

---

# Trabajo de Fin de Grado

## **INFLUENCIA DE LAS MUTACIONES DE LA ENZIMA LUCIFERASA EN LA BIOLUMINISCENCIA DE LA OXILUCIFERINA**

Miguel Ángel Bayo Valero

Departamento de Química

Facultad de Ciencias - Universidad Autónoma de Madrid

Curso 2022/2023

---

---

## Índice

1. Abstract.....	1
2. Resumen.....	1
3. Introducción.....	2
4. Objetivos.....	5
5. Plan de trabajo.....	6
6. Métodos de cálculo.....	6
6.1 Dinámica Molecular Clásica.....	7
6.2 Métodos de cálculo de energía libre de interacción.....	10
6.3 Teoría del Funcional de la Densidad y su dependencia con el tiempo.....	11
7. Resultados y discusiones.....	13
7.1 Detalles computacionales.....	13
7.1.1 Parametrización del campo de fuerzas por el método de Seminario.....	13
7.1.2 Dinámica Molecular Clásica.....	14
7.1.3 Cálculos QM/MM.....	14
7.2 Equilibrado y convergencia de la trayectoria.....	15
7.3 Energía libre de interacción.....	17
7.3.1 Análisis de la trayectoria.....	17
7.3.2 Análisis de clústeres.....	21
7.4 Espectros de emisión.....	24
8. Conclusiones.....	25
9. Bibliografía.....	26

---

---

## 1. Abstract

Bioluminescent systems have always drawn attention of researchers thanks to their infrequent light emission feature, which is produced from a chromophore interacting with an enzyme (luciferin/luciferase system). This flamboyant characteristic is now found to present a wide range of promising applications in the field of biomedicine, with features that go from bioimaging of mammalian tissues to monitorization of the metabolic development of cancer and tumorous cells in a non-invasive way. The development of new and upgraded applications requires a minute study of those bioluminescent systems in order to control their bioluminescence.

This investigation project addresses a computational study of the photophysics of the complex oxyluciferin/luciferase bioluminescent system and the influence of an amino acid mutation in its bioluminescence that could have consequences in its possible biomedical applications by using Classical Molecular Dynamics (MD) and hybrid quantum mechanics/molecular mechanics calculations. A reparametrization of the force field together with a MD simulation of the natural system (wild type, wt) have been performed in order to calculate the binding free energy between luciferin and the enzyme using the *Molecular Mechanics with Generalized Born Surface Area* (MMGBSA) classical method applying two different protocols. The first one take into account hundreds of geometries from the classical trajectory in the free energy calculation. In the second protocol, a cluster analysis that classify the geometries in few groups by structural similarities (clusters) is carried out, and only one representative geometry from each cluster is chosen to perform the free energy calculation, so it is much more efficient from the computational point of view. It is confirmed that the second protocol, even being greatly efficient, provides almost equal results as the first one. Moreover, the binding free energy has been decomposed in different contributions from the closest amino acids to the chromophore in order to identify the residue that presents the greatest contribution to the free energy. This residue, which turns out to be phenylalanine (PHE244), has been picked to mutated by an alanine (ALA), i.e. the mutation PHE244ALA was executed. Following the mutation, another MD simulation was performed and the binding free energy and its decomposition by amino acids were calculated again so as to the differences with respect the not-mutated system could be analyzed. Lastly, the emission spectra of the luciferin interacting with both (wt and mutated) proteins have been calculated using the time dependent density functional theory and, in particular, the CAM-B3LYP/6-311G(2d,p) protocol. Both emission spectra have turned out to be closely akin, but a slight displacement towards to color red is caused because of the mutation. Future studies should investigate other mutations able to improve the emission properties of bioluminescent molecules to achieve new biomedical applications.

## 2. Resumen

Los sistemas bioluminiscentes siempre han atraído la atención de los investigadores gracias a su excepcional cualidad de emisión de luz, que se produce desde un cromóforo interaccionando con una enzima (sistema luciferina/luciferasa). Se ha encontrado que esta llamativa característica presenta una amplia gama de aplicaciones prometedoras en el campo de la

---

biomedicina, con funciones que van desde la bioimagen de tejidos de mamíferos hasta la monitorización del desarrollo metabólico de células cancerosas y tumorales de forma no invasiva. El desarrollo de nuevas y mejoradas aplicaciones requiere estudiar en detalle dichos sistemas bioluminiscentes para así poder controlar su bioluminiscencia.

