudio. No obstante, se ha reportado⁹ la participación de moléculas de agua en los pasos del mecanismo de la adición de Michael, por lo que se ha trabajado con el archivo en bruto, que no contiene otras moléculas que interfieran.

Se separan las líneas del archivo que se refieren a N3 de las que hacían referencia al agua de cristalización y M^{Pro}, obteniendo así dos ficheros que contienen las dos moléculas por separado. El fichero de la proteína, que no necesita más tratamiento, se envía al programa GROMACS para obtener la topología de M^{Pro} y las aguas de cristalización, eligiendo el campo de fuerzas CHARMM36, que se ajusta bien al sistema.

Una vez obtenida la topología de la proteína, se procede a obtener la del ligando. El fichero anterior se edita para añadir los hidrógenos al esqueleto carbonado que se obtuvo del PDB inicial. Se edita el fichero de texto que se obtiene para el programa no confunda las etiquetas que definen las partes de N3 y se guardan los cambios. Este fichero se envía a la herramienta online SwissParam²² del Instituto Suizo de

Bioinformática, donde se procesa hasta obtener varios archivos, de los que interesa el PDB y el ITP.

Una vez obtenidos estos ficheros, se puede editar la topología de la proteína que se ha obtenido previamente para añadir la topología del ligando que proporciona SwissParam, con lo que se tiene la topología del complejo. En este punto, se restringen además las posiciones del ligando y la proteína para facilitar el cálculo.

El siguiente paso es crear las PBC, la caja en la que se simulará la molécula, y solvatar la caja, añadiendo iones Cl^- o Na^+ que neutralizarán la carga del sistema. Una vez añadidos, el sistema ya está listo para realizar la minimización energética, para lo cual se prepara un fichero que contiene las órdenes para que el programa realice un steepest descent de un máximo de 50000 pasos de 0.1 kJ/mol de energía, aplicando cut-offs. Una vez terminada, se comprueban los ficheros de salida y se visualizan para comprobar que el resultado es adecuado.

A continuación, se realiza una segunda restricción de posiciones al inhibidor y se modifica la topología del complejo para que se tenga en cuenta. Además, para obtener un buen control de la temperatura, se establecen dos parejas de moléculas: una formada por el ligando y la proteína y otra formada por el disolvente y los iones de solvatación. Así, los termostatos se aplican de manera que a cada grupo de moléculas similares se le aplica el termostato más adecuado, lo que facilita el cálculo. Una vez resuelto este punto, se prepara un fichero con las órdenes para realizar la simulación NVT a 300K. Una vez terminada la simulación NVT, se prepara el fichero para la posterior simulación NPT y se lanza.

Con estos pasos se ha obtenido un sistema que representa un complejo proteína-ligando minimizado energéticamente a la temperatura de 300K, por lo que el sistema está listo para realizar la etapa de producción. Este cálculo es mucho más largo que los dos anteriores, ya que se pasa de 50.000 a 5.000.000 de pasos.

Una vez obtenido el cálculo de Dinámica molecular, se obtienen dos ficheros, XTC y TPR, con los que ya se puede trabajar para obtener las distancias, ángulos y número de enlaces de hidrógeno del complejo proteína-ligando, que son el objeto de estudio.

PROGRAMAS Y HERRAMIENTAS UTILIZADAS

A continuación, se detallará brevemente las características de las herramientas informáticas más importantes utilizadas para la realización de este trabajo.

GROMACS²¹

Como se ha mencionado anteriormente, GROMACS (Groningen MACHine for Chemical Simulation) es un software que ha servido para realizar los cálculos de Dinámica Molecular de este Trabajo de Fin de Grado. Se diseñó principalmente para realizar cálculos en sistemas de lípidos, ácidos nucleicos y proteínas, aunque su aplicación se ha extendido a otras áreas con el paso del tiempo y la mejora de los cálculos de interacciones no enlazantes.

Este software es gratuito sin perder calidad o rendimiento en los cálculos, por lo que usuarios de todo el mundo lo utilizan habitualmente. Este uso tan extendido implica que se actualiza continuamente, y además existe una gran cantidad de foros, tutoriales y herramientas en Internet que facilita el uso de GROMACS. La existencia de tanta información ayuda a solventar los obstáculos que un usuario poco experimentado puede encontrar, como ha sucedido en la realización de este trabajo.

BIOVIA DISCOVERY STUDIO²³

Es una herramienta de modelado y simulación de moléculas, especialmente orientada a las proteínas y la optimización de fármacos. Permite obtener y visualizar macromoléculas, así como datos de minimización energética, enlaces, interacciones, esquemas 2D,

Es una herramienta muy potente que se ha utilizado para la visualización del sistema M^{Pro}-N3, puesto que proporciona una imagen definida e interactiva con la que se han estudiado los sitios activos, interacciones ligando-proteína, secuencias de aminoácidos importantes en el sistema, etc. También es un software gratuito, por lo que, al igual que ocurría con GROMACS, hay gran cantidad de recursos que han facilitado su manejo.

Visual Molecular Dynamics (VMD)²⁴

Es otra herramienta de visualización, que se ha utilizado principalmente durante la fase de los cálculos de Dinámica y Mecánica Molecular para corroborar que los ficheros obtenidos ofrecían datos coherentes, así como para analizar las trayectorias de la dinámica.

SWISSPARAM²²

Es una herramienta online, ofrecida por el Instituto Suizo de Bioinformática, que ha servido para realizar la parametrización del inhibidor N3. El fichero que contiene la estructura cristalina del sistema M^{Pro}-N3 consideraba el inhibidor y la proteína como un todo, por lo que ha sido necesario utilizar SwissParam para obtener los ficheros derivados de la parametrización de N3. Posteriormente, se combinaron con el fichero de M^{Pro} para obtener un archivo que contiene el complejo parametrizado. La web incorpora un tutorial muy sencillo para el uso de esta herramienta, por lo que resulta muy fácil de utilizar.

RESULTADOS

Una vez realizado el cálculo y obtenidos los ficheros necesarios para el análisis del sistema, el primer paso es la comprobación de que la simulación se ha llevado a cabo correctamente. Para ello, se atiende, principalmente, a la representación de la RMSD (Root Mean Square Deviation) respecto al tiempo de simulación, que ha sido de 100 ns.

