

Trabajo de Fin de Grado

**INFLUENCIA DEL ENTORNO
PROTEICO DE LA ENZIMA
LUCIFERASA EN EL ESPECTRO DE
EMISIÓN DE LA OXILUCIFERINA**

Autor: Manuel Sánchez-Beato Puebla

Curso 2023-2024

Departamento de Química

ÍNDICE

1. Abstract	2
2. Resumen	3
3. Introducción.....	4
4. Objetivos.....	7
5. Plan de trabajo	8
6. Métodos	8
6.1. Dinámica Molecular Clásica	9
6.2. Campo de Fuerzas.....	10
6.3. Condiciones Periódicas de Contorno	11
6.4. Propiedades Termodinámicas	12
6.5. Energía Libre de Interacción	12
6.6. Teoría del Funcional de la Densidad y su Variante Dependiente del Tiempo.....	13
6.7. Mecánica Cuántica/Mecánica Molecular	14
7. Resultados y Discusión.....	15
7.1. Montaje y Descripción del Sistema	15
7.2. Dinámica Molecular.....	16
7.3. Cálculo de la Energía Libre de Interacción.....	19
7.3.1. Cálculo de la Energía Libre de Interacción con Geometrías Equidistantes	20
7.3.2. Cálculo de la Energía Libre de Interacción con Geometrías Representativas de los Clústeres.....	23
7.4. Espectro de Emisión.....	26
8. Conclusiones	27
9. Bibliografía.....	28

1. Abstract

There are compounds in nature capable of producing the bioluminescence phenomenon, which consists of the emission of a photon as a result of a chemical reaction between a chromophore and an enzyme. D-luciferin is a chromophore, which can be found in some firefly species, that has attracted researchers' attention due to its properties. A wide variety of applications have been developed in the field of biomedicine for the luciferin/luciferase system, including tissue bioimaging, which is one of the most promising applications. The development of bioluminescence applications requires the study of the chromophores in order to control their emission wavelength.

This work examines the interactions between the protein environment and the oxyluciferin emitter, specifically using the oxyluciferin-keto (-1) tautomer. The goal is to identify the amino acids that interact most with the chromophore, in order to potentially mutate them in the future to modulate the emission wavelength. To do so, a computational simulation of the system was evolved by classical Molecular Dynamics (MD) in the first state. Then, the binding free energy between the enzyme and the substrate was computed using the Molecular Mechanics with Generalized Born Surface Area (MMGBSA) method by selecting different frames by two different methods. The first protocol computes the binding free energy by considering a large number of equidistant geometries along the trajectory, while the second protocol groups the trajectory geometries into clusters, based on the similarity of the chromophore structure, and calculates a representative geometry for each cluster for which the free energy of interaction is computed.

The importance of the protein environment is studied by decomposing the binding free energy into individual contributions from each amino acid, considering those residues that are in contact with oxyluciferin more than 5% of the time along the trajectory. In this way, the amino acids with the greatest energy contribution were identified, and the interactions between the amino acid and the chromophore were characterized. Finally, by using mixed Quantum Mechanics/Molecular Mechanics (QM/MM) calculations, in which the QM region was described by Time-Dependent Density Functional Theory (TD-DFT), the emission spectrum of the oxyluciferin/luciferase system was calculated, showing a good agreement with the experimental one.

In future studies, mutations of the amino acids with the highest contribution to the binding free energy can be analyzed to study the effect on the emission spectrum. The ideal goal would be to achieve a bathochromic shift of the emitted wavelength so that it is in line with the transparency window in mammalian tissues and improves performance in bioimaging applications.

2. Resumen

Existen compuestos en la naturaleza capaces de producir el fenómeno de bioluminiscencia, que consiste en la emisión de un fotón como resultado de una reacción química entre un cromóforo y una enzima. La D-luciferina es un cromóforo, que se puede encontrar en algunas especies de luciérnagas, que ha atraído la atención de los investigadores debido a sus propiedades. Se han desarrollado una gran variedad de aplicaciones en el campo de la biomedicina para el sistema D-luciferina/luciferasa, siendo la aplicación en bioimagen de tejidos una de las aplicaciones más prometedoras. El desarrollo de aplicaciones de la bioluminiscencia requiere el estudio de los cromóforos empleados para modular su longitud de onda de emisión.

Este trabajo estudia las interacciones entre el entorno proteico y la oxiluciferina en el primer estado excitado, concretamente usando el tautómero oxiluciferina phenolato-ceto (-1). El objetivo es determinar los aminoácidos que más interactúan con la oxiluciferina, para, en un futuro, realizar mutaciones sobre ellos para modular la longitud de onda de emisión. Para ello, se realizó una simulación computacional del sistema sobre la que se han realizado cálculos de Dinámica Molecular clásica (MD) en el primer estado excitado. Después, se calculó la energía libre de interacción entre la enzima y el sustrato utilizando el método *Molecular Mechanics With Generalized Born Surface Area* (MMGBSA), seleccionando diferentes geometrías mediante dos protocolos distintos. El primer protocolo computa la energía libre de interacción usando un gran número de geometrías equidistantes a lo largo de la trayectoria, mientras que en el segundo protocolo se agrupan las geometrías de la trayectoria en clústeres, bajo el criterio de similitud estructural del cromóforo, sobre los que se calcula una geometría representativa de cada uno que se usan para calcular la energía libre de interacción.

La importancia del entorno proteico se estudia mediante la descomposición de la energía libre de interacción en contribuciones individuales de cada aminoácido, teniendo en cuenta aquellos residuos que pasan más de un 5% del tiempo en contacto con el cromóforo a lo largo de la trayectoria. De esta forma se ha podido identificar los aminoácidos con mayor contribución energética y caracterizar las interacciones que se dan entre cada aminoácido y el cromóforo. Finalmente, se realizaron cálculos mixtos de Mecánica Cuántica y Mecánica Molecular (QM/MM), donde se usó la Teoría del Funcional de la Densidad Dependiente del Tiempo (TD- DFT) para describir la región QM, para el espectro de emisión del sistema oxiluciferina/luciferasa, el cual concuerda con el espectro experimental.

En el futuro, se pueden analizar mutaciones de los aminoácidos más contribuyentes a la energía libre de interacción con el fin de estudiar el efecto que tendría en el espectro de emisión. Lo ideal sería conseguir un desplazamiento batocrómico de la longitud de onda emitida, de forma que esté acorde con la ventana de transparencia en tejidos de mamíferos y mejore el rendimiento en bioimagen.

3. Introducción

La bioluminiscencia es un proceso biológico mediante el cual organismos vivos son capaces de emitir energía en forma de luz visible. Este proceso de emisión es el resultado de una reacción química que tiene lugar dentro del organismo, y que generalmente implica la oxidación de un sustrato denominado luciferina por una enzima llamada luciferasa. El color que se visualiza depende de la energía del fotón emitido, lo cual varía dependiendo del organismo y el sistema luciferina/luciferasa que éste utilice.^[1]

Hasta la fecha se ha determinado la estructura de nueve luciferinas,^[1] de las cuales el ejemplo más representativo corresponde a la D-luciferina (Figura 1a), molécula que podemos encontrar en insectos como la luciérnaga norteamericana (*Photinus pyralis*), escarabajos, como el escarabajo chasqueador (*Pyrophorus plagiophthalmus*) y otro tipo de insectos como el gusano ferroviario (*Phrixothrix hirtus*). Otros cromóforos bioluminiscentes son la coelenteracina (Figura 1b) y la cypiridina luciferina^[2] (Figura 1c), ambos son tripéptidos modificados capaces de emitir en la región de luz azul (entorno a los 490 nm). Dichas moléculas se encuentran en la mayoría de los organismos marinos que presentan bioluminiscencia. También existen otros organismos capaces de sintetizar luciferinas como el krill y los organismos dinoflagelados, además de los hongos con luciferina fúngica^[3] (Figura 1e) y bacterias con luciferina bacteriana (Figura 1d).

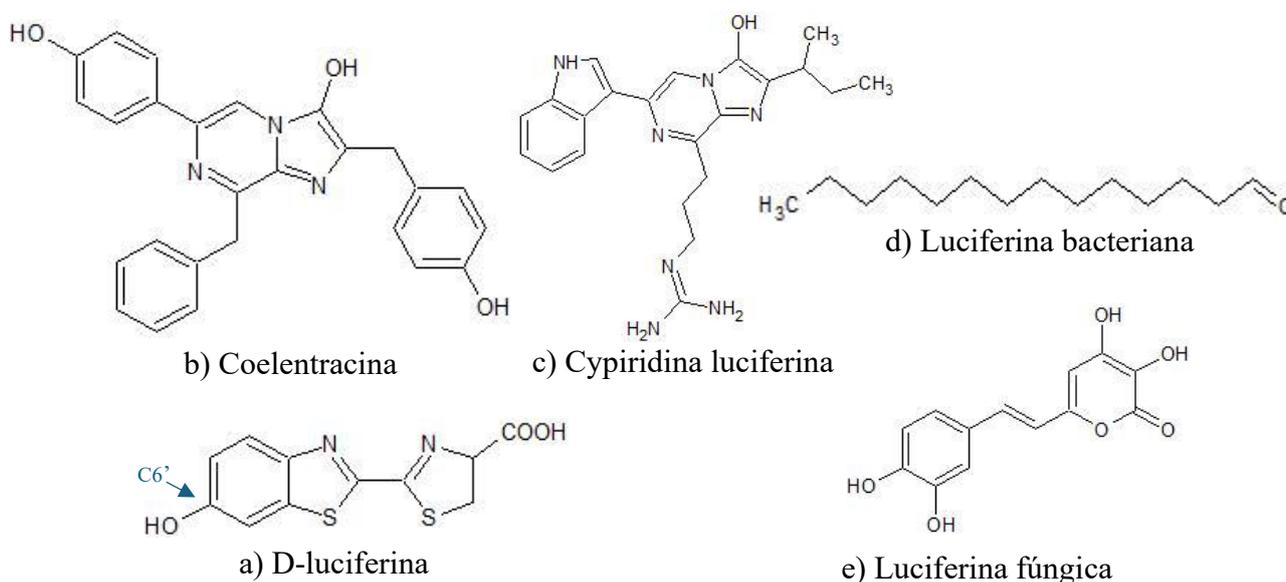


Figura 1. Representación de distintos cromóforos: a) Coelenteracina b) D-luciferina c) Cypiridina luciferina d) Luciferina bacteriana e) Luciferina fúngica.^[1]

Pese a la gran cantidad de posibles aplicaciones que tienen estos sistemas, las luciferinas naturales en disolución presentan la posibilidad de dimerización en presencia de un medio básico, lo que puede afectar a la reacción de bioluminiscencia.^[4] Además, para el uso en aplicaciones de bioimagen de tejidos de mamíferos la luz que mejor penetra los tejidos biológicos se encuentra en la región del rojo

(entorno a los 750 nm) y del infrarrojo cercano, por lo que es un inconveniente que la longitud de onda de emisión de las luciferinas naturales se encuentre en la región del espectro visible del verde-amarillo (520-580 nm). Es por estas cuestiones que se han llevado a cabo estudios con luciferinas sintéticas con el objetivo de aumentar su estabilidad y desplazar la emisión de luz hacia el rojo. Las aminoluciferinas fueron los primeros derivados sintéticos de la D-luciferina,^[5] los cuales se sintetizaron cambiando el grupo hidroxilo en el carbono 6' en el fragmento de benzotiazol (ver Figura 1a). Se ha comprobado que la adición del grupo amino no conlleva la pérdida de la bioluminiscencia; además, la emisión de luz sufre un desplazamiento hacia longitudes de onda mayores respecto a la D-luciferina natural.^[6] Si bien en un principio se pensó que la presencia del grupo benzotiazol era esencial para que la bioluminiscencia tuviera lugar, hay estudios que muestran que son posibles las modificaciones de dicho fragmento. Esto ha llevado a la investigación de derivados de la D-luciferina con distintos sistemas aromáticos,^[7] que ha llevado al descubrimiento de cromóforos que emiten en el rango desde el azul hasta el rojo en función de la longitud de la cadena π conjugada. Por ejemplo, en derivados de cumarina y quinoleno^[8] se observan desplazamientos hipsocrómicos (desplazamiento hacia menor longitud de onda) y batocrómicos (desplazamiento hacia mayor longitud de onda) respectivamente en la frecuencia del fotón emitido.