Este proyecto de investigación aborda un estudio computacional de la fotofísica del complejo oxiluciferina/luciferasa y la influencia de una mutación de un aminoácido en su bioluminiscencia, que podría tener consecuencias en sus posibles aplicaciones biomédicas mediante el uso de la Dinámica Molecular Clásica (MD) y cálculos híbridos de mecánica cuántica y mecánica molecular (QM/MM). Se ha realizado una reparametrización de los campos de fuerza y una simulación MD del sistema natural (wild type, wt) con el fin de calcular la energía libre de interacción entre la luciferina y la enzima utilizando el método clásico *Molecular Mechanics with Generalized Born Surface Area* (MMGBSA) aplicando dos protocolos diferentes. El primero tiene en cuenta cientos de geometrías de la trayectoria clásica en el cálculo de energía libre. En el segundo protocolo, se realiza un análisis de clúster que clasifica las geometrías en unos pocos grupos según similitudes estructurales (clústeres), y escoge solamente una geometría representativa de cada clúster para el cálculo de energía libre, de manera que es mucho más eficiente desde el punto de vista computacional. Se ha comprobado que el segundo protocolo, aun siendo muy eficiente, proporciona resultados prácticamente iguales al primero. Además, la energía libre de interacción se ha descompuesto en las distintas contribuciones procedentes de los aminoácidos más cercanos al cromóforo, con el fin de identificar el residuo que tiene un mayor contribución a la energía libre. Este residuo, que ha sido una fenilalanina (PHE244), se ha escogido para mutarlo por una alanina (ALA), es decir, se ha llevado a cabo la mutación PHE244ALA. Después de la mutación, se ha realizado otra simulación MD y se ha calculado de nuevo la energía libre de interacción y su descomposición por aminoácidos para analizar las diferencias con respecto al sistema sin mutar. Por último, los espectros de emisión de la luciferina interaccionando con ambas proteínas (wt y mutada) se han calculado utilizando la teoría del funcional de densidad dependiente del tiempo y, en particular, el protocolo CAM-B3LYP/6-311G(2d,p). Los dos espectros de emisión han resultado ser muy similares, con un ligero desplazamiento hacia el rojo causado por la mutación. Los estudios futuros deberían investigar otras mutaciones capaces de mejorar las propiedades de emisión de las moléculas bioluminiscentes para lograr nuevas aplicaciones biomédicas.

### **3. Introducción**

La bioluminiscencia es un fenómeno de emisión de luz originado por reacciones químicas catalizadas por enzimas.<sup>[1]</sup> La interacción enzima-sustrato (luciferasa/luciferina) produce la oxidación de este último a oxiluciferina en estado electrónico excitado que, al relajarse al estado fundamental, emite radiativamente luz que cae dentro del espectro visible con un elevado rendimiento cuántico.

En la naturaleza existen un gran número de sistemas bioluminiscentes, de los cuales más de 30 han sido estudiados y analizados. Estos sistemas incluyen la D-luciferina (Figura 1) presente

---

en luciérnagas, algunos escarabajos y lombrices, como es el caso del gusano del ferrocarril. Este último presenta la capacidad de emitir a diferentes longitudes de onda en distintas partes del cuerpo debido a la interacción de diversas luciferasas con la luciferina. Por ejemplo, presenta emisión en verde amarillento en los costados alrededor de  $\lambda_{\text{max}} = 540\text{-}580\text{ nm}$  y rojo en la cabeza entorno a  $\lambda_{\text{max}} = 565\text{-}620\text{ nm}$ ).<sup>[2]</sup> Otros cromóforos bioluminiscentes son la *Cypridina* y la coelenterazina, encontradas en crustáceos marinos invertebrados como luciérnagas de mar y gambas de aguas profundas, respectivamente. Ambos son tripéptidos modificados (la coelenterazina se construye a partir de dos residuos de tirosina y uno de fenilalanina, mientras que la *Cypridina* a partir de triptófano, isoleucina y arginina), por lo que poseen una estructura similar, y por tanto, emisión parecida en la región del azul.<sup>[2, 3]</sup> También cabe destacar las luciferinas bacterianas (aldehído mirístico) presentes en bacterias del medio acuático como *Vibrio fischeri* que conviven en simbiosis con especies como el pez piña japonés (*Monocentris japonicus*)<sup>[4]</sup> y que emiten luz azul (~490 nm).<sup>[3]</sup> Otras luciferinas a mencionar son las fúngicas, con un pico máximo de emisión a 530 nm, presentes en hongos de las familias *Mycena* y *Omphalotus*, entre otras,<sup>[5]</sup> y las presentes en krill marino y organismos dinoflagelados.<sup>[3]</sup>

Este proyecto de investigación se centra en el estudio teórico de la emisión de bioluminiscencia por parte de la oxiluciferina como molécula emisora final en el mecanismo bioluminiscente de la D-luciferina, reflejado en la Figura 1. Analizando el mecanismo, se observa que el sustrato reacciona con adenosín trifosfato (ATP), junto con iones magnesio y la proteína luciferasa para dar el intermedio de reacción luciferil adenilato (2). Este presenta dos vías principales de reacción, una oscura y una lumínica. La primera ocurre con un 20% de rendimiento y comienza con la reacción del luciferil adenilato (2) con oxígeno molecular, para dar peróxido de hidrógeno junto con un éster conjugado conocido como dehidroluciferil adenilato (3a), el cual presenta una molécula de adenosín monofosfato (AMP) unida al oxígeno en posición alfa al carbonilo. El AMP entra en contacto con pirofosfato liberando una molécula de ATP, y deja como producto final dehidroluciferina (4a).