La Gráfica 1 representa la evolución temporal de la RMSD. Idealmente, la representación debería mostrar una etapa de estabilización muy rápida a partir del tiempo cero manteniéndose la gráfica alrededor de este un valor hasta el final del intervalo del tiempo. Sin embargo, se aprecia una gráfica con dos picos muy marcados en las regiones de 40-45 y 77-82 ns.

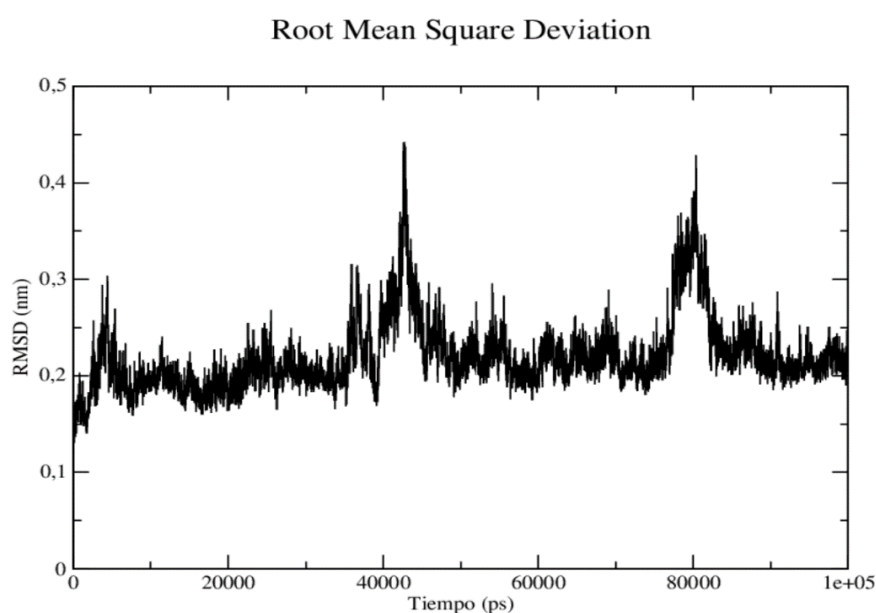


Gráfico 1. Evolución de la RMSD con el tiempo

No obstante, estos dos picos no tienen por qué indicar una mala simulación, sino que puede que, por movimientos naturales del complejo, alguno de los valores se aleje de la norma.

Por seguridad, se observaron las estructuras en la totalidad de dichos intervalos utilizando VMD. Se comprobó que el complejo no tomaba ninguna conformación extraña ni presentaban anomalías que pudieran deberse a la simulación. Se consideró que las fluctuaciones de la gráfica corresponden a dos intervalos de estructuras en las

que la hendidura del sitio activo está ligeramente más abierta que en el resto de la simulación, hecho que no afecta a la validez de los datos obtenidos. Prueba de ello son estas dos imágenes del complejo, que representan las estructuras del complejo en los tiempos exactos de los dos valores máximos de la gráfica de RMSD, y en las que se puede comprobar que las estructuras no son apenas distinguibles de la de la *Figura 5*.

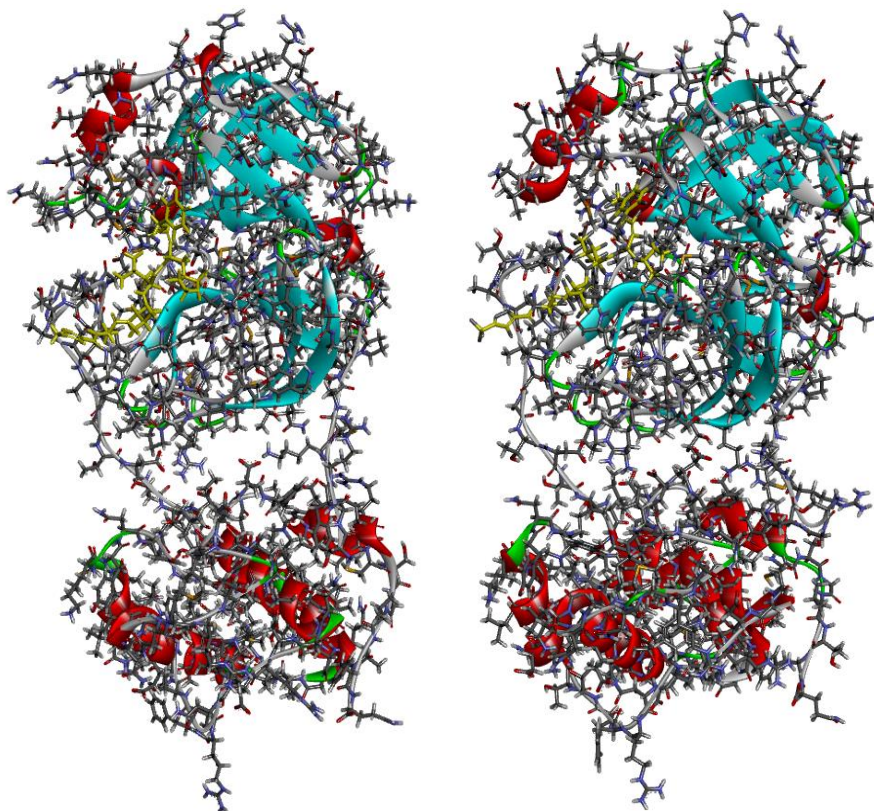


Figura 18. Estructuras de los dos máximos de la Gráfica 2

A la vista de estos resultados, se puede aseverar que el sistema se comporta de manera estable, por lo que los datos obtenidos de su análisis son válidos.

Tras comprobar que la simulación se ha realizado de manera correcta y, por tanto, los datos de la trayectoria son válidos, se procedió a analizar los enlaces de hidrógeno del complejo en el sitio activo.