Las luciferasas son las enzimas que catalizan las reacciones que producen la bioluminiscencia, y en ellas también se encuentra una gran variabilidad. En la naturaleza, encontramos tres grandes familias^[5] de insectos capaces de producir luciferasas naturales que interactúan con la D-luciferina produciendo la emisión de luz. Por un lado, está la luciferasa de las luciérnagas, que además fue la primera en ser descubierta en la especie de luciérnaga norteamericana (*Photinus pyralis*), perteneciente a la familia *Lampyridae*. Esta luciferasa es capaz de interactuar con la D-luciferina emitiendo luz entre los 540-570 nm, correspondiéndose a colores verdes amarillentos.^[5] Por otro lado, se encuentra la luciferasa de los escarabajos chasqueadores (como el *Pyrophorus plagiophthalmus*) de la familia *Elateridae*, cuya emisión de luz abarca el rango de los 540 a 600 nm encontrando una gama de colores desde el verde hasta el naranja.^[5,9] Por último, los gusanos ferroviarios (como el *Phrixothrix hirtus*) constituyen la tercera familia, denominada *Phengodidae*, capaz de sintetizar una luciferasa natural. El rango de emisión de luz de dicha luciferasa se encuentra entre los 540 nm (verde) y los 630 nm (rojo).^[5]

Como se puede observar en la Figura 2, en el entorno enzimático la D-luciferina (1) reacciona con adenosín trifosfato (ATP), resultando en una molécula de luciferil adenilato (2), siendo éste un intermedio importante del mecanismo de bioluminiscencia, a partir del cual se pueden seguir dos rutas de reacción.^[5] Un 80% del luciferil adenilato sigue la ruta lumínica, denominada así porque se produce la emisión de luz. En este camino de reacción se oxida el luciferil adenilato siguiendo el

mecanismo de transferencia de un electrón^[10] para transformarse en la forma peróxido (3). La pérdida del AMP genera la dioxietanona (4), sobre la que se produce una descarboxilación obteniendo la estructura final de esta vía, la oxiluciferina en el primer estado electrónico excitado en sus formas ceto (5) la cual puede posteriormente tautomerizar a su forma enol (6). Al relajarse de forma radiativa produce la oxiluciferina en su estado fundamental, emitiendo un fotón (7). El 20% de luciferil adenilato sigue la ruta oscura, denominada así porque no se produce emisión de luz, en la que el intermedio de la reacción se oxida produciendo peróxido de hidrógeno y dehidroluciferil adenilato (8) el cual, tras una fosforilación, origina la dehidroluciferina (9) y una molécula de ATP.

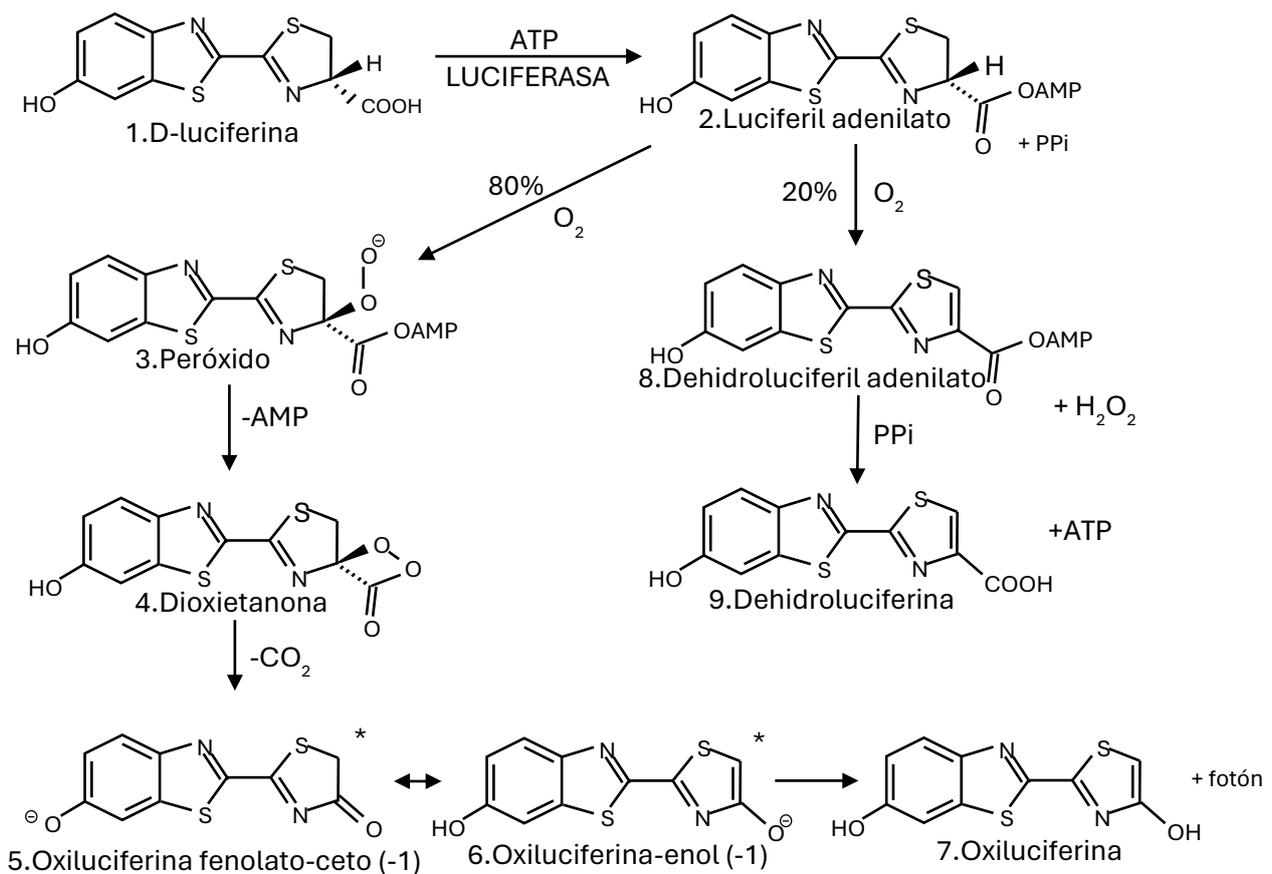


Figura 2. Mecanismo de emisión de luz de la D-luciferina.^[5]

El estudio del sistema D-luciferina/luciferasa es interesante debido a su gran variedad de aplicaciones.^[5] Algunas de ellas residen en la participación del ATP en el mecanismo de bioluminiscencia, pudiendo así determinar la concentración de ATP en una célula y estudiar el metabolismo mitocondrial. Esto permite llevar a cabo estudios de células tumorales, ya que el desarrollo de un tumor implica la privación de glucosa de sus células, lo que conlleva que se altere la ruta metabólica de la célula para la producción de ATP, produciendo una disminución en la concentración de ATP e induciendo la muerte celular.^[11] En medicina, la bioluminiscencia también se usa para la obtención de imágenes de tejidos vivos o procesos celulares de manera no invasiva, siendo

que la gran mayoría de estudios que emplean bioimagen se centran en el estudio del cáncer. Los procesos más importantes en el desarrollo de un tumor, como la metástasis o la respuesta a un fármaco, se deben a interacciones entre las células y sus alrededores. Utilizando técnicas de bioimagen, es posible obtener una representación visual de estos procesos, lo que proporciona información valiosa con la que estudiar el desarrollo de un tumor. Este proceso se ve favorecido gracias a la alta selectividad de la bioluminiscencia, la ausencia de autofluorescencia y poco ruido de fondo. Otras aplicaciones del sistema D-luciferina/luciferasa se basan en el estudio de interacciones proteína-proteína o proteína-ligando mediante el método BRET (*Bioluminescence Resonance Energy Transfer*), o el uso de la oxiluciferina en bioimagen como marcador de la expresión génica.^[12,13]

Este proyecto usa como cromóforo la D-luciferina oxidada por la luciferasa de una luciérnaga japonesa (*Nipponoluciola cruciata*) a oxiluciferina mediante el mecanismo descrito en la Figura 2. En concreto, se ha usado el tautómero OLU (estructura 5 de la Figura 2). La peculiaridad de este tautómero reside en que posee rutas de relajación efectivas desde estados excitados superiores hacia el estado excitado S_1 , a diferencia de los otros tautómeros que sufren de conversión interna a un estado unido por dipolo, haciendo que su fluorescencia sea menos efectiva.^[14] Para utilizar cromóforos en bioimagen es importante la longitud de onda de emisión, la ventana de transparencia de tejidos biológicos se encuentra en el rojo e infrarrojo cercano. Por tanto, es interesante llevar a cabo estudios donde se altere el entorno proteico del cromóforo de forma que se obtenga un desplazamiento batocrómico en la longitud de onda de emisión que concuerde con la ventana de transparencia. Dichas modificaciones solo se pueden llevar a cabo de manera racional si se conoce de manera precisa el tipo de interacciones presentes en el complejo formado por el cromóforo y la proteína. La caracterización computacional de dichas interacciones es el objeto del presente trabajo.

4. Objetivos

El objetivo general de este trabajo es el estudio y modelización computacional de las interacciones presentes en el sistema OLU/luciferasa.

Objetivos específicos:

- Simulación de la dinámica del sistema OLU/luciferasa con el programa *AMBER*.
- Cálculos de la energía libre de interacción entre la OLU y los aminoácidos de la proteína usando el método MMGBSA (Molecular Mechanics with Generalized Born Surface Area).
- Simulación de los espectros de emisión del cromóforo OLU mediante cálculos cuánticos TD-DFT (Teoría del Funcional de la Densidad Dependiente del Tiempo) con el programa *Gaussian09*, a nivel CAM-B3LYP/6-311(2d,p).

- Aprendizaje del uso del sistema operativo *Linux*, programas de química computacional como *AMBER* y *Gaussian09*, y programas de visualización como *VMD*.

5. Plan de Trabajo

Se han realizado las siguientes actividades del trabajo fin de grado:

Semanas 1-7: Búsqueda bibliográfica sobre el sistema luciferina/luciferasa y métodos de cálculo empleados (TD-DFT) y dinámica molecular (Molecular Dynamics, MD)).

Semanas 3 y 4: Aprendizaje en el uso del sistema operativo *Linux* y los programas a emplear: *Gaussian09* (cálculos cuánticos), *VMD* (visualización de la dinámica del sistema), *AMBER* (simulación de la MD).