La segunda vía ocurre con un 80% de rendimiento y comienza con la oxidación del intermedio de reacción luciferil adenilato (2) mediante un mecanismo de transferencia monoelectrónico para obtener un peróxido (3b), que rápidamente pierde un AMP para cerrar el ciclo y formar una dioxietanona (4b). Con el fin de liberar la tensión del anillo de cuatro miembros, se libera  $\text{CO}_2$  y se genera la oxiluciferina ceto (5). Esta última, formada inicialmente en forma neutra, pierde un protón via transferencia protónica en estado excitado (a día de hoy aún se desconoce la especie aceptadora de este protón) dando lugar a la oxiluciferina ceto(-1) (5a) electrónicamente excitada, en equilibrio tautomérico con el correspondiente enol (oxiluciferina enol(-1) (5b) excitado), capaz de relajarse al estado fundamental emitiendo luz que cae en el visible y ganando el protón perdido anteriormente. En este estudio se modeliza la emisión de la oxiluciferina en su forma ceto (5a) ya que la energía de activación de la reacción de transferencia de protón en estado excitado para obtener la forma enol (5b) es muy elevada.

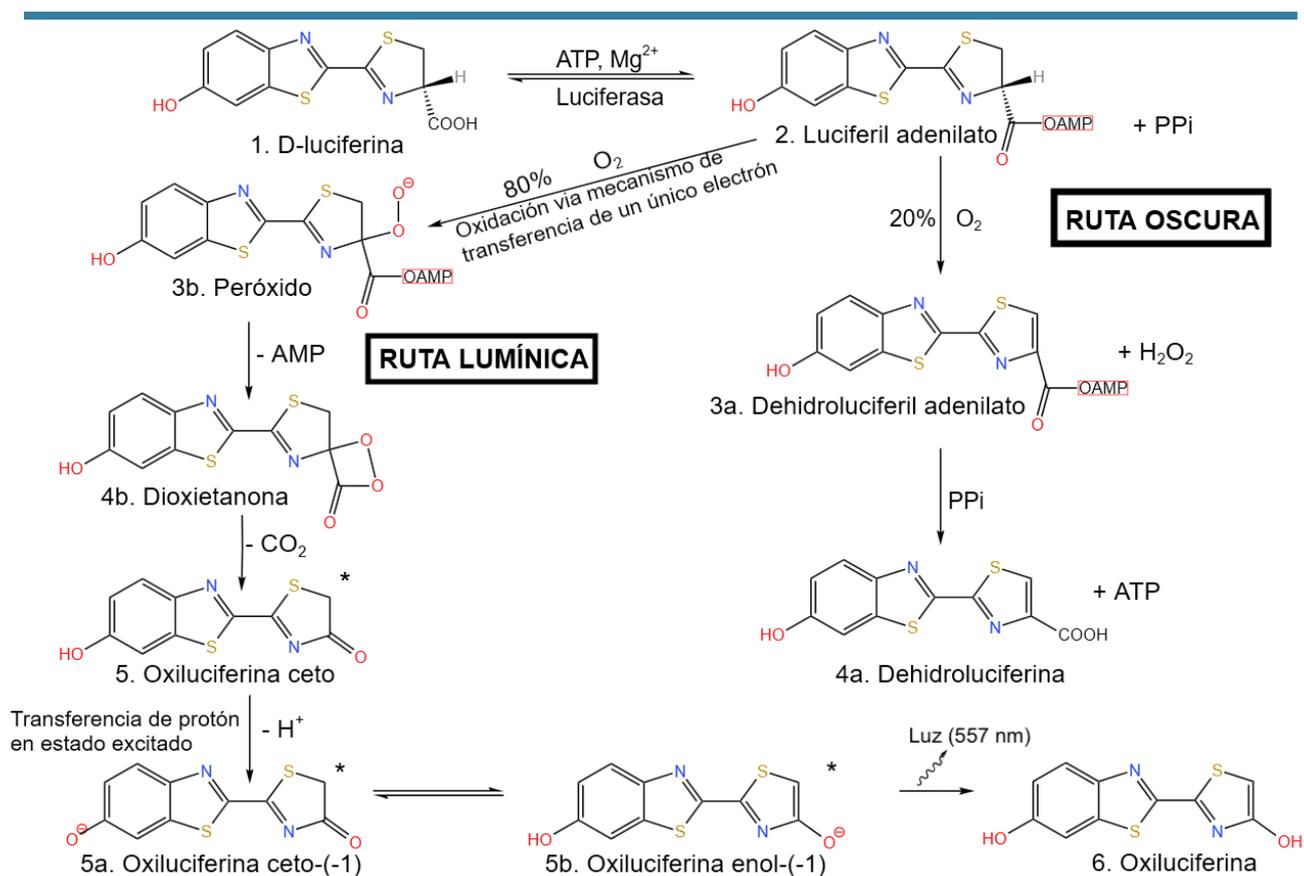


Figura 1. Mecanismo de la bioluminiscencia de la D-luciferina.