El comando *gmx hbond* permite a GROMACS analizar la trayectoria que ha generado la dinámica molecular, obteniéndose varias representaciones entre las que se encuentran el número de enlaces de hidrógeno, la distribución de ángulos de esos enlaces de hidrógeno, y la distribución de distancias entre el donador y el aceptor del enlace de hidrógeno. Además, permite obtener un archivo index en el que se encuentran todos los enlaces hidrógeno que se producen tanto entre la proteína y el ligando como los enlaces intermoleculares de la proteína.

A continuación, se presentan dos gráficas, que corresponden a la distribución de distancias Donor-Aceptor y los ángulos Hidrógeno...Donor-Aceptor. De acuerdo con la bibliografía²⁵, se aplicó un cut-off de 0,35 nm a la distancia y de 30° al ángulo del enlace de hidrógeno.

Distribucion de distancias de Enlace de Hidrogeno

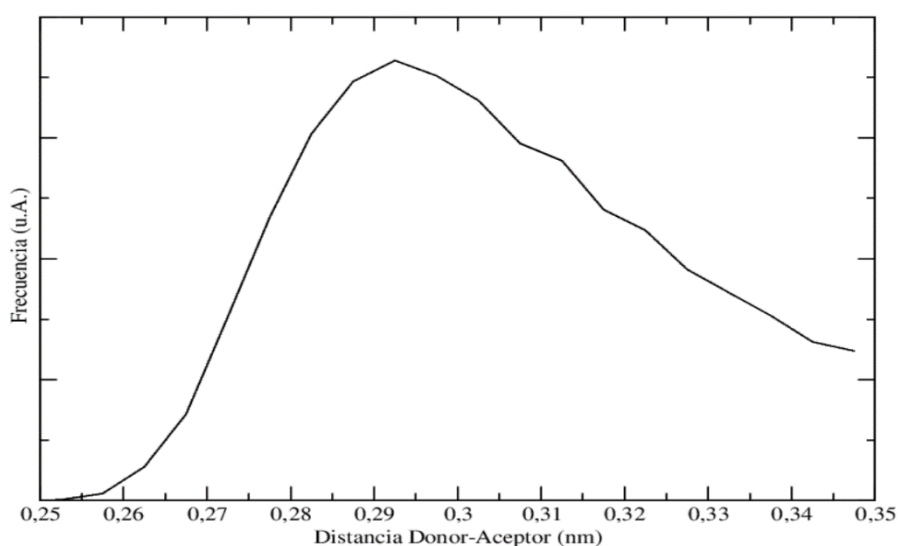


Gráfico 2. Distribución de distancias donador-aceptor en los enlaces de hidrógeno

Distribucion de angulos de Enlace de Hidrogeno

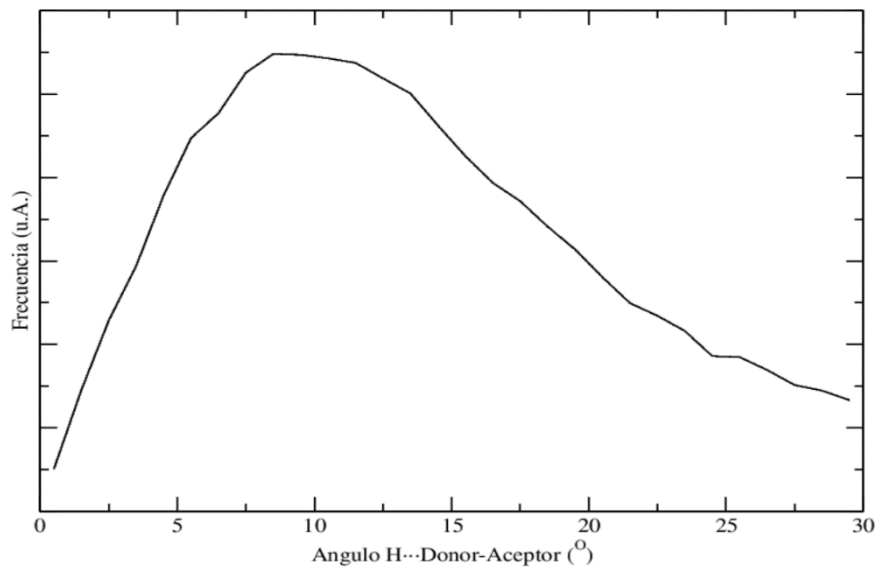


Gráfico 3. Distribución de ángulos de los enlaces de hidrógeno

Los máximos de estas representaciones corresponden a 0,29 nm y 8,5°, respectivamente.

El siguiente gráfico representa el cambio del número de enlaces de hidrógeno al avanzar en la trayectoria.

Numero de enlaces de Hidrogeno

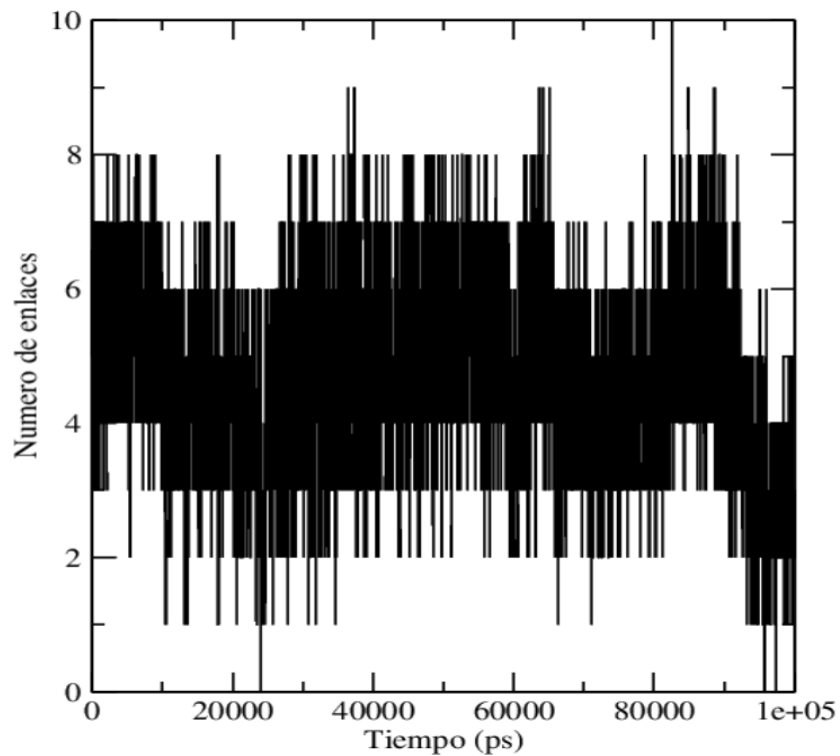


Gráfico 4. Evolución temporal del número de enlaces de hidrógeno

A la vista del gráfico, el complejo pasa por varias conformaciones, que generan mayor o menor número de enlaces de hidrógeno en la región del sitio activo. Esto algo normal, ya que es un sistema dinámico donde los átomos se mueven continuamente, lo que provoca que estas interacciones se formen y rompan constantemente a lo largo del tiempo.