Semanas 5 y 6: Montaje del sistema y simulación de la dinámica clásica del sistema OLU/luciferasa solvatado.

Semanas 7 y 8: Análisis de las propiedades de la dinámica simulada: temperatura, densidad y energía total entre otras.

Semana 9: Análisis de la función de distribución radial de los aminoácidos alrededor del cromóforo para identificar los aminoácidos más relevantes.

Semanas 10 y 11: Análisis de las geometrías de la MD para obtener geometrías representativas (análisis de clustering) y cálculo de la energía libre de interacción entre la OLU y la proteína.

Semana 12: Cálculo de los espectros de emisión de la OLU interaccionando con la proteína.

Semanas 13, 14 y 15: Redacción del informe final y preparación de la defensa del trabajo.

6. Métodos

En este trabajo se han utilizado diferentes métodos de cálculo. Para la simulación de la dinámica del complejo enzima/sustrato se ha recurrido a ecuaciones de MD clásica. Para la minimización de la energía y obtención de la geometría de equilibrio así como las cargas tanto del cromóforo como de la molécula de AMP, necesarias para calcular las interacciones electrostáticas dentro del campo de fuerzas, se han empleado cálculos de mecánica cuántica (*Quantum Mechanics*, QM) basados en la teoría del funcional de la densidad (*Density Functional Theory*, DFT). El análisis de la dinámica simulada se llevó a cabo mediante el método MMGBSA,^[15] con el que se calcularon las energías libres de interacción del sistema. Por último, para calcular el espectro de emisión se recurrió a cálculos de mecánica cuántica/mecánica molecular (QM/MM) usando TD-DFT.

6.1. Dinámica Molecular Clásica

Un sistema biológico, como es el caso del complejo luciferina/luciferasa, está formado por un gran número de átomos (del orden de 10^4 átomos). La simulación de la evolución temporal de un sistema tan grande requiere el uso de métodos aproximados, como es el caso de la MD clásica. En la MD clásica se usan las ecuaciones de Hamilton para describir la evolución del sistema a lo largo del tiempo, teniendo en cuenta tanto la energía potencial como la energía cinética y considerando un sistema conservativo, de forma que la energía potencial tan solo depende de la posición de los átomos.

$$\frac{dp_i}{dt} = -\frac{dV}{dq_i} \quad [1] \qquad \frac{dq_i}{dt} = \frac{dT_i}{dp_i} \quad [2]$$

En las Ecuaciones [1] y [2], p_i es el momento lineal del átomo i , t es el tiempo, V es la energía potencial, q_i son las coordenadas del átomo i , y T_i es la energía cinética del átomo i .

Las ecuaciones de Hamilton no tienen solución exacta, una forma de resolverlas es mediante el uso de soluciones numéricas obtenidas mediante el uso de expansiones de Taylor truncadas en el tercer término (segundo orden). Esta resolución se consigue con el algoritmo de Velocidad de Verlet^[16] (Ecuaciones [3] y [4], en las que m y v son la masa y la velocidad respectivamente de cada átomo).

$$q(t + \Delta t) = q(t) + \Delta t v(t) - \frac{\Delta t^2}{2m} \frac{dV(t)}{dq} \quad [3]$$

$$v(t + \Delta t) = v(t) - \frac{\Delta t}{2m} \left(\frac{dV(t + \Delta t)}{dq} + \frac{dV(t)}{dq} \right) \quad [4]$$

Para su resolución se requiere del cálculo del gradiente de la energía potencial (término $\frac{dV(t)}{dq}$), esto significa que para calcular q_t y v_t se necesita conocer las propiedades en tiempos anteriores, por lo tanto, se deben definir las coordenadas iniciales (q_0) y velocidades iniciales (v_0). El cálculo del gradiente de la energía potencial se puede llevar a cabo mediante el uso de un campo de fuerzas (*Force Field*, FF), resolviendo la ecuación de Schödinger con métodos *ab initio*,^[17] o mediante la combinación de ambos métodos en cálculos QM/MM. En este trabajo se ha escogido la primera opción debido al gran tamaño del sistema. Las coordenadas iniciales se toman de un banco de datos de proteínas (*Protein Data Bank*, PDB), mientras que para las velocidades iniciales se usa una distribución de Maxwell-Boltzmann asumiendo equilibrio térmico en la temperatura, T , a la que se realiza la simulación (Ecuación [5]). Siendo k_B la constante de Boltzmann:

$$P(v) = \left(\frac{m}{2\pi k_B T} \right)^{3/2} 4\pi v^2 e^{-\frac{mv^2}{2k_B T}} \quad [5]$$

6.2. Campo de Fuerzas

La energía potencial dicta el comportamiento de un sistema, ya que ésta describe la interacción de todas las partículas del mismo. Un campo de fuerzas es una ecuación que correlaciona la energía potencial de un sistema con sus coordenadas internas (Ecuación [6]). Los FF más usados dividen la energía potencial en dos tipos de interacciones: interacciones enlazantes, que se componen de interacciones entre átomos que forman enlaces, ángulos y diedros (interacciones 1-2, 1-3 y 1-4 respectivamente), e interacciones no enlazantes entre átomos más lejanos.

$$V_{total} = \sum_{i=1}^{N_{enlace}} V_{enlace} + \sum_{i=1}^{N_{ángulo}} V_{ángulo} + \sum_{i=1}^{N_{diedro}} V_{diedro} + \sum_{i=1}^{N_{no\ enlazante}} V_{no\ enlazante} \quad [6]$$

Las interacciones de átomos dentro de enlaces y ángulos están descritas por potenciales armónicos (Ecuación [7]), esto implica que solo son correctos para geometrías cercanas a la geometría de equilibrio, por lo que no se pueden usar para la descripción de la ruptura o la formación de un enlace. Las interacciones de diedros se dan entre átomos separados por tres enlaces, y son descritas usando un potencial de Fourier (Ecuación [8]). Las interacciones no enlazantes se componen de interacciones coulombicas (Ecuación [9]), interacciones de van der Waals e interacciones de repulsión, estas dos últimas están descritas por un potencial de Lennard-Jones (Ecuación [10]).

$$V_{armónico} = \frac{1}{2}k_b(x - x_0)^2 \quad [7]$$

$$V_{diedro} = k_t(1 + \cos(n\omega - \gamma)) \quad [8]$$

$$V_{Coulomb} = \frac{Q_i Q_j}{4\pi\epsilon r_{ij}} \quad [9]$$

$$V_{Lennard-Jones} = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad [10]$$

En la ecuación [7], k_b es la constante de fuerza, y x corresponde a r , distancia entre átomos para V_{enlace} , o a Θ , ángulo entre átomos para $V_{ángulo}$, mientras que x_0 es el valor de equilibrio de estas coordenadas. En la Ecuación [8], ω es el ángulo diedro, y k_t, n, γ son parámetros que definen la forma del potencial. En la Ecuación [9], Q es la carga de cada átomo, r_{ij} es la distancia entre átomos, y ϵ es la permeabilidad eléctrica en el vacío. En la Ecuación [10], ϵ_{ij} es la profundidad del pozo de energía y σ_{ij} corresponde a la separación entre átomos cuando el potencial de Leonard-Jones es nulo, son parámetros que definen la forma del potencial. De esta forma la Ecuación [6] se puede escribir como la Ecuación [11], que corresponde a la ecuación de un FF.

$$\begin{aligned}
V_{total} = & \sum_{i=1}^{N_{enlace}} \frac{1}{2} k_b (r - r_0)^2 + \sum_{i=1}^{N_{angulo}} \frac{1}{2} k_a (\theta - \theta_0)^2 + \sum_{i=1}^{N_{diedro}} k_t (1 + \cos(n\omega - \gamma)) \\
& + \sum_{i>j}^{N_{no\ enlazante}} \frac{Q_i Q_j}{4\pi\epsilon r_{ij}} + 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad [11]
\end{aligned}$$

6.3. Condiciones Periódicas de Contorno

La representación computacional de un sistema biológico es imposible debido a la gran cantidad de átomos que se ven implicados. Debido a esto, tan solo se representa de forma explícita un sector del espacio, denominado celda primaria, de forma que se resuelven las ecuaciones de MD en ella. Esta misma celda se repite en las tres direcciones del espacio, de forma que todo movimiento que tiene lugar en la celda primaria también ocurrirá en las réplicas de la celda. Además, para mantener el principio de conservación de la materia y energía, en el caso de que un átomo salga de la celda, otro idéntico, con la misma dirección y velocidad entrará por el lado opuesto de la celda. Esto se denomina condiciones periódicas de contorno (*Periodic Boundary Conditions*, PBC), y se usan en simulaciones de MD. De no usar la PBC en los cálculos, computacionalmente se consideraría que la celda representada está aislada, de forma que no existe nada a su alrededor. En esta situación, se daría un gran error en los cálculos, ya que hay interacciones de largo alcance que se pueden encontrar fuera de la celda que no se tendrían en cuenta. Las PBC ayudan a solventar este problema, de forma que al crear un espacio de celdas de imitación se pueden tener en cuenta dichas interacciones de largo alcance. Si bien esto es una forma de reducir el error de cálculo, no es una representación fiel a la realidad pero, como representar un sistema biológico de forma completamente realista es imposible, trabajar bajo condiciones PBC es una buena aproximación.

Una replicación de celdas infinita implica un número de interacciones a tener en cuenta demasiado grande, por lo que para hacer viable el cálculo computacional se define un radio suficientemente grande, de forma que tan solo se tienen en cuenta las interacciones que ocurren dentro de este (radio de cutoff, R_{cutoff}). Las interacciones de van der Waals son de corto alcance, por lo que decaen rápidamente, esto significa que cerca del R_{cutoff} su contribución será cercana a cero y, por tanto, se puede aplicar una corrección cerca de R_{cutoff} para hacer que en el límite la contribución sea cero. En el caso de las interacciones electroestáticas de Coulomb, estas son interacciones de largo alcance, lo que significa que decaen muy despacio con la distancia. Esto supone un problema, ya que a la distancia de R_{cutoff} su contribución es aún considerable, lo que produciría una discontinuidad en la energía potencial al igualar a cero su valor en la distancia R_{cutoff} . Para solucionar este problema se emplea el sumatorio de Ewald para transformar las interacciones de largo alcance en interacciones de

corto alcance, con las que podemos aplicar el R_{cutoff} seleccionado. Es importante indicar que para que el método del sumatorio de Ewald funcione de forma correcta, la celda ha de ser neutra.

6.4. Propiedades Termodinámicas

Para llevar a cabo una simulación de un sistema biológico con MD hay que tener en cuenta las propiedades termodinámicas. Las simulaciones se pueden llevar a cabo manteniendo con un valor constante el número de partículas (N), el volumen (V) y la temperatura (T) en el llamado conjunto canónico NVT, o manteniendo el número de partículas (N), la presión (P) y la temperatura (T) constantes en el conjunto isobárico-térmico NPT.^[18] Para el control de dichos valores se usan algoritmos matemáticos denominados termostatos y baróstatos.

La temperatura está relacionada con la energía cinética del sistema, por lo que los termostatos controlan esta propiedad mediante la modificación del momento lineal de los átomos. Encontramos varios modelos de termostatos como el de Berendsen^[19] y el de Andersen. En este proyecto se ha usado el termostato de Langevin,^[18] en el cual el momento lineal de los átomos se ve afectado por fuerzas de fricción y colisiones entre átomos. Esto se consigue modificando las ecuaciones de Hamilton para describir el movimiento, de forma que se tiene en cuenta la viscosidad del disolvente y colisiones aleatorias entre átomos.