Dada la inestabilidad de las luciferinas naturales en disolución al contacto con el oxígeno, a su corta longitud de onda con respecto a la ventana de transparencia biomédica (intervalo rojo e IR cercano en el cual la luz es capaz de atravesar tejidos biológicos a gran profundidad)<sup>[3]</sup> y la posibilidad de dimerización en presencia de bases<sup>[6]</sup> se han desarrollado diferentes tipos de análogos sintéticos para mejorar las propiedades fotofísicas (Figuras 2B-G). Aquí se incluyen derivados aminados (Figuras 2B-F), los cuales presentan un desplazamiento de la emisión hacia el rojo, también conocido como desplazamiento batocrómico, al aumentar el grado de sustitución en la posición 6' del fragmento benzotiazol gracias a la capacidad dadora de electrones del grupo amino (Figuras 2B y C), aunque también se produce una disminución del rendimiento cuántico de emisión. Además, es posible acentuar el desplazamiento hacia longitudes de onda más largas de estos derivados aminados al combinarlos con una sustitución del heteroátomo de azufre por uno de selenio (Figura 2D). Otros tipos de análogos sintéticos son los que poseen grupos aromáticos dado que la presencia del benzotiazol no es necesaria para la emisión lumínica. En el caso de los derivados aromáticos de cumarina y quinoleno (Figuras 2E y F), se observan desplazamientos hipsocrómicos y batocrómicos, respectivamente. La sustitución del grupo benzotiazol por un grupo aromático junto a un sistema  $\pi$  conjugado (Figura 2G) promueve la formación de cromóforos que emiten desde azul a rojo en función de la extensión de la cadena. Cada olefina aumenta la longitud de onda de emisión en aproximadamente 100 nm ( $\lambda = 440$  nm para  $n=1$ ,  $\lambda = 530$  nm para  $n=2$  y  $\lambda = 640$  nm para  $n=3$ ). La intensidad de emisión de estos compuestos aumenta lentamente hasta que permanece constante aunque, aun así, es bastante

baja. Sin embargo, se ha encontrado que la baja intensidad que presentan este tipo de compuestos aumenta considerablemente con la adenilación.<sup>[3]</sup>

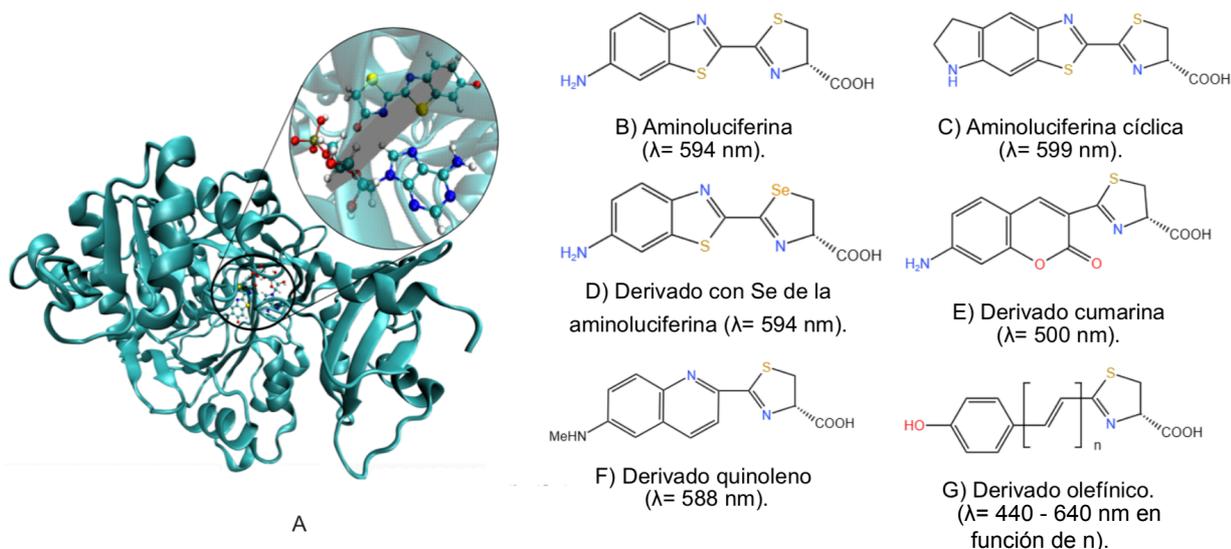


Figura 2. A) Estructura del complejo D-luciferina/luciferasa. B-G) Derivados sintéticos y su longitud de onda de emisión.