Hay un máximo de diez enlaces de hidrógeno y un mínimo de cero entre M^{pro} y N3, si bien hay muy pocas estructuras que presenten menos de dos enlaces de hidrógeno o más de ocho. La mayoría de las estructuras de la trayectoria poseen entre cuatro y seis enlaces de hidrógeno. El siguiente histograma muestra el conteo de cada número de enlaces en todo el intervalo de tiempo, indicando que la situación que más se da en el complejo es la de cinco enlaces de hidrogeno.

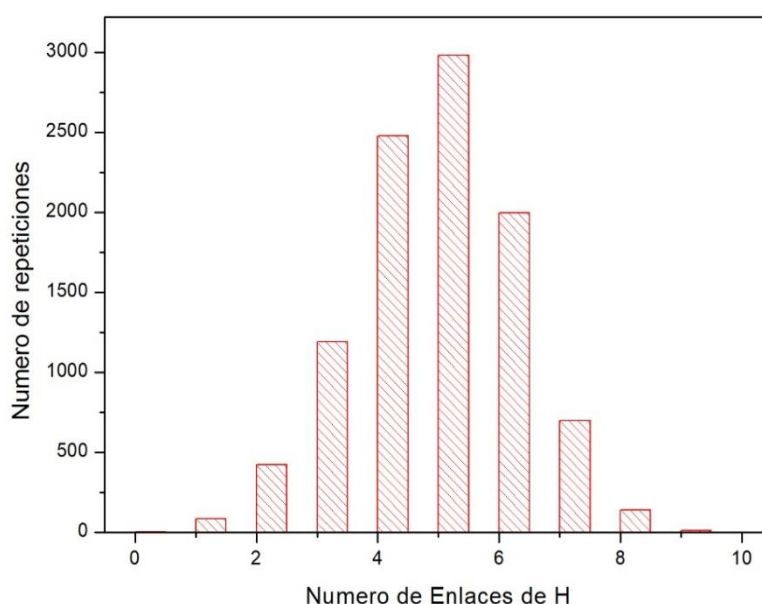


Gráfico 5. Histograma de distribución del número de enlaces de hidrógeno

Para ilustrar algunas de las estructuras por las que ha pasado el complejo durante esta simulación, se ha tomado la situación que presenta cinco enlaces de hidrógeno para estudiar la estructura e identificar los aminoácidos que participan en los enlaces de hidrógeno.

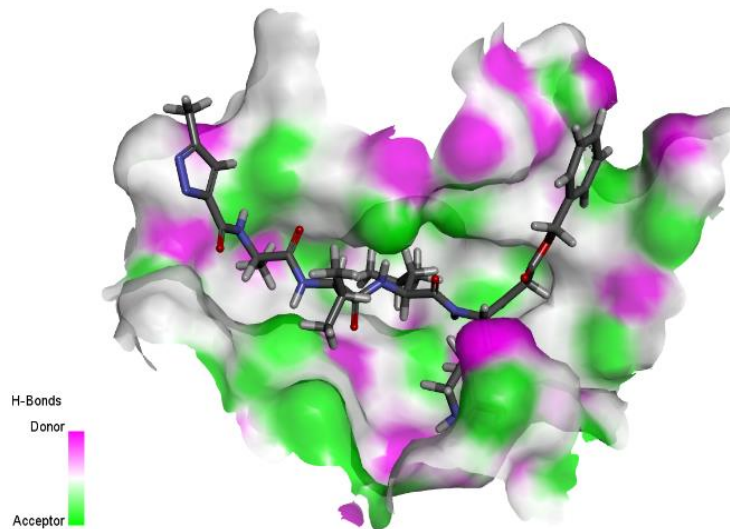


Figura 19. Estructura 3D del sitio activo y su comportamiento donador-aceptor de hidrógeno

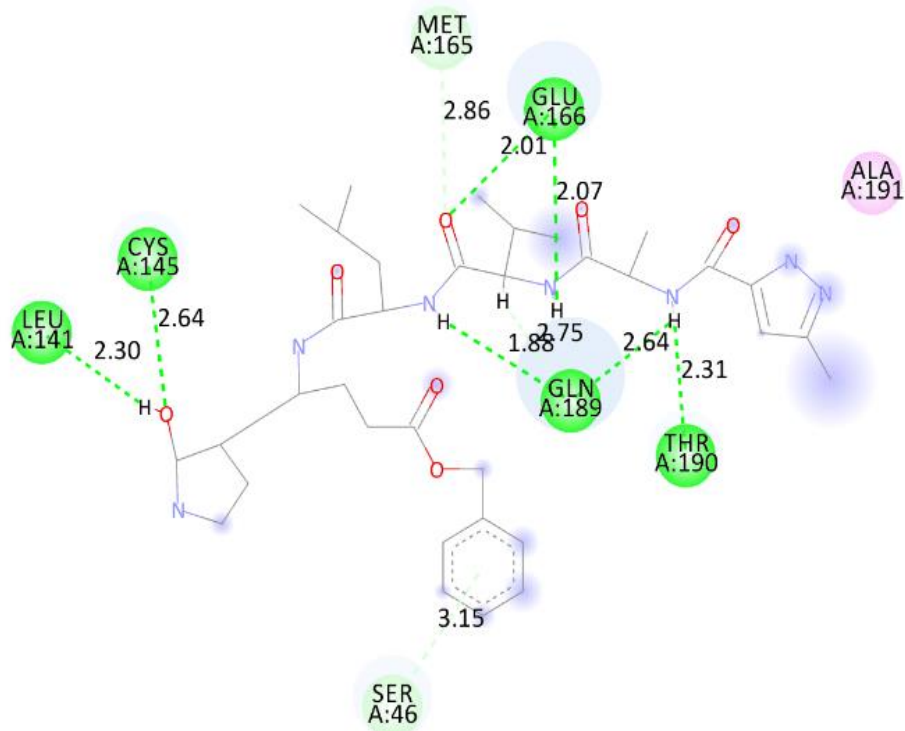


Figura 20. Diagrama 2D de los enlaces de hidrógeno presentes en la estructura de la Figura 19