La presión y el volumen del sistema están relacionadas mediante la ecuación de estado, de forma que modificando el volumen de la celda se controla la presión. En este proyecto se ha utilizado el baróstato de Berendsen,^[19] en el cual los vectores que forman la celda unidad se ven modificados por un factor dependiente de la presión.

6.5. Energía Libre de Interacción

Una vez que se ha llevado a cabo la dinámica del sistema luciferina/luciferasa, el siguiente paso de este proyecto es el cálculo de la energía libre de interacción entre el cromóforo y la enzima. La energía libre de interacción es una medida de la fortaleza con la que interaccionan la proteína y el cromóforo. Para su cálculo se ha usado el método MMGBSA,^[15] que tiene en cuenta la diferencia de las energías libres de interacción entre el complejo ($\Delta G_{complejo}$), la proteína ($\Delta G_{receptor}$) y el ligando ($\Delta G_{ligando}$) (Ecuación [12]). Cada uno de estos términos se calculan teniendo en cuenta tres posibles contribuciones (Ecuación [13]): la contribución de energía en fase gas (ΔE_g), que incluye las energías de interacción electroestáticas y de van der Waals; la energía de solvatación (ΔG_{solv}), que a su vez está descompuesta en la aportación de la energía de solvatación polar y la energía de solvatación no polar. La última aportación incluye el término de la entropía ($T\Delta S$), que no se tiene en cuenta porque habitualmente su aportación es baja y requiere un coste computacional importante.

$$\Delta G_{binding} = \Delta G_{complejo} - \Delta G_{ligando} - \Delta G_{receptor} \quad [12]$$

$$\Delta G_i = \Delta G_g + \Delta G_{solv} - T\Delta S \quad [13]$$

6.6. Teoría del Funcional de la Densidad y su Variante Dependiente del Tiempo

Además de la energía libre de interacción, también se ha calculado el espectro de emisión del sistema usando cálculos QM/MM, donde la región del cromóforo está descrita mediante QM usando la TD-DFT. La teoría del funcional de la densidad fue formulada en base a dos teoremas propuestos por Hohenberg y Kohn.^[20] El primer teorema afirma que cualquier propiedad del estado fundamental puede ser calculada a partir de la densidad electrónica ($\rho(r)$), para ello ésta debe ser siempre positiva y su integración en todo el espacio debe corresponder al número total de electrones. Además, la densidad electrónica también determina de manera unívoca el potencial externo al que están sometidos los electrones, que se corresponde con la energía de interacción electrón-núcleo dentro de la aproximación de Born-Oppenheimer. En esta aproximación se mantienen los núcleos de los átomos en una posición fija al considerar que estos no se ven afectados por el movimiento de los electrones, mientras que los electrones se adaptan al movimiento de los núcleos. De forma esta forma, la energía electrónica se puede escribir de la siguiente manera:

$$E_e(\rho) = T_e(\rho) + E_{ee}(\rho) + E_{ext}(\rho) \quad [14]$$

Donde E_e es la energía electrónica; T_e la energía cinética electrónica; E_{ee} la energía de interacción entre electrones y E_{ext} el potencial externo. El símbolo ρ entre paréntesis denota la dependencia del término de la densidad electrónica. Esta ecuación se considera dentro de la aproximación de Born-Oppenheimer, lo que implica que la energía cinética de los núcleos (T_N) es nula y el término que describe la interacción entre núcleos (V_{NN}) sea constante, por lo que no aparecen en ella.

El segundo teorema de Hohenberg y Kohn dice que la densidad electrónica exacta del estado fundamental es aquella que minimiza la energía, de forma que $\frac{dE_e(\rho)}{d\rho} = 0$. Esto implica el uso del teorema variacional, donde cualquier energía calculada siempre será igual o mayor a la energía exacta. El problema es que no se conoce la relación entre $T_e(\rho)$ y la densidad electrónica, y por lo tanto no se puede obtener la energía de forma exacta. Para solventar este problema se usa el método de Kohn-Sham. En este método se supone un sistema de referencia en donde los electrones no interactúan, por lo que su energía ($E_{e,r}$) si se puede calcular de forma exacta. La diferencia energética entre el sistema real y el de referencia se denomina energía de correlación e intercambio (E_{xc}); y corresponden con aquellas aportaciones energéticas que se ignoran en el sistema de referencia para poder calcularse de forma exacta.

$$E_{xc}(\rho) = E_e(\rho) - E_{e,r}(\rho) \quad [15]$$

Se han desarrollado funcionales con el objetivo de describir E_{xc} , estos funcionales son parametrizados de forma que puedan reproducir propiedades calculadas con métodos más exactos o experimentales. En la formulación de la DFT, este se ve descrito como un teorema variacional, pero en la práctica, la necesidad de usar la energía de correlación e intercambio aproximada hace que DFT no sea variacional.

La DFT tan solo puede ser usada para calcular el estado fundamental de un sistema, esto ha llevado a la formulación de una DFT dependiente del tiempo (TD-DFT).^[17] El teorema de Runge-Gross^[21] propone la existencia de una relación unívoca entre la densidad electrónica dependiente del tiempo ($\rho(r, t)$) y el potencial externo dependiente del tiempo. Esta dependencia permite la reformulación de las ecuaciones de la DFT de forma que no solo permite el cálculo de la energía en el estado fundamental, si no que además también permite el cálculo de estados excitados. Para el cálculo de los espectros de emisión del sistema se ha usado TD-DFT en cálculos QM/MM.

6.7. Mecánica Cuántica/Mecánica Molecular

Para el cálculo del espectro de emisión se usaron cálculos QM/MM, tratando todo el complejo menos el cromóforo con MM (región externa) y este último con QM (región interna). En la Figura 3 podemos ver la proteína en azul, parte calculada con MM, y en rojo el cromóforo, parte calculada con QM.

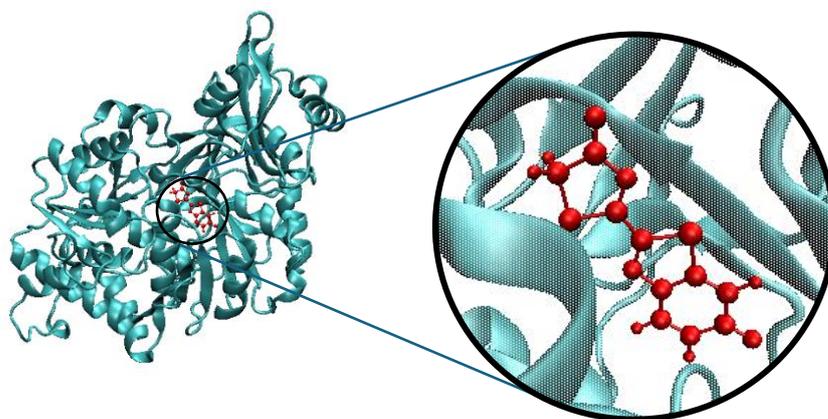


Figura 3. Representación del complejo enzima/sustrato diferenciando la región interna y la región externa en los cálculos QM/MM. En azul la proteína y en rojo el cromóforo.

En los cálculos QM/MM se asume que todos los cambios electrónicos se centran en una parte del sistema denominada región interna, de forma que esta región se trata a nivel QM mientras que el resto del sistema con MM. Siendo la energía calculada de forma aditiva mediante la siguiente ecuación:

$$E = E^{QM}(I) + E^{MM}(O) + E^{QM/MM}(I + O) \quad [16]$$

La interacción electrostática entre ambas regiones es calculada mediante el uso de un hamiltoniano efectivo (\hat{H}^{eff}) (Ecuación [17]). \hat{H}^{eff} incluye la interacción entre las cargas de la región externa (q_i^O) y el potencial electrostático de la región interna (\hat{V}^I) (Ecuación [18]). De esta forma la función de onda de la región interna se ve polarizada por las cargas clásicas de la región externa, pero las cargas de la región interna no afectan a las cargas de la región externa. Esto se conoce como *electrostatic embedding*, donde las cargas de la región externa se cogen de un FF como puntos fijos.

$$\hat{H}^{eff} |\psi_I\rangle = (\hat{H}_I^O + \hat{H}_{I/O}) |\psi_I\rangle = E |\psi_I\rangle \quad [17]$$

$$\hat{H}_{I/O} = \hat{H}_{I/O}^{eff} = \sum_i q_i^O (r_i) \hat{V}^I(r_i) \quad [18]$$

En este proyecto se ha utilizado el *electrostatic embedding*, pero también es importante remarcar que hay otras formas de realizar el cálculo con QM/MM.^[22] En el *mechanical embedding* el término que describe la interacción entre ambas regiones es completamente calculado con un FF, y por lo tanto la función de onda de la región interna no se ve afectada por la región externa. Otra forma sería usar *polarizable embedding*, donde el hamiltoniano de interacción entre ambas regiones es ampliado con un término de polarización, lo que requiere un FF polarizable. Esto produce que ambas regiones se vean polarizadas la una por la otra.

7. Resultados y Discusión

En este proyecto se ha realizado una simulación de MD del sistema luciferasa/oxiluciferina, para estudiar la interacción entre la proteína y el cromóforo. Para ello, se ha calculado la energía libre de interacción mediante el método MMGBSA.^[15] El cálculo ha sido realizado de dos formas: (i) tomando geometrías equiespaciadas a lo largo de la trayectoria, y (ii) clasificando la dinámica en clústeres de geometrías similares de la OLU y tomando las geometrías representativas de cada clúster. Más adelante se discutirá la exactitud de ambas aproximaciones. También se ha calculado el espectro de emisión del cromóforo mediante cálculos QM/MM utilizando la TD-DFT.