En este estudio se modeliza la emisión de la oxiluciferina en su forma ceto (5a) porque conocer en detalle su fotofísica permitirá comprender cómo puede usarse en ciertas aplicaciones. Una de las aplicaciones más relevantes tiene que ver con la reactividad de las moléculas de ATP localizadas intracelularmente, ya que son responsables de la producción de energía de las células en las mitocondrias y, por ende, de la vida de las mismas. Esta relación es ampliamente estudiada dada su aplicabilidad biomédica en la bioimagen de tejidos de mamíferos con luz roja e IR cercana y el seguimiento del desarrollo metabólico de tumores y cánceres de manera no invasiva con elevada sensibilidad y especificidad. Dado que la síntesis de ATP en este tipo de células ocurre mediante glicólisis anaeróbica en lugar de fosforilación oxidativa aeróbica, también conocido como efecto Warburg,<sup>[8, 9]</sup> una disminución significativa de la cantidad de ATP puede ser indicativo de disfunción mitocondrial asociada al cáncer.<sup>[1 - 3, 6, 7]</sup> Además, presenta ausencia de autofluorescencia y poco ruido de fondo, por lo que es utilizada en estudios *in vivo* e *in vitro*. Además de las ya mencionadas, el complejo D-luciferina/luciferasa (Figura 2A) presenta un amplio rango de aplicaciones. Entre ellas cabe destacar, la detección de fármacos, drogas, bacterias y virus causantes de enfermedades infecciosas, estudio de interacciones proteína-proteína y proteína-ligando, análisis de metabolitos involucrados en comunicación y señalización celular,<sup>[2]</sup> creación de medicamentos anticarcinogénicos y análisis citotóxicos antimicrobianos, entre otras.<sup>[3]</sup>

#### 4. Objetivos

Este proyecto de investigación tiene dos objetivos principales. El primero de ellos, más general, consiste en el estudio y modelización computacional de la fotofísica del complejo oxiluciferina/luciferasa, junto con la caracterización teórica de los espectros de emisión del cromóforo con la ayuda de programas de computación biofísica.

---

El segundo objetivo, más específico, radica en el estudio del efecto de la mutación de aminoácidos de una proteína natural (wild type, wt) en la energía libre de interacción y consecuentemente, su relación con la bioluminiscencia, lo cual podría proporcionar información relevante para futuras aplicaciones biomédicas.

Además de estos objetivos concretos relativos al estudio de la bioluminiscencia de proteínas, se buscan otros transversales como el aprendizaje y la familiarización con Amber20, el programa de dinámica molecular; Gaussian, de química cuántica, los visualizadores Gaussview, Pymol y Visual Molecular Dynamics (VMD) y con el sistema operativo Linux. Asimismo, se pretende comprender cómo se desarrolla la investigación en Química Teórica a través de la obtención y razonamiento crítico de resultados y lectura bibliográfica.

## 5. Plan de trabajo

**Semanas 1 - 6:** Búsqueda y lectura bibliográfica sobre el tema a tratar, el sistema luciferina/luciferasa y análogos y los métodos computacionales empleados (DFT y Dinámica Molecular Clásica o MD, por sus siglas en inglés).

**Semanas 2 y 3:** Aprendizaje sobre el manejo del sistema operativo *Linux* y programas informáticos *Gaussian*, *Gaussview*, *Amber20*, *Pymol* y *VMD* mediante la realización de tutoriales.

**Semanas 3 y 4:** Montaje y reparametrización del sistema mediante cálculos cuánticos con *Gaussian09* y los paquetes *antechamber* y *tLeap* de *Amber20*.

**Semanas 5 y 6:** Cálculo de la energía libre de interacción total ( $\Delta G_{TOT}$ ) mediante análisis MMGBSA (Molecular Mechanics with Generalized Born and Surface Area) de la proteína wt. Estudio desglosado de las contribuciones de los aminoácidos más cercanos al centro activo con el fin de hallar el de mayor contribución y mutarlo.

**Semanas 8 y 9:** Análisis de *clustering*, tanto para la geometría wt como la mutada, en donde las geometrías obtenidas de la dinámica se clasifican en varios grupos o *clusters* según sus similitudes estructurales.

**Semanas 9 - 11:** Cálculo de la energía libre de interacción para la geometría representativa de cada *cluster*. El objetivo es comprobar si estas geometrías representativas proporcionan el mismo resultado que el cálculo anterior realizado con cientos de geometrías. Esto se hará tanto para la proteína wt como para la mutada.