El comando *gmx hbond* permite también obtener un diagrama de recurrencia de enlaces de hidrógeno en el que se representan todos los enlaces de hidrógeno que se producen entre el ligando y la proteína, etiquetados de acuerdo con el orden establecido en el archivo index. En la representación, para cada uno de los nanosegundos de la simulación se marca una raya roja en aquellas etiquetas cuyo enlace de hidrógeno esté presente para ese tiempo concreto. Este gráfico es particularmente ilustrativo, puesto que

permite obtener la estabilidad de los enlaces de hidrógeno que se producen en el complejo a lo largo de todo el tiempo de la simulación.

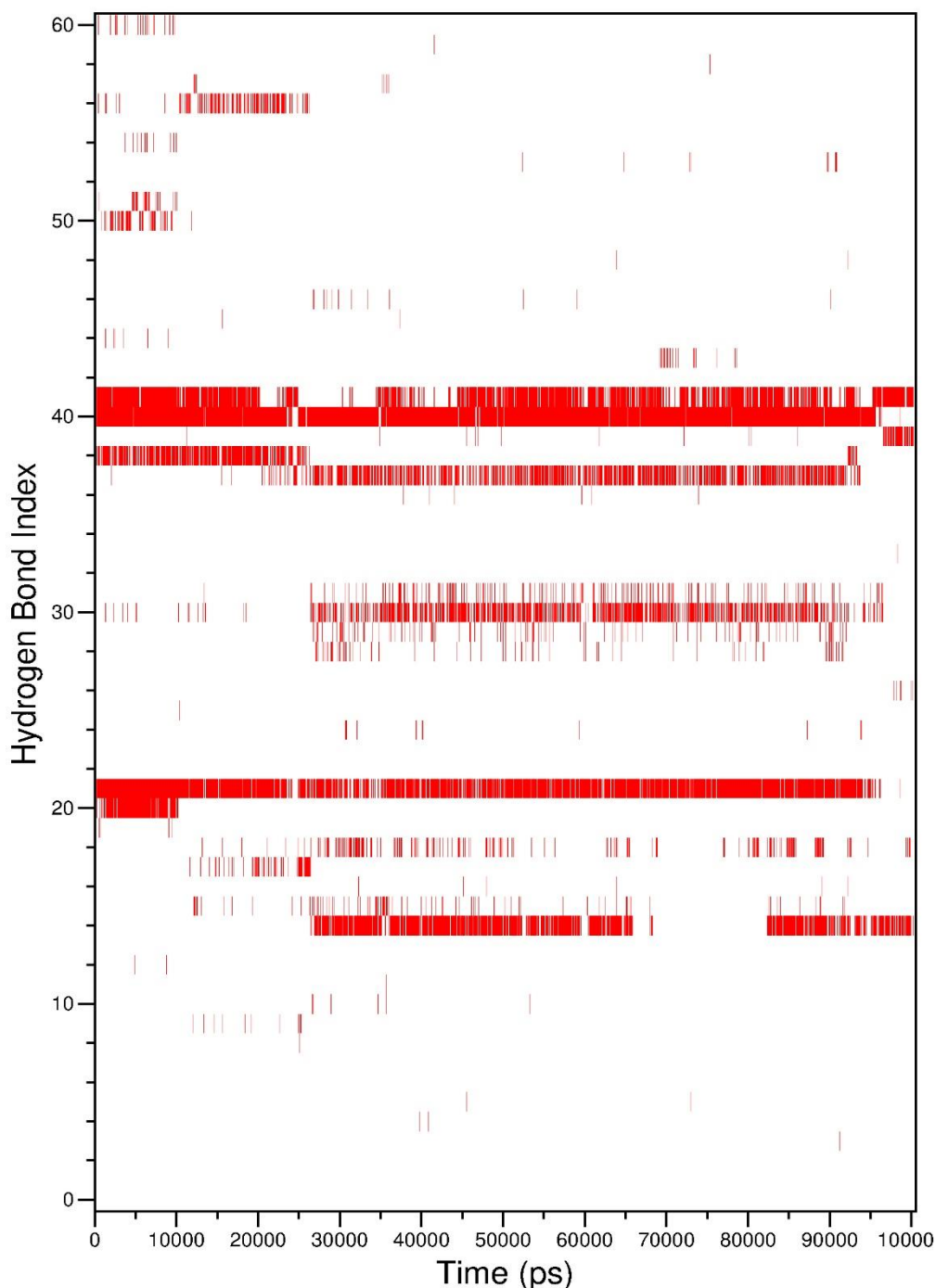


Gráfico 6. Diagrama de recurrencia de enlaces de hidrógeno

Utilizando un programa auxiliar a GROMACS²⁶, se ha obtenido la siguiente tabla, que recopila los enlaces de hidrógeno más estables en el tiempo, así como información sobre los aminoácidos de la proteína y átomos del ligando que participan en esas interacciones. También se proporciona el porcentaje del tiempo de simulación en el que esos enlaces de hidrógeno están presentes.