7.1. Montaje y Descripción del Sistema

Para realizar los cálculos lo primero es el montaje del sistema luciferasa/oxiluciferina. Para ello, la estructura de la proteína luciferasa ha sido obtenida de la base de datos de proteínas (PDB ID: 2D1R).^[23] Este sistema corresponde con la estructura cristalina termoestable de la luciferasa de la luciérnaga japonesa. Además, la proteína está interaccionando con una molécula de oxiluciferina y una molécula de AMP. Estos dos últimos residuos mencionados dan información acerca de la posición del sitio activo de la proteína. Los archivos descargados del PDB^[23] no contienen la posición de los hidrógenos, por lo que es necesario protonar tanto los ligandos como la proteína. La estructura de la

oxiluciferina fue sustituida por la correspondiente a OLU, mientras que el AMP fue protonado acorde al pH fisiológico, con carga -1. Además, fue necesario realizar un cambio en la proteína substituyendo los residuos CSO (S-hidroxicisteína) por CYS (cisteína) y también se añadieron los hidrógenos correspondientes. Tras este proceso se solvató el sistema, añadiendo moléculas de agua alrededor del complejo sustrato/enzima creando una celda con forma de octaedro truncado, imponiendo que la máxima distancia entre cualquier átomo del complejo y una molécula de agua fuera de 12 Å. La celda tiene aspecto de octaedro truncado ya que es la forma geométrica más parecida a una esfera (que es más eficiente en términos de número de moléculas) que no deja espacios vacíos entre celdas contiguas cuando se aplique posteriormente PBC. Es importante la neutralidad del sistema para el cálculo de MD, por lo que añadimos cuantos cationes de Na⁺ sean necesarios para la neutralidad de la celda. Para la solvatación y neutralización del sistema se usó el programa TLEAP.^[24]

Para describir el sistema se emplea un campo de fuerzas, cuya parte más importante es la energía de Coulomb, que depende de las cargas. Por ello, calculamos las cargas tanto de la geometría optimizada del AMP como de la OLU de manera cuántica mediante el método de Merz-Kollman^[25] en un cálculo DFT con el funcional CAM-B3LYP^[26] y la base 6-311(2d,p).^[27,28,29] Además, el resto de parámetros del campo de fuerzas necesarios para el AMP se han obtenido del General Amber Force Field (GAFF).^[30] En el caso del cromóforo, se han utilizado parámetros que describen el primer estado excitado electrónico. Estos parámetros han sido desarrollados por el grupo de investigación MoBioChem mediante el ajuste de energías TD-CAM-B3LYP/6-311G(2d,p). Las moléculas de agua y el Na⁺ se describieron según el modelo TIP3P^[31] y la luciferasa con el campo de fuerzas FF19SB.^[32]

7.2. Dinámica Molecular

Los cálculos de MD comienzan con una minimización para buscar la geometría que corresponda a la energía mínima del sistema bajo condiciones PBC. Este proceso se llevó a cabo en 10000 pasos, de forma que se inicia con el método *steepest descent* y tras 5.000 pasos se cambia al método *conjugate gradient*. En la Figura 4a, se encuentra la representación de la energía frente al número de pasos. En ella se puede apreciar que al final de la minimización la energía se encuentra estabilizada en los -258735 kcal/mol.

Posteriormente se realizó una termalización de 500.000 pasos con un tiempo de paso de 2 femtosegundos a volumen constante (condiciones dentro del conjunto canónico NVT) hasta llegar a los 300 K utilizando el termostato de Langevin^[18] para controlar la temperatura (Figura 4b). En la Figura 4c se recoge la energía total de las fases de termalización, así como su descomposición en energía cinética y energía potencial. Durante la termalización encontramos dos etapas distinguibles en las representaciones Figura 4b y 4c: en la primera etapa se produce el calentamiento, donde un

aporte de energía hace que la temperatura del sistema aumente de forma constante hasta alcanzar la temperatura deseada de 300K. La segunda etapa comienza una vez alcanzados los 300 K de temperatura, en ella cesa el aporte energético y por tanto la energía del sistema y la temperatura se estabilizan.

Por último, tiene lugar la fase de producción dentro del conjunto isobárico-térmico NPT, en la que se mantiene la temperatura constante a 300 K usando el termostato de Langevin^[18] y aplicando el baróstato de Berendsen,^[19] se controla la presión para que sea constante e igual a 1 bar. Este proceso se llevó a cabo en 250.000.000 pasos, con tamaño de paso de 2 femtosegundos, lo que corresponde a 500 nanosegundos de simulación. Al aplicar estas restricciones es el volumen el que varía para controlar la densidad ya que, al estar en un entorno biológico, la densidad del sistema ha de ser cercana a 1 g/cm^3 (Figura 4d), lo que corresponde a la densidad del agua.

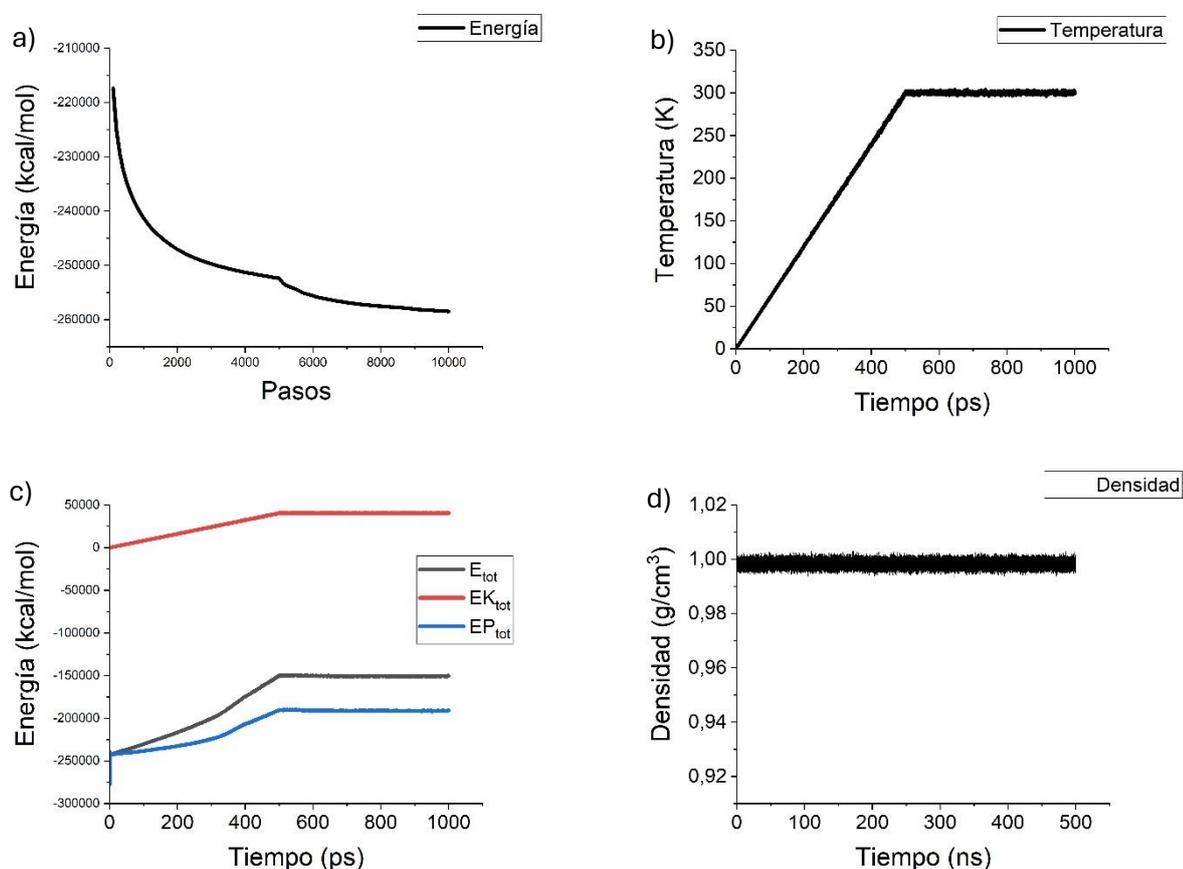


Figura 4. Variación de las propiedades del sistema durante las diferentes fases de la simulación. a) Disminución de la energía durante la minimización. b) Aumento de la temperatura durante la termalización. c) Variación de las energías total (E_{tot}), cinética (EK_{tot}) y potencial (EP_{tot}) durante la termalización. d) Variación de la densidad durante la producción.

Además, se usó la raíz de desplazamiento cuadrático medio (*Root Mean Square Deviation*, RMSD) para evaluar la estabilidad de la proteína a lo largo de la dinámica. El RMSD es un valor que nos

indica el desplazamiento que han tenido los átomos con respecto de la posición inicial, con la Ecuación [19]:

$$RMSD = \sqrt{\frac{1}{n} \sum_{i=1}^n ||P_i - Q_i||^2} \quad [19]$$

Donde n es el número total de átomos, P_i es la posición inicial del átomo i, Q_i es la posición del átomo i a tiempo t. En la Figura 5 se encuentra RMSD para la simulación completa, obtenida con el programa CPPTRAJ.^[33] Al comienzo de la simulación, se puede observar un alto crecimiento en el valor del RMSD, lo que indica que el sistema no es estable ya que la proteína presenta grandes oscilaciones. Por el contrario, al final de la dinámica el valor se estabiliza. En esta representación encontramos que el valor del RMSD crece hasta un punto cercano a los 75 ns, lo que corresponde a la geometría número 1500 (marcado con una línea azul en la Figura 5). Para tener certeza de que es sistema está completamente estabilizado se ha decidido tomar como punto de partida para el análisis la dinámica a partir de los 100 ns, lo que corresponde a la geometría número 2000 de la dinámica (marcado con una línea roja en la Figura 5), es decir, el análisis de la dinámica, el cálculo de ΔG y el cálculo del espectro se hacen entre las geometrías 2001 y 10000, o lo que es lo mismo, entre los 100 y 500 ns.

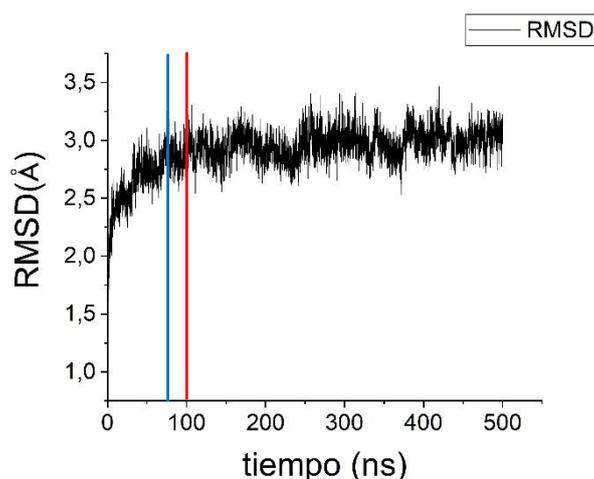


Figura 5. Representación del RMSD a lo largo de la fase de producción.

Para poder realizar la descomposición de la energía por aminoácidos es necesario conocer cuáles son los que se encuentran cerca del cromóforo. Este cálculo es muy costoso realizarlo teniendo en cuenta toda la proteína, por lo que se necesita una distancia límite (cutoff) entorno al cromóforo para reducir el tiempo de cálculo. De esta forma tan solo se tienen en cuenta los aminoácidos que espacialmente están cerca del cromóforo. El cutoff lo establecemos mirando la función de distribución radial (*Radial Distribution Function*, RDF), que nos da información de las esferas de coordinación que se forman alrededor de un centro. En este caso con la RDF vemos como varía la posición de los carbonos alfa de los aminoácidos entorno a la OLU (Figura 6).

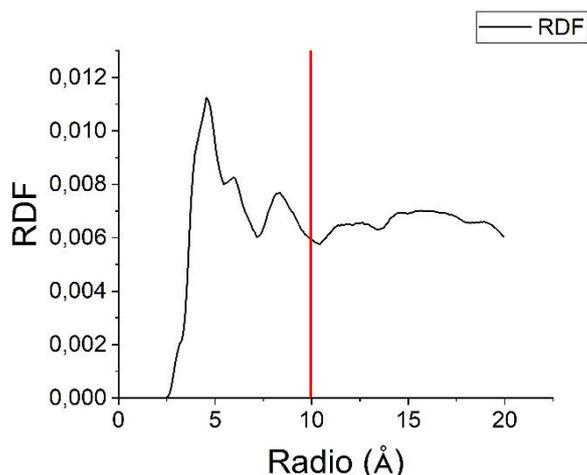


Figura 6. Representación de variación de la posición de los carbonos alfa de los aminoácidos en función del radio entorno a la OLU con la RDF.