**Semanas 11 - 13:** Cálculo de los espectros de emisión de la luciferina interaccionando con la proteína wt y mutada.

**Semanas 13 - 16:** Redacción del manuscrito final y preparación de la defensa pública

## 6. Métodos de cálculo

Durante la realización de este proyecto de investigación se ha recurrido a diferentes métodos de cálculo. Se ha utilizado desde dinámica molecular clásica (MD, por sus siglas en inglés) hasta mecánica cuántica (QM, por sus siglas en inglés) pasando por cálculos híbridos de mecánica cuántica/mecánica molecular (QM/MM, por sus siglas en inglés). Además, se han realizado

cálculos de la energía libre de interacción mediante el método MMGBSA (Molecular Mechanics with Generalized Born Surface Area) para realizar un estudio completo del sistema luciferasa/luciferina natural junto con el sistema luciferasa mutada/luciferina.

### 6.1 Dinámica Molecular Clásica

El hecho de estudiar el proceso biológico de la bioluminiscencia del sistema luciferina/luciferasa presenta un problema, el elevado número de átomos que conforman el sistema (del orden de  $10^4$ ). Por ello, se realiza MD para muestrear todas las configuraciones que son térmicamente accesibles y que proporcionan información relevante sobre las propiedades del sistema. Para ello, han de resolverse las ecuaciones de movimiento de Hamilton (Ecuaciones [1] y [2]) cuando se supone que el sistema es conservativo, es decir, la energía potencial depende solamente de las posiciones de los átomos y no del tiempo.

$$\frac{dp_i}{dt} = - \frac{dV(q_i)}{dq_i} \quad [1]$$

$$\frac{dq_i}{dt} = \frac{dT_i(p_i)}{dp_i} \quad [2]$$

donde  $p$  es el momento lineal,  $t$  el tiempo,  $q$  las coordenadas de cada átomo,  $V$  la energía potencial y  $T$  la energía cinética.

Sin embargo, estas ecuaciones no tienen solución exacta y deben resolverse de manera numérica utilizando distintos métodos. Al aplicar expansiones de Taylor truncadas a segundo orden, puede construirse el llamado algoritmo de velocidad de Verlet para resolver MD:

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

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

donde  $m$  es la masa de cada átomo,  $\Delta t$  el paso de tiempo y  $v$  la velocidad.

La resolución de este sistema requiere unas coordenadas ( $q(t)$ ) y velocidades ( $v(t)$ ) iniciales que son desconocidas y necesitan ser aproximadas. Las coordenadas se pueden obtener a partir de resultados experimentales o modelos teóricos mientras que las velocidades se toman de una distribución de Maxwell-Boltzmann a una temperatura dada, suponiendo equilibrio térmico. También se necesita conocer la expresión del gradiente de energía potencial ( $\frac{dV(t)}{dq}$ ), el

cual puede calcularse mediante la aplicación de un campo de fuerzas (FF, por sus siglas en inglés), resolviendo la ecuación de Schrödinger con métodos *ab initio*, o combinando estos dos últimos en cálculos QM/MM. En este trabajo se ha optado por la aplicación de campos de fuerza. Éstos son un conjunto de ecuaciones simples que relacionan la energía potencial de un sistema con sus coordenadas internas, es decir, con la geometría de cada una de las moléculas que lo conforma. Para ello, se expresa la energía potencial en función de interacciones enlazantes (de enlace, ángulos y diedros) y no enlazantes (atractivas y repulsivas) de acuerdo con la expresión:

$$V_{TOT} = \sum_i^{N_{enlace}} V_{enlace} + \sum_i^{N_{ángulo}} V_{ángulo} + \sum_i^{N_{diedro}} V_{diedro} + \sum_i^{N_{no\ enlazante}} V_{no\ enlazante} \quad [5]$$

Las interacciones entre átomos enlazados (interacciones 1-2) y ángulos entre átomos (interacciones 1-3) se describen mediante potenciales armónicos, por lo que únicamente son correctos entorno a la geometría de equilibrio y a temperaturas no muy elevadas, además, no describen ni la formación ni la ruptura de enlaces por lo que no pueden ser usados para describir reacciones químicas. Las interacciones entre átomos separados por 3 enlaces (interacciones 1-4), diedros o torsión se describen mediante un potencial de Fourier. Las interacciones no enlazantes engloban las interacciones coulómbicas, las dipolo-dipolo atractivas y las de repulsión, ambas descritas por un potencial de Lennard-Jones. La Ecuación [5] puede reescribirse de manera completa tal que:

$$V_{TOT} = \sum_{i=1}^{N_{enlace}} \frac{1}{2} k_b (r - r_0)^2 + \sum_{i=1}^{N_{ángulo}} \frac{1}{2} k_a (\theta - \theta_0)^2 + \sum_{i=1}^{N_{diedro}} k_t (1 + \cos(n\omega - \gamma)) + \sum_{i>j}^{N_{no\ enlazante}} \frac{c_i c_j}{4\pi\epsilon r_{ij}} + 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad [6]$$