Tabla 1. Resultados de la gráfica de recurrencia de enlaces de hidrógeno

ETIQUETAS	ÁTOMO 1	ÁTOMO 2	ÁTOMO 3	% RECURRENCIA
14	N (ASN 142)	H (ASN 142)	O4 (LIG)	45,66
20	N E2 (HIS 163)	H (HIS 163)	N6 (LIG)	8,41
21	N E2 (HIS 163)	H (HIS 163)	O8 (LIG)	83,47
30	N (ALA 191)	H (ALA 191)	O1 (LIG)	42,03
37	N (LIG)	H4 (LIG)	O (ARG 188)	45,04
38	N (LIG)	H4 (LIG)	O (GLN 189)	19,72
40	N (LIG)	H9 (LIG)	O (GLN192)	92,92
41	N1 (LIG)	H18 (LIG)	O (GLU 166)	66,77

INTERPRETACIÓN DE LOS RESULTADOS Y COMPARACIÓN CON DATOS BIBLIOGRÁFICOS

En el caso de estudio, la media de cinco enlaces de hidrógeno entre el N3 y M^{pro} implica que este inhibidor está fuertemente anclado al sitio activo, lo que significa que el enlace entre el inhibidor y la proteína está muy favorecido.

El número de enlaces de hidrógeno no es el único criterio a la hora de considerar un inhibidor apto o no apto para un buen enlace con la proteína. Los aminoácidos concretos a los que se ancla el inhibidor también son un factor importante en la estabilidad de la interacción ligando proteína. Atendiendo a la *Tabla 1*, se infiere que los aminoácidos cuyo enlace de hidrógeno posee un porcentaje de recurrencia mayor participan más activamente en la fijación del ligando al sitio activo, por lo que estos aminoácidos representan los puntos de anclaje a los que se debe prestar atención a la hora de diseñar nuevos ligandos con capacidad inhibitoria de la M^{pro}. Así pues, se ha obtenido que los puntos de anclaje del ligando más relevantes son: Glicina 192 (92,92%), histidina 163 (83,47 y 8,41%), glutamato 166 (66,77%), asparagina 142 (45,66%), arginina 188 (45,04%), alanina 191 (45,04%) y glutamina 189 (19,72%). Los porcentajes de recurrencia indican que la glicina 192, histidina 163 y glutamato 166 son los aminoácidos cuyos enlaces de hidrógeno con el ligando más se repiten a lo largo de la simulación, con porcentajes por encima del 50%. ASN 142, ARG 188 Y ALA 191 son también significativos, con porcentajes entre el 42 y el 46%. Estos enlaces de hidrógeno más significativos se producen entre un nitrógeno y un oxígeno, hecho que se repite para todas las interacciones, con la excepción del enlace NHN de la etiqueta 20 (HIS 163, 8,41%). En el caso de la histidina, además, se observan dos enlaces de hidrógeno distintos, si bien es cierto que el segundo enlace es menos recurrente.

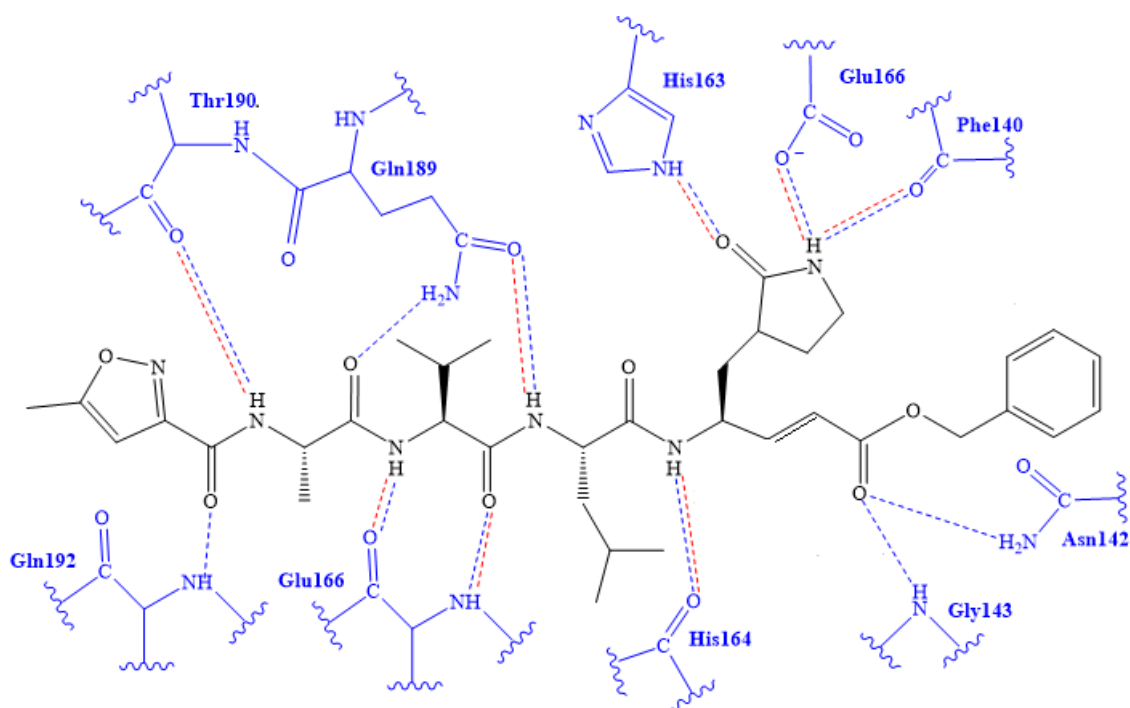


Figura 21. Diagrama de enlaces de hidrógeno en el sitio activo de la proteasa¹⁰

Esta lista de aminoácidos coincide con los aminoácidos que se enumeran en la bibliografía para interacciones de inhibidores con el sitio activo²⁷⁻²⁹, corroborando que son los anclajes moleculares a los que se debe prestar atención si se desea diseñar inhibidores de la proteasa M^{pro}. GLU 166 y GLN 192 se repiten particularmente como los aminoácidos de la proteína que presentan interacciones más significativas con el inhibidor descrito en cada caso, lo que confirma los datos obtenidos a este respecto. ALA 191, ASN 142 y ARG 188, por su parte, no tienen tanta representación en estos artículos como los dos aminoácidos anteriores. Sin embargo, es cierto que los datos obtenidos en este trabajo pueden considerarse más simples que los de las referencias, puesto que solo se refieren a un tipo de interacción entre el ligando y la proteína. Aun así, es revelador que, atendiendo a una sola de las características de este sistema (los enlaces de hidrógeno), los resultados obtenidos sí que coinciden, en mayor o menor medida, con estudios más exhaustivos y que abordan otra serie de conceptos además del enlace de hidrógeno.