En la Figura 6 se ven tres esferas de coordinación bien definidas, cuyos picos se encuentran en los 4,55 Å, 5,95 Å y 8,35 Å respectivamente. Debido a que las tres primeras esferas de coordinación se encuentran prácticamente al completo dentro de un radio de 10 Å, es este valor el que se ha seleccionado como *cutoff* (marcado en rojo en la Figura 6).

7.3. Energía Libre de Interacción

En el apartado anterior se ha usado el RMSD para determinar la ventana temporal de la MD donde la estructura de la proteína estaba estabilizada, concluyendo que esto se da en los últimos 400 ns. Sobre dicha ventana se ha estudiado la energía libre de interacción mediante cálculos MMGBSA.^[15] Para ello es importante determinar el número mínimo de geometrías necesarias para que el cálculo sea representativo para la dinámica al completo, es decir, para que el cálculo esté convergido. Para ello, se han seguido dos metodologías: análisis de la energía libre de interacción tomando geometrías equidistantes en toda la dinámica estabilizada; y agrupación de la dinámica estabilizada en diferentes números de clústeres para llevar a cabo el análisis de la energía libre de interacción con una geometría representativa de cada clúster.

Con el cálculo MMGBSA^[15] se obtiene la energía libre de interacción total como la diferencia de energía entre el complejo (OLU/luciferasa) y ambos monómeros cuando están libres. A su vez, esta energía total se puede dividir en contribuciones de cuatro tipos de interacciones. Interacciones electrostáticas, que son calculadas siguiendo un potencial de Coulomb, e interacciones de Van der Waals, que se calculan con un potencial de Leonard-Jones. La solvatación polar, que se calcula como la diferencia de energía entre la interacción del complejo con el disolvente y cada uno de los monómeros con el disolvente, y la solvatación no polar que tiene en cuenta la diferencia energética en las interacciones hidrofóbicas entre el disolvente y el complejo, y el disolvente y cada uno de los monómeros.

Además, también se ha llevado a cabo un estudio de la descomposición de la energía libre de interacción por aminoácidos. Para determinar los residuos sobre los que llevar a cabo el análisis de la descomposición, se analizan los últimos 400 ns de la trayectoria, para determinar los aminoácidos que pasan un mayor tiempo cerca del cromóforo durante la simulación. Se ha usado el programa CPPTRAJ^[24] para determinar la fracción del tiempo que pasa cada residuo en contacto con el cromóforo, y se han escogido para el análisis todos aquellos aminoácidos dentro del *cutoff* de 10 Å que pasan más de un 5% de la trayectoria en contacto con la OLU. Siendo 23 aminoácidos en total: ARG220, ARG339, HIE247, PHE252, MET251, GLY256, MET198, THR348, ILE351, TYR342, VAL368, GLN340, THR529, TYR257, LEU528, GLY230, HIE223, HIE246, ASN231, ASN199, ALA319, TYR229.

En la Figura 7, se puede ver una representación de la OLU junto a los aminoácidos en consideración. En verde se ve la proteína, en el centro la OLU, en rojo los aminoácidos más importantes que se comentarán en los apartados siguientes, y en azul el resto de los aminoácidos analizados.

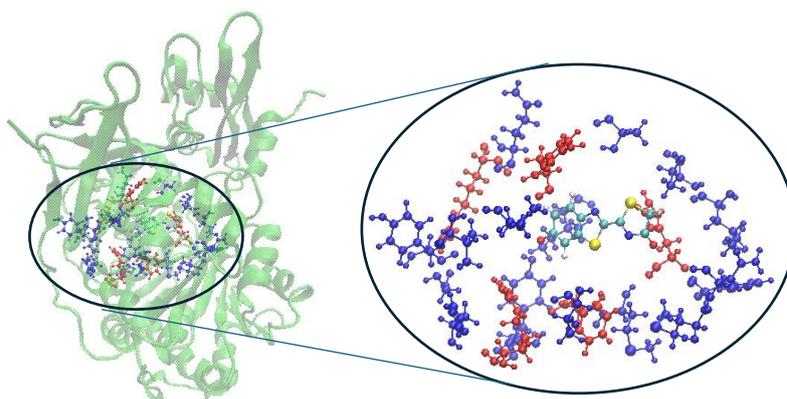


Figura 7. Representación de la OLU y los aminoácidos con más de un 5% de interacción durante la simulación, siendo representados en azul y rojo los aminoácidos más relevantes.

Los aminoácidos estudiados tienen una gran variedad de cadenas laterales. Encontramos aminoácidos con cadenas laterales ácidas como asparagina (ASN) y glutamina (GLN), o básicas como arginina (ARG) e histidina (HIE). También están presentes cadenas laterales aromáticas como el caso de fenilalanina (PHE) y tirosina (TYR), cadenas alifáticas en leucina (LEU), glicina (GLY), isoleucina (ILE), valina (VAL) y alanina (ALA), y cadenas laterales polares como el caso de treonina (THR).

7.3.1. Cálculo de la Energía Libre de Interacción con Geometrías Equidistantes

En esta primera metodología, se han utilizado hasta un máximo de 500 geometrías equidistantes de los últimos 400 ns de la trayectoria. En la Figura 8 se puede ver la variación de la energía libre de interacción total (ΔG_T) frente al número de geometrías utilizado para calcularla (# Geometrías). Se

aprecia que cuando se utilizan pocas geometrías para calcular la ΔG_T la energía presenta grandes oscilaciones. Mientras que, a partir de las 200 geometrías, la ΔG_T se estabiliza entorno a las -33,89 kcal/mol. Este valor ha sido calculado como la media aritmética de los últimos cinco valores, donde se aprecia en la Figura 8 que la energía libre de interacción empieza a tomar un valor similar para cada uno de ellos, considerando así este valor como referencia de la convergencia del cálculo de la energía libre de interacción.

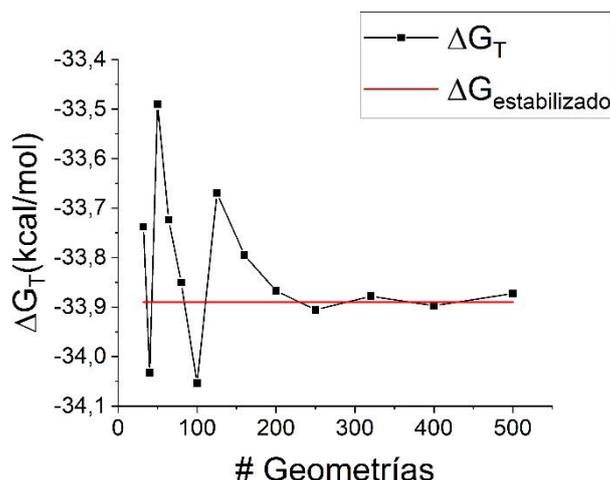


Figura 8. Energía libre de interacción en función del número de geometrías seleccionadas para su cálculo.

Como ΔG_T se estabiliza a partir de las usan 200 geometrías, la Figura 9 muestra la descomposición de la ΔG_T para ese número de geometrías para los aminoácidos más cercanos al cromóforo. Es decir, se muestra la contribución de cada aminoácido a la energía de interacción total. Es importante comentar que en la Figura 9 aparecen los aminoácidos ARG220 y ARG339, que son dos argininas, cuyos valores de interacción electrostática y solvatación polar se salen del eje representado, por lo que se ha incluido una tabla con sus descomposiciones energéticas (Tabla 1).

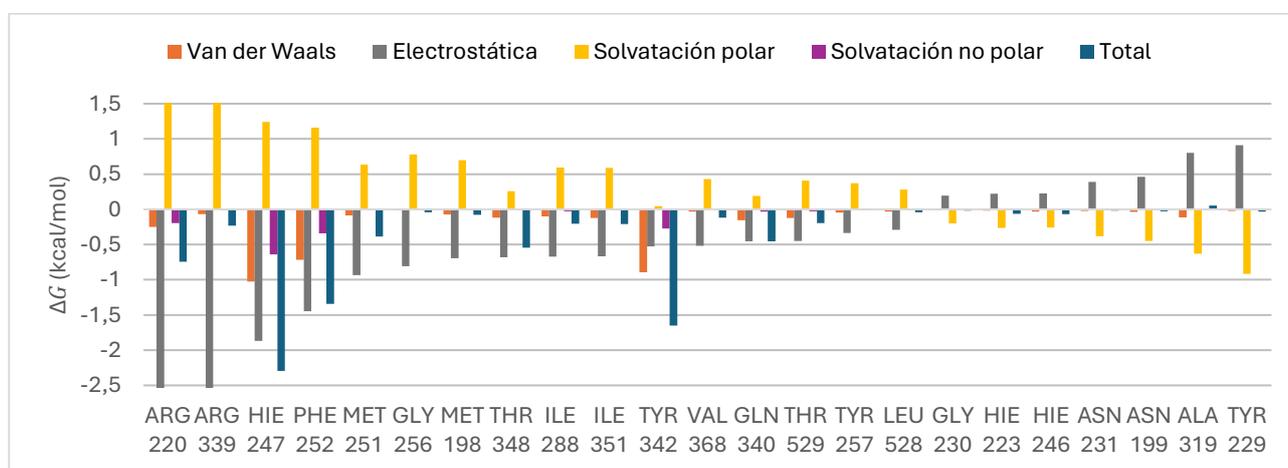


Figura 9. Descomposición de la energía libre de interacción total por aminoácido tomando 200 geometrías equidistantes de los últimos 400 ns de la trayectoria.

Tabla 1. Descomposición de la contribución energética de los residuos más importantes. Todos los valores de la tabla se dan en unidades de kcal/mol.

	Van der Waals	Electrostática	Solvatación polar	Solvatación no polar	Total
HIE 247	-1,02	-1,87	1,24	-0,64	-2,29
TYR 342	-0,89	-0,53	0,04	-0,27	-1,65
PHE 252	-0,72	-1,45	1,16	-0,34	-1,34
ARG 220	-0,25	-18,6	18,3	-0,20	-0,74
ARG 339	-0,068	-13,1	12,9	-0,0086	-0,23

Al realizar los cálculos MMGBSA con la descomposición por aminoácido, se puede observar una gran disparidad entre las contribuciones de los distintos aminoácidos. En la Figura 9, se ve que los aminoácidos con mayor aportación energética total son HIE247 que corresponde a una histidina, TYR342 que corresponde a una tirosina y PHE252 que corresponde a una fenilalanina. La mayor contribución de HIE247 y PHE252 proviene de interacciones electroestáticas, aunque se ven debilitadas por la energía de solvatación polar. Para el residuo TYR342 su principal aportación se encuentra en la energía de interacción de van der Waals. La histidina es un aminoácido cuya cadena lateral contiene un grupo imidazol. Es posible la formación de enlaces de hidrógeno entre el hidrógeno enlazado al nitrógeno del grupo imidazol y un oxígeno de la OLU, lo que contribuye la aportación de la interacción electrostática, que es el término que describe los puentes de hidrógeno en el FF utilizado.