donde  $k_b$ ,  $k_a$  son las constantes de fuerza de los potenciales que describen los enlaces y ángulos, respectivamente,  $r$  es la distancia de enlace,  $r_0$  es la distancia de enlace en la geometría de equilibrio,  $\theta$  es el ángulo entre dos átomos,  $\theta_0$  es el ángulo entre dos átomos en la geometría de equilibrio,  $k_t$  es la altura de la barrera energética necesaria para rotar el diedro,  $n$  es el número de mínimos de energía al rotar los átomos que conforman el ángulo diedro,  $\omega$  es el ángulo diedro formado entre 4 átomos,  $\gamma$  es el ángulo que determina la posición del mínimo,  $c_i$  y  $c_j$  corresponden a las cargas de los átomos  $i$  y  $j$ ,  $\epsilon$  es la permisividad eléctrica en el vacío,  $r_{ij}$  es la distancia entre los átomos  $i$  y  $j$ ,  $\epsilon_{ij}$  corresponde a la profundidad del pozo de energía del potencial de Lennard-Jones y  $\sigma_{ij}$  corresponde con la separación entre átomos cuando la energía potencial es nulo. Dado que realizar cálculos de dinámica molecular para un sistema macroscópico es computacionalmente inviable, se aplican condiciones periódicas de contorno (PBC, por sus siglas en inglés). Esto consiste en la repetición periódica de la celda unidad a lo largo del espacio tridimensional, de manera que si un átomo sale por un lado de la celda, otro igual y con el mismo vector de velocidad entra por el opuesto para tener conservación de la materia. Esto permite realizar los cálculos asegurando que un mayor porcentaje de átomos interactúan entre sí en lugar de con el vacío en los límites de la celda.

---

Dado que se aplican condiciones periódicas de contorno y la celda unidad se repite periódicamente, se calculan las interacciones no enlazantes entre átomos dentro de una cota o región, en lugar de para todo el sistema, para que el cálculo sea viable desde el punto de vista económico. Sin embargo, esta aproximación presenta el problema de la continuidad de energía con la distancia. En el caso del potencial de Lennard-Jones, como es de corto alcance, decae a 0 muy rápidamente y cerca del límite de la cota presenta interacciones muy bajas y fáciles de suavizar, por lo que se impone que la energía en el límite de la cota sea nula mediante la adición de una función de corrección que hace que el potencial disminuya suavemente desde un poco antes del límite de la cota. En el caso de las interacciones electrostáticas de Coulomb, como son de largo alcance, presenta una contribución muy significativa en el límite de la cota por lo que aplicar una función similar a la utilizada para el potencial de Lennard-Jones conduciría a mucho error. Por esto, se recurre a la sumatoria de Ewald, que transforma las interacciones de largo alcance en unas de corto alcance al aplicar una distribución gaussiana de signo opuesto sobre cada carga puntual del sistema en el espacio real, apantallándolas. Seguidamente, en el espacio recíproco se añaden distribuciones de carga de signo opuesto a las de las primeras distribuciones añadidas para contrarrestarlas y poder computar las interacciones reales de largo alcance.

Las simulaciones de dinámica molecular clásica de un sistema aislado pueden realizarse dentro de distintos conjuntos termodinámicos, como el canónico NVT donde el número de partículas  $N$ , el volumen  $V$  y la temperatura  $T$  se conservan y permanecen constantes o el isobárico-isotérmico (NPT) donde el número de partículas  $N$ , la presión  $P$  y la temperatura  $T$  permanecen constantes.<sup>[31]</sup> Sin embargo, es también de suma importancia el hecho de incluir un control de las propiedades termodinámicas del sistema dado su interés experimental, para poder asegurar que la energía del sistema de estudio se conserva y que las propiedades termodinámicas que se quieren controlar permanezcan constantes. Concretamente, se discutirá el control de la temperatura y la presión, que están directamente relacionadas con la velocidad y el volumen del sistema de estudio. Existen varios métodos de control que se pueden aplicar a las propiedades medibles del sistema como: <sup>[32]</sup>

1.) Control diferencial: Se fija una propiedad termodinámica a un valor y no se producen fluctuaciones entorno a ese valor medio.

2.) Control proporcional: Las variables, acopladas a una propiedad termodinámica, se corrigen en cada paso o cada cierto número de pasos en función de una constante de acoplamiento.

3.) Control integral: Se expande el hamiltoniano añadiendo nuevos términos que representan el efecto de un sistema externo que fija el estado que se pretende alcanzar. La evolución temporal de estos términos viene regida por las ecuaciones de movimiento hamiltonianas (ecuaciones [1] y [2]).

4.) Control estocástico: Los valores de las variables acopladas a las propiedades termodinámicas se propagan con ecuaciones de movimiento modificadas para poder controlar cada propiedad de manera aleatoria.

---

Existe una gran variedad de mecanismos de control de temperatura de un sistema, como pueden ser el termostato de Andersen, el termostato de Berendsen y el termostato de Langevin. En este trabajo se utilizó este último.