Cabe destacar que estas referencias estudian ligandos concretos, con estructuras que difieren de la de N3, y el hecho de que no se obtengan exactamente los mismos datos que se han obtenido en este trabajo no significa que los resultados mostrados hasta ahora sean erróneos. Sin embargo, que aparezcan nombrados en los sitios activos de

Mpro en experimentos con otros ligandos sí que es significativo a la hora de dar valor a los resultados obtenidos en este trabajo.

Considerando el mecanismo de la reacción descrito en la introducción y atendiendo al *Gráfico 7* de recurrencia de enlaces de hidrógeno, cabe destacar que la cisteína 145 no es especialmente activa en los enlaces de hidrógeno detectados. El porcentaje de recurrencia del único enlace de hidrógeno que se ha computado en el que participa la CYS 145 (etiqueta 15) es solamente de un 0,2%, lo cual puede parecer contradictorio si se tiene en cuenta que la CYS 145 es precisamente el aminoácido en el que el inhibidor se va a enlazar con la proteína. La explicación de este hecho reside en que, como se ha explicado, el mecanismo zwitteriónico comienza con la cesión del protón de la CYS 145 a la HIS 41 de la proteína. El mismo razonamiento es aplicable a la HIS 41 (0,02% de recurrencia). Los cálculos computacionales que se han realizado se refieren exclusivamente a las interacciones ligando-proteína, por lo que una interacción intramolecular de la proteína como es la de la díada catalítica cisteína/histidina no es analizada mediante este método.

Atendiendo a los anclajes más relevantes, se puede comprobar que los enlaces de hidrógeno más significativos que se producen entre el ligando y la proteína afectan únicamente a la parte no activa del ligando en el enlace químico que provoca la inhibición de la proteasa. Esto concuerda con el concepto de “warhead” que se cita en la bibliografía¹⁰, y que define al ligando que inhibe a la proteasa como una molécula con dos partes: una parte no reactiva que fija al ligando en aminoácidos concretos del sitio activo y la región warhead, que posee un grupo de electrófilo responsable del enlace con M^{pro}. Modificando en la warhead el tipo de grupo electrófilo, se puede conseguir una inhibición reversible o irreversible. En el caso de estudio, no se dan interacciones no enlazantes significativas entre el grupo electrófilo de la warhead (doble enlace) y los aminoácidos estudiados, como se ha explicado anteriormente, lo que corrobora los resultados.

En lo referente a distancias y ángulos del enlace de hidrógeno, las *Gráficas 1* y *2* ponen de manifiesto que la mayoría de los enlaces de hidrógeno tienen ángulos entre los 2-25° y distancias de entre 0,27 y 0,35 nm. La bibliografía al respecto²⁵ indica que el motivo de tomar 3,5 Å (0,35 nm) como máximo de distancia de enlace se debe a que ésta es la

distancia entre el H y el O en el modelo Single Point Charge (SPC) de la molécula de agua a la que aparece el primer mínimo en la función de distribución radial. El valor más frecuente de la distancia del enlace es de 0,29 nm, que se encuentra dentro del máximo que marca la referencia.

En cuanto a los ángulos de enlace de hidrógeno, la situación es similar. En el programa de análisis se considera que el valor mínimo del ángulo H-Donor...Aceptor es de 30°.

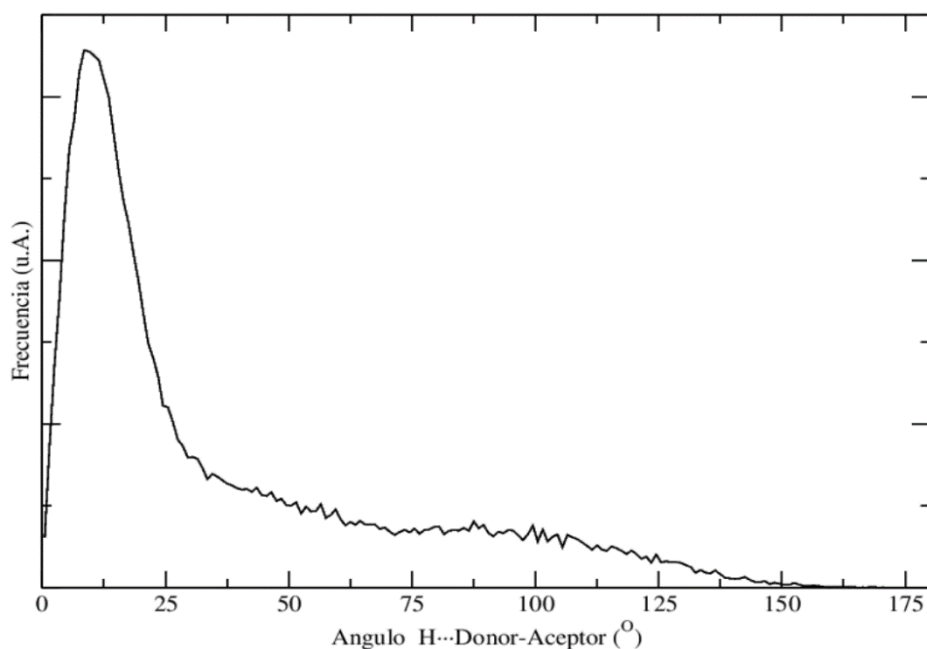


Gráfico 7. Distribución de ángulos del enlace de hidrógeno sin restricción

La Gráfica 8 muestra la distribución de enlaces de hidrógeno del sistema sin aplicar la restricción de los 30° al ángulo, es decir, con un ángulo libre entre 0 y 180°.

El marcado máximo de la representación se sitúa entre 2 y 30°, con un máximo en 8,5°. Este máximo del ángulo de enlace de hidrógeno coincide exactamente con el que proporciona la gráfica con valores restringidos, lo que indica que el criterio de 30° es perfectamente aplicable a la búsqueda de este tipo de interacciones. La bibliografía²⁵ explica que el enlace de hidrógeno es más fuerte cuanto más se acerca el ángulo a los 180°.