En cuanto a los aminoácidos de arginina (ARG220 y ARG339), encontramos que la ARG220 es el cuarto aminoácido con mayor contribución energética total mientras que la ARG339 tiene una aportación menor. La peculiaridad de estos dos residuos de arginina es la magnitud de la energía de interacción electrostática, siendo de -18,6 y -13,1 kcal/mol respectivamente. La cadena lateral de la arginina termina en un grupo guanidina que contiene una carga neta positiva a pH fisiológico, por lo que pueden producirse fuertes interacciones electrostáticas con la OLU, lo que explica sus valores de interacción electrostática y solvatación polar. En la Figura 10 se representan dos geometrías donde los aminoácidos ARG334 (a) y ARG220 (b) están cercanos a la OLU, de forma que sus cadenas laterales (cargadas positivamente) se orientan hacia la OLU pudiendo resultar en interacciones electrostáticas. También se encuentra una geometría donde se representa una interacción entre HIE247 y la oxiluciferina formando un enlace de hidrógeno entre el hidrógeno enlazado al nitrógeno de la histidina y el oxígeno del cromóforo.

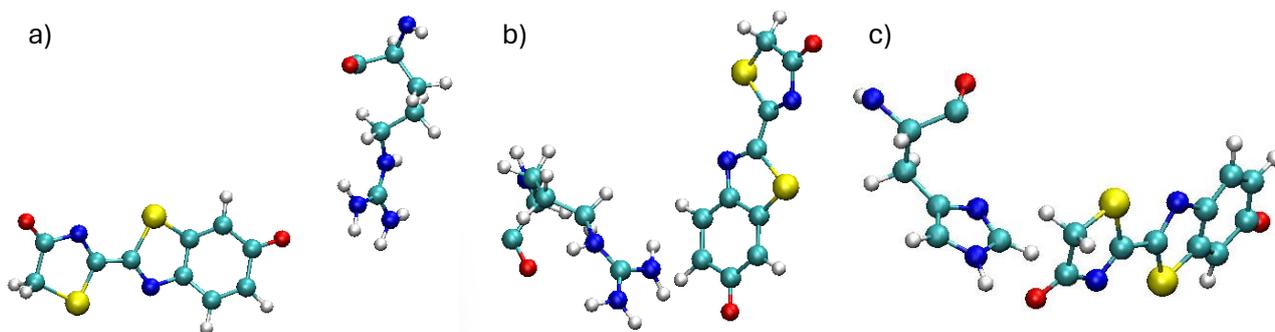


Figura 10. Representación de la OLU junto a ARG334 (a), ARG220 (b) y HIE247 (c). Los átomos de carbono se representan en color azul claro, nitrógeno en color azul oscuro, oxígeno en color rojo, azufre en color amarillo e hidrógenos en color blanco.

7.3.2. Cálculo de la Energía Libre de Interacción con Geometrías Representativas de los Clústeres

Para calcular propiedades sobre la trayectoria entera se requiere de un tiempo de cálculo relativamente alto. En el apartado anterior se tomaban geometrías equidistantes, de forma que se buscaba el menor número de geometrías equidistantes necesarias para que el cálculo de una propiedad sobre ellas equivaliese al de toda la trayectoria. De la misma manera, en este apartado se discutirá la agrupación de las 8000 geometrías que componen los últimos 400 ns de la trayectoria en clústeres, siguiendo un criterio de similitud estructural de la OLU. De ellos se calcula una geometría representativa, con el objetivo de hallar el menor número de geometrías representativas necesario sobre el que calcular la ΔG_T para que corresponda a la propiedad de la trayectoria completa.

Para este análisis, se separan las 8000 geometrías de la trayectoria en distintos números de clústeres, entre 1 a 20. De estos clústeres se escoge la geometría representativa, que corresponde al centroide del clúster, sobre la que se calcula la energía de interacción. En este caso hay que tener en cuenta que no todos los clústeres van a ser del mismo tamaño, es decir, van a incluir un número diferente de geometrías, por lo que su aportación energética será distinta. Debido a esto, la forma de calcular la energía de interacción total para cada conjunto de clústeres es mediante un promedio ponderado en donde se tiene en cuenta la población del clúster:

$$\Delta G_T^n = \sum_{i=1}^n \Delta G_{T,i} * F_i \quad [20]$$

En la Ecuación [20] n es el número de clústeres considerado, $\Delta G_{T,i}$ es la energía de interacción de la geometría representativa del clúster i y F_i es el peso del clúster i , que se calcula como el número de

geometrías dentro de ese clúster dividido entre el número de geometrías total. De esta forma se ha calculado el valor de ΔG_T^n para n entre 1 y 20 clústeres que aparece representado en la Figura 11.

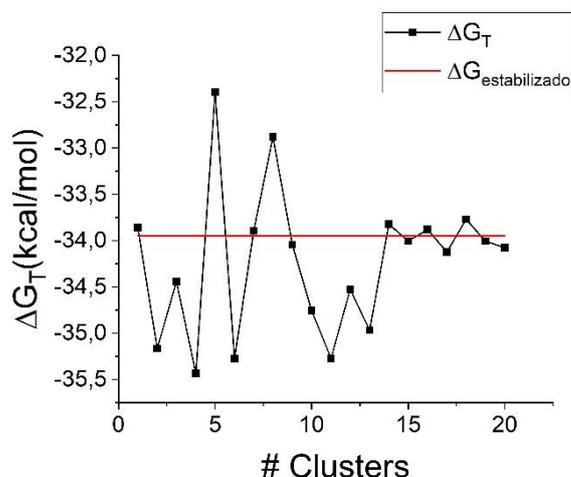


Figura 11. Energía libre de interacción en función del número de clústeres. La línea roja representa el valor de $\Delta G_{estabilizado} = -33,89$ kcal/mol, tomado del apartado 7.3.1.

En la Figura 11 se observa una oscilación de los valores de ΔG_T para números de clusters bajos, indicando que aún no se ha alcanzado la convergencia. Esto es debido a que las geometrías representativas tomadas de cada clúster no son representativas de la trayectoria completa. Sin embargo, a partir de los 14 clústeres la energía libre de interacción se empieza a estabilizar entorno a las -33,89 kcal/mol, igual que en el cálculo de la energía libre de interacción con geometrías equidistantes. El valor de $\Delta G_{estabilizado}$ se ha tomado del apartado 7.3.1., calculando la energía de interacción para 200 geometrías a lo largo de toda la trayectoria y se encuentra marcado en la Figura 11 con una línea roja.

Para el análisis de la descomposición de la energía libre de interacción por aminoácidos se ha escogido el conjunto de 15 clústeres, ya que se observa en la Figura 11 la energía libre de interacción está convergida. Como en el apartado anterior, se ha usado el método MMGBSA^[15] y se ha calculado la descomposición para los aminoácidos que pasan más de un 5% de la dinámica cerca del cromóforo.

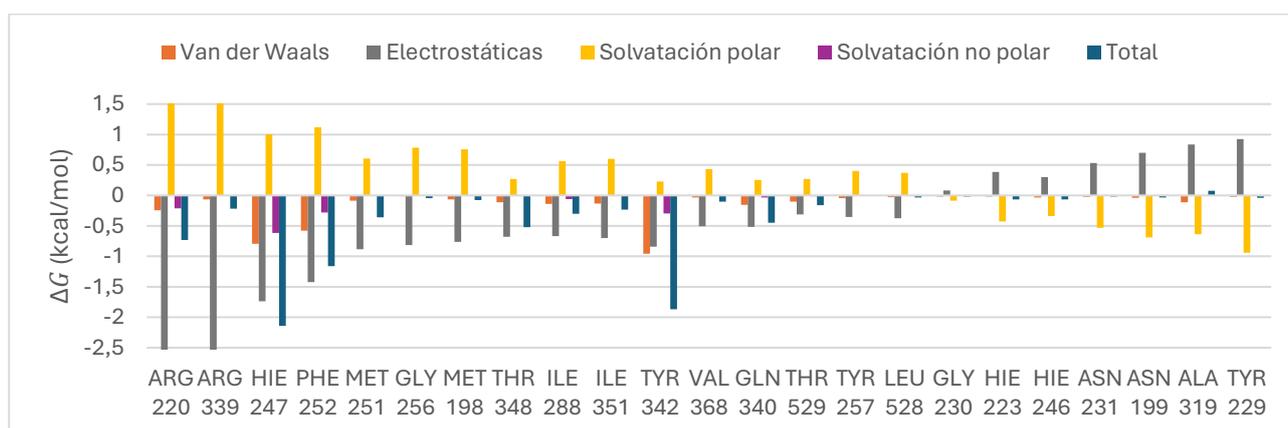


Figura 12. Descomposición de la energía libre de interacción total por residuo tomando 15 geometrías representativas de los últimos 400 ns de la trayectoria.

Tabla 2. Descomposición de la contribución energética de los residuos más importantes en el análisis de los 15 clústeres. Todos los valores se dan en unidades de kcal/mol.

	Van der Waals	Electrostática	Solvatación polar	Solvatación no polar	Total
HIE 247	-0,80	-1,74	1,00	-0,62	-2,14
TYR 342	-0,96	-0,84	0,23	-0,30	-1,87
PHE 252	-0,58	-1,42	1,12	-0,28	-1,16
ARG 220	-0,25	-18,6	18,3	-0,21	-0,73
ARG 339	-0,063	-13,3	13,1	-0,0041	-0,22

En la Figura 12, los residuos ARG220 y ARG339 tienen los valores de interacción electrostática y solvatación polar fuera del eje, por lo que su descomposición energética se recoge en la Tabla 2. Al igual que en el apartado 7.3.2., los aminoácidos HIE247, TYR342, PHE252 y ARG220 son los que mayor energía libre de interacción total presentan, por lo que para poder realizar una buena comparación se ha recogido la descomposición de la energía de dichos residuos en la Tabla 2.

El aminoácido HIE247 es el que mayor aportación total tiene y, al igual que en el caso de la PHE252, la mayor parte de dicha contribución energética viene dada por las interacciones electrostáticas. Es importante mencionar que las interacciones de van der Waals que presentan dichos aminoácidos también son relevantes. Ambos se ven desestabilizados por la energía de solvatación polar. Por otro lado, la TYR342 es el segundo residuo con mayor energía libre de interacción total, su mayor contribución se da mediante interacciones de van der Waals, aunque las interacciones electrostáticas tienen un valor parecido.

En cuanto a los residuos de arginina ARG220 y ARG339, si bien las energías libres de interacción totales que presentan concuerdan con el resto de aminoácidos, las interacciones electrostáticas presentan de nuevo un valor muy negativo, con una energía de solvatación muy alta y positiva que la debilita. Ambos métodos de cálculo muestran interacciones electrostáticas fuertes entre estas dos argininas y la OLU, con valores similares al comparar los valores de dichas interacciones en las Tablas 1 y 2.

Los resultados obtenidos por el análisis de geometrías representativas de los 15 clústeres muestran una gran concordancia con los obtenidos en el análisis de geometrías equiespaciadas. Esto indica que las 15 geometrías representativas también son representativas de los 400 ns de la trayectoria completa. Computacionalmente hablando, el cálculo de 15 geometrías representativas es considerablemente más barato que el cálculo de 200 geometrías equidistantes. Como los resultados obtenidos son

estadísticamente idénticos, en futuras investigaciones es preferible el uso del cálculo de la energía libre de interacción mediante geometrías representativas, consiguiendo abaratar costes.

7.4. Espectro de Emisión

El fenómeno de bioluminiscencia dado por la OLU se da cuando pasa del estado excitado S_1 al estado fundamental S_0 (transición $S_1 \rightarrow S_0$), estados que corresponden con una transición electrónica desde el orbital LUMO (*Lowest Unoccupied Molecular Orbital*) al HOMO (*Highest Occupied Molecular Orbital*). Para estudiar la bioluminiscencia del complejo OLU/luciferasa se ha calculado su espectro de emisión en el rango UV-Vis para 100 geometrías (Figura 13) mediante cálculos QM/MM basados en la TD-DFT. El resultado de este cálculo es un valor de energía, que para obtener una banda se posiciona una función gaussiana sobre él.