El termostato de Langevin<sup>[26]</sup> es un sistema de control estocástico que asegura que la temperatura sea constante con el fin de evitar expansiones anómalas del sistema. Esto se consigue mediante la introducción simulada del sistema en un baño externo con partículas ficticias que colisionan con las partículas del sistema, de manera que se consigue modificar el momento gracias a fuerzas aleatorias de fricción.

Con respecto al control de la presión existen numerosos barostatos como el de Monte Carlo, Nosé-Hoover y Berendsen. En este trabajo se utilizó este último. El barostato de Berendsen<sup>[27]</sup> es un sistema de control proporcional que garantiza presión constante mediante el reescalado continuo de la caja de solvatación del sistema.

## 6.2 Método de cálculo de energía libre de interacción

Tras correr las simulaciones de dinámica molecular, se calculó la energía libre de interacción entre la oxiluciferina y la enzima para cientos de geometrías escogidas de la dinámica. Esta energía se calculó siguiendo el método MMGBSA, que es un método de cálculo basado en campos de fuerza que usa una combinación de energías de mecánica molecular (MM) como fuerzas de van der Waals, fuerzas electrostáticas, solvatación polar y solvatación no polar,<sup>[23]</sup> para calcular la energía libre de interacción ( $\Delta G_{bind}$ ) entre complejo unido ( $\Delta G_{cp}$ ) y receptor no unido ( $\Delta G_{rp}$ ) y ligando ( $\Delta G_{lp}$ ) en solución tal que:<sup>[21]</sup>

$$\Delta G_{bind} = \Delta G_{cp} - \Delta G_{rp} - \Delta G_{lp} \quad [7]$$

Cada uno de estos términos se descompone en tres contribuciones: la energía calculada con MM en fase gas ( $\Delta E_{MM}$ ), la energía de solvatación ( $\Delta G_{solv}$ ) y la entropía a una temperatura dada  $T(\Delta S)$ .

$$\Delta G_{(cplrp/lp)} = \Delta E_{MM} + \Delta G_{solv} - T\Delta S \quad [8]$$

El primer término se calcula mediante campos de fuerza (ecuaciones [5 y 6]) y el término de solvatación se desdobra en solvatación polar ( $\Delta G_{solv\ polar}$ ) y no polar ( $\Delta G_{solv\ no\ polar}$ ). Las contribuciones polares se calculan mediante aproximaciones generalizadas de Born, que consideran el disolvente como un material continuo sin estructura molecular,<sup>[22-24]</sup> mientras que las no polares se calculan como función lineal del área de superficie accesible al disolvente.<sup>[24, 25]</sup> Dado el alto coste computacional del cálculo de la entropía y su leve relevancia en los cálculos, el término  $T\Delta S$  no se tiene en cuenta.

$$\begin{aligned} \Delta G_{(cplrp/lp)} = & \Delta E_{enlace} + \Delta E_{ángulo} + \Delta E_{diedro} + \Delta E_{electrostática} + \Delta E_{vdW} + \Delta G_{solv\ polar} \\ & + \Delta G_{solv\ no\ polar} - T\Delta S \end{aligned} \quad [9]$$

### 6.3 Teoría del Funcional de la Densidad y su dependencia del tiempo

Tras calcular la energía libre de interacción, se calcularon los espectros de emisión para ambas proteínas (wt y mutada) recurriendo a cálculos QM/MM. Estos cálculos se basan en la Teoría del Funcional de la Densidad (DFT, por sus siglas en inglés), la cual se centra en dos teoremas propuestos por Hohenberg y Kohn.<sup>[33]</sup> El primero de ellos dicta que la energía total de un sistema puede expresarse como un funcional de la densidad electrónica y que cualquier propiedad del estado fundamental puede ser calculado a partir de la densidad electrónica. Esta debe ser siempre positiva, su integración debe proporcionar el número de electrones y, además, determina de manera inequívoca el potencial externo al que están sometidos los electrones. De esta manera la energía electrónica del sistema, al expresarse como funcional de la densidad, se puede dividir en tres términos: la energía cinética de los electrones ( $T_e[\rho]$ ), la interacción repulsiva interelectrónica ( $E_{ee}[\rho]$ ) y un potencial externo que recoge la interacción entre los electrones y los núcleos ( $E_{ext}[\rho]$ ). Cabe mencionar que la interacción internuclear se considera constante dentro de la aproximación de Born-Oppenheimer.

$$E_e[\rho] = T_e[\rho] + E_{ee}[\rho] + E_{ext}[\rho] \quad [10]$$

El segundo teorema de Hohenberg y Kohn afirma que la densidad electrónica exacta del estado fundamental es aquella que minimiza la energía, es decir, la aplicación del teorema variacional al estudio del funcional de la densidad. Sin embargo, dado que la relación entre  $T_e[\rho]$  y la densidad electrónica es desconocida, no se puede obtener la energía de forma exacta. Es por ello que