RESUMEN Y CONCLUSIONES

Se han estudiado mediante Dinámica Molecular las interacciones entre la proteasa principal del SARS-CoV-2, M^{pro}, y el inhibidor N3. Se ha contextualizado este virus, así como la enfermedad que provoca, y se ha explicado su mecanismo de infección.

Se ha hecho una explicación teórica de los fundamentos de la Mecánica y Dinámica Moleculares y se han resumido los pasos seguidos para la obtención de una simulación de dinámica molecular.

Se ha obtenido información sobre los enlaces de hidrógeno que forma N3 con M^{pro}, y se ha estudiado su número, distancia donador-aceptor, ángulo y aminoácidos que participan en ellos. Se ha comparado esta información con la bibliografía, comprobando que las conclusiones sobre los aminoácidos importantes y los enlaces de hidrógeno que se forman coinciden con los datos obtenidos de la simulación.

El hecho de que la mecánica molecular no permita el análisis de enlaces químicos no significa que no se puedan obtener conclusiones significativas observando únicamente los enlaces de hidrógeno. Los métodos de docking molecular para evaluar potenciales inhibidores utilizan el número de enlaces de hidrógeno que el ligando establece con la proteína como uno de los principales factores a la hora de determinar si un inhibidor está suficientemente estabilizado mediante estos enlaces de hidrógeno como reaccionar con M^{pro} 27-29. Así, incluso un número relativamente bajo de enlaces de hidrógeno en el sitio activo pueden representar una situación suficientemente favorable como para que los inhibidores que estudian estos artículos se consideren útiles para limitar la actividad de esta proteasa.

BIBLIOGRAFÍA

1. <https://www.who.int/es/emergencias/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19>.
2. Zheng, J. SARS-coV-2: An emerging coronavirus that causes a global threat. *International Journal of Biological Sciences* 16, 1678–1685 (2020).
3. Soto, G. P. Bases Genéticas y Moleculares del COVID-19 (SARS-CoV-2). Mecanismos de Patogénesis y de Respuesta Inmune Genetic and Molecular Basis of COVID-19 (SARS-CoV-2) Mechanisms of Pathogenesis and Immune. *Int. J. Odontostomat* vol. 14 (2020).
4. Sars-cov-, I. de. INFORMACIÓN CIENTÍFICA-TÉCNICA. (2021).
https://www.mscbs.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov/documentos/20210621_MICROBIOLOGIA.pdf
5. Duan, G. Intuition on virology, epidemiology, pathogenesis, and control of COVID-19. *Novel Research in Microbiology Journal* 4, 955–967 (2020).
6. Yang, H. et al. Design of wide-spectrum inhibitors targeting coronavirus main proteases. *PLoS Biology* 3, (2005).
7. <https://www.rcsb.org/structure/6LU7>
8. Jin, Z. et al. Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* 582, 289–293 (2020).
9. Ramos-Guzmán, C. A., Ruiz-Pernía, J. J. & Tuñón, I. Unraveling the SARS-CoV-2 Main Protease Mechanism Using Multiscale Methods. *ACS Catalysis* 10, 12544–12554 (2020).
10. Arafet, K. et al. Mechanism of inhibition of SARS-CoV-2 Mpro by N3peptidyl Michael acceptor explained by QM/MM simulations and design of new derivatives with tunable chemical reactivity. *Chemical Science* 12, 1433–1444 (2021).
11. J. Cramer, C. *Essentials of Computational Chemistry: Theories and Models, 2nd Edition*. (Wiley, 2004).
12. https://cmm.cit.nih.gov/intro_simulation/intro_simulation.pdf
13. Waidyasooriya, H. M., Hariyama, M. & Kasahara, K. An FPGA accelerator for molecular dynamics simulation using OpenCL. *International Journal of Networked and Distributed Computing* 5, 52–61 (2017).
14. Nutt, D. R. & Smith, J. C. Molecular dynamics simulations of proteins: Can the explicit water model be varied? *Journal of Chemical Theory and Computation* 3, 1550–1560 (2007).

15. Gkeka, P. *Molecular dynamics studies of peptide-membrane interactions: insights from coarse-grained models*. (2010). <http://hdl.handle.net/1842/4652>.
16. https://upload.wikimedia.org/wikipedia/commons/thumb/8/8c/Bitruncated_Cubic_Honeycomb.svg/500px-Bitruncated_Cubic_Honeycomb.svg.png
17. <https://slideplayer.com/slide/6947592/>
18. <https://morioh.com/p/e5b706f79119>
19. <https://culturacientifica.com/2017/09/12/la-distribucion-velocidad-de-moleculas-maxwell/>
20. Kim, T. W. et al. Protein folding from heterogeneous unfolded state revealed by time-resolved X-ray solution scattering. *Proceedings of the National Academy of Sciences of the United States of America* 117, 14996–15005 (2020).
21. <https://manual.gromacs.org>
22. <https://www.swissparam.ch/>
23. <https://www.addlink.es/productos/discovery-studio>
24. <https://www.ks.uiuc.edu/Research/vmd/>
25. Jeffrey, G. A. *An introduction to hydrogen bonding*. (Oxford University Press, 1997).
26. <https://osf.io/3wunt/>
27. Gurung, A. B., Ali, M. A., Lee, J., Farah, M. A. & Al-Anazi, K. M. Unravelling lead antiviral phytochemicals for the inhibition of SARS-CoV-2 Mpro enzyme through in silico approach. *Life Sciences* 255, 117831 (2020).
28. Ibrahim, M. A. A., Abdeljawaad, K. A. A., Abdelrahman, A. H. M. & Hegazy, M. E. F. Natural-like products as potential SARS-CoV-2 Mpro inhibitors: in-silico drug discovery. *Journal of Biomolecular Structure and Dynamics* 0, 1–13 (2020).
29. Khan, S. L., Siddiqui, F. A., Jain, S. P. & Sonwane, G. M. Discovery of Potential Inhibitors of SARS-CoV-2 (COVID-19) Main Protease (Mpro) from Nigella Sativa (Black Seed) by Molecular Docking Study. *Coronaviruses* 2, 384–402 (2020).