En el espectro de emisión calculado del cromóforo se aprecia un pico de emisión a 2,37 eV, que es un valor que se asemeja al valor experimental de 2,21 eV, ya que la diferencia está dentro del error de método aceptado. El valor de la energía del fotón emitido corresponde a una longitud de onda de 523 nm, que entra dentro del rango de color verde. Esta longitud de onda de emisión queda fuera de la ventana de transparencia biomédica (zona del rojo e infrarrojo cercano), pero en futuros estudios se pueden considerar mutaciones en el cromóforo o la proteína para producir sistemas con una longitud de onda de emisión adecuada. Teniendo en cuenta los resultados de los apartados anteriores, una mutación sobre los aminoácidos HIE247, TYR342, PHE252, ARG220 o ARG334 puede generar un gran desplazamiento de la longitud de onda de emisión.

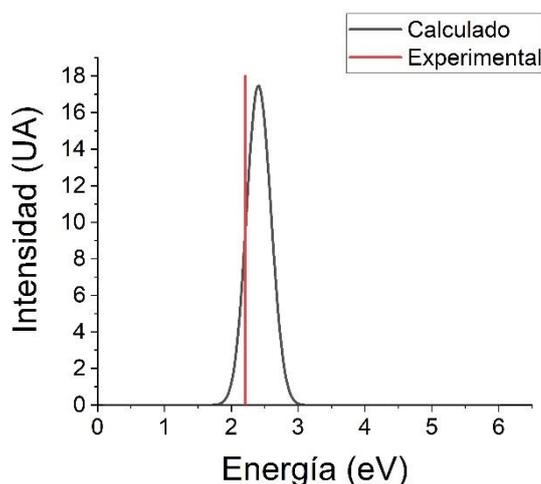


Figura 13. Espectro de emisión teórico del sistema OLU/luciferasa, calculado para 100 geometrías. En rojo se ha representado el valor de emisión experimental a 2,21 eV.^[23]

8. Conclusiones

En este proyecto se ha llevado a cabo un estudio computacional del sistema OLU/luciferasa mediante cálculos de MD clásica, energía libre de interacción y TD-DFT/MM. Sobre la trayectoria calculada mediante MD se ha estudiado la energía libre de interacción entre la OLU y la proteína con el método MMGBSA, teniendo en cuenta las contribuciones de interacciones electrostáticas, interacciones de van der Waals, energía de solvatación polar y energía de solvatación no polar. Este proceso se ha llevado a cabo siguiendo dos protocolos: análisis de geometrías equidistantes y análisis de geometrías representativas de clústeres. Además, se ha estudiado la descomposición de la energía libre de interacción por aminoácidos, que ha servido para analizar los aminoácidos más cercanos al cromóforo y las interacciones aminoácido-cromóforo más fuertes. Por último, se ha calculado el espectro de emisión del sistema OLU/luciferasa usando cálculos QM/MM en donde la región QM se ha descrito con TD-DFT. Tras haber analizado los resultados obtenidos se han llegado a las siguientes conclusiones:

1. Con la simulación de MD clásica para el sistema OLU/luciferasa y tras haber calculado el RMSD para estudiar la diferencia entre la posición inicial y cada geometría de la trayectoria, se concluye que la estructura converge pasados 100 ns de simulación, con un tramo de 400 ns estabilizados de trayectoria.
2. El análisis de la energía libre de interacción en el tramo de 400 ns estabilizados de trayectoria mediante el análisis de estructuras equiespaciadas muestra que se necesita un mínimo de 200 geometrías equiespaciadas para que converja la energía libre de interacción total.
3. El análisis de la energía libre de interacción en el tramo de 400 ns estabilizados de trayectoria mediante el análisis de geometrías representativas de clústeres formados siguiendo un criterio de similitud estructural en la OLU, muestra que se requiere de un mínimo de 15 geometrías representativas para calcular una energía libre de interacción total representativa de toda la trayectoria.
4. El análisis de la descomposición de la energía libre de interacción por aminoácidos muestra, por los dos métodos previamente mencionados, que los cuatro aminoácidos con mayor contribución son: ARG220, HIE247, PHE525 y TYR 342; siendo HIE247 el aminoácido con mayor contribución total debido a interacciones electrostáticas por la posible formación de un enlace de hidrógeno con la OLU.
5. Al comparar los métodos de análisis de la energía libre de interacción se observa una consistencia en los resultados obtenidos, tanto en la convergencia de la energía libre de interacción en -33,89 kcal/mol, como en los resultados de la descomposición por aminoácidos. Por lo que se concluye que al dar resultados estadísticamente idénticos, el método de análisis

de clústeres es más eficiente, debido a que este requiere menor tiempo de cálculo con lo que se pueden abaratar costes en futuras investigaciones.

6. El cálculo del espectro de emisión del sistema OLU/luciferasa muestra un pico de emisión en los 2,37 eV, que es suficientemente cercano al valor experimental de 2,21 eV, quedando dentro del error aceptable por el método de cálculo.

Este es un estudio preliminar de la influencia del entorno proteico sobre la OLU. En estudios posteriores sería interesante analizar posibles mutaciones en los aminoácidos más influyentes en la energía de interacción enzima/sustrato, como lo son HIE247, TYR342, PHE252, ARG220 o ARG334, con el objetivo controlar la longitud de onda de emisión.

9. Bibliografía

- [1] Fleiss, A., & Sarkisyan, K. S. (2019). *Current Genetics*, 65(4), 877-882.
- [2] Shimomura, O., Goto, T., & Hirata, Y. (1957). *Bulletin of the Chemical Society of Japan*, 30(8), 929-933.
- [3] Stevani, C. V., Oliveira, A. G., Mendes, L. F., Ventura, F. F., Waldenmaier, H. E., Carvalho, R. P., & Pereira, T. A. (2013). *Photochemistry and Photobiology*, 89(6), 1318-1326.
- [4] Al-Handawi, M. B., Polavaram, S., Kurlevskaya, A., Commins, P., Schramm, S., Carrasco-López, C., & Naumov, P. (2022). *Chemical Reviews*, 122(16), 13207-13234.
- [5] Kaskova, Z. M., Tsarkova, A. S., & Yampolsky, I. V. (2016). *Chemical Society Reviews*, 45(21), 6048-6077.
- [6] White, E. H., Wörther, H., Seliger, H. H., & McElroy, W. D. (1966). *Journal of the American Chemical Society*, 88(9), 2015-2019.
- [7] Branching, B. R., Hayward, M. M., Bamford, S., M Brennan, & Lajiness, E. J. (1989). *Photochemistry and Photobiology*, 49(5), 689-695.
- [8] Takakura, H., Sasakura, K., Ueno, T., Urano, Y., Terai, T., Hanaoka, K., & Nagano, T. (2010). *Chemistry An Asian Journal*, 5(9), 2053-2061.
- [9] Wood, K. V., Lam, Y. A., & McElroy, W. D. (1989). *Journal of Bioluminescence and Chemiluminescence*, 4(1), 289-301
- [10] Branchini, B. R., Behney, C. E., Southworth, T. L., Fontaine, D. M., Gulick, A. M., Vinyard, D. J., & Brudvig, G. W. (2015). *Journal of the American Chemical Society*, 137(24), 7592-7595.
- [11] Lee, H., Woo, S. M., Jang, H., Kang, M., & Kim, S. Y. (2022, November). *Seminars in Cancer Biology* (Vol. 86, pp. 347-357). Academic Press
- [12] Nogawa, M., Yuasa, T., Kimura, S., Tanaka, M., Kuroda, J., Sato, K., & Maekawa, T. (2005). *The Journal of Clinical Investigation*, 115(4), 978-985.

- [13] Contag, C. H., Spilman, S. D., Contag, P. R., Oshiro, M., Eames, B., Dennery, P., & Benaron, D. A. (1997). *Photochemistry and Photobiology*, 66(4), 523-531.
- [14] Patel, A. M., Henley, A., Parkes, M. A., Assmann, M., Worth, G. A., Anderson, J. C., & Fielding, H. H. (2020). *Physical Chemistry Chemical Physics*, 22(34), 19022-19032
- [15] Miller III, B. R., McGee Jr, T. D., Swails, J. M., Homeyer, N., Gohlke, H., & Roitberg, A. E. (2012). *Journal of Chemical Theory and Computation*, 8(9), 3314-3321
- [16] Grubmüller, H.; Heller, H.; Windemuth, A.; Schulten, K., (1991). *Mol. Simul.*, 6 (1–3), 121–142.
- [17] Dreuw, A., & Head-Gordon, M. (2005). *Chemical Reviews*, 105(11), 4009-403
- [18] Davidchack, R. L., Handel, R., & Tretyakov, M. V. (2009). *The Journal of Chemical Physics*, 130(23).
- [19] Berendsen, H. J., Postma, J. V., Van Gunsteren, W. F., DiNola, A. R. H. J., & Haak, J. R. (1984). *The Journal of Chemical Physics*, 81(8), 3684-3690.
- [20] Hohenberg, P., & Kohn, W. J. P. R. (1964). *Physical Review*, 136(1964), B864.
- [21] Runge, E., & Gross, E. K. (1984). *Physical Review Letters*, 52(12), 997.
- [22] Nogueira, J. J., & González, L. (2018). *Annual Review of Physical Chemistry*, 69, 473-497.
- [23] Nakatsu, T., Ichiyama, S., Hiratake, J., Saldanha, A., Kobashi, N., Sakata, K., & Kato, H. (2006). *Nature*, 440(7082), 372-376.
- [24] Case, D. A., Duke, R. E., Walker, R. C., Skrynnikov, N. R., Cheatham III, T. E., Mikhailovskii, O., & ET, A. (2022). 235-271
- [25] Singh, U. C., & Kollman, P. A. (1984). *Journal of Computational Chemistry*, 5(2), 129-145.
- [26] Yanai, T., Tew, D. P., & Handy, N. C. (2004). *Chemical Physics Letters*, 393(1-3), 51-57.
- [27] Francl, M. M., Pietro, W. J., Hehre, W. J., Binkley, J. S., Gordon, M. S., DeFrees, D. J., & Pople, J. A. (1982). *The Journal of Chemical Physics*, 77(7), 3654-3665.
- [28] Krishnan, R. B. J. S., Binkley, J. S., Seeger, R., & Pople, J. A. (1980). *The Journal of Chemical Physics*, 72(1), 650-654.
- [29] McLean, A. D., & Chandler, G. S. (1980). *The Journal of Chemical Physics*, 72(10), 5639-5648.
- [30] Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., & Case, D. A. (2004). *Journal of Computational Chemistry*, 25(9), 1157-1174.
- [31] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). *The Journal of Chemical Physics*, 79(2), 926-935.
- [32] Tian, C., Kasavajhala, K., Belfon, K. A., Raguette, L., Huang, H., Migués, A. N., & Simmerling, C. (2019). *Journal of Chemical Theory and Computation*, 16(1), 528-552.
- [33] Roe, D. R., & Cheatham III, T. E. (2013). *Journal of Chemical Theory and Computation*, 9(7), 3084-3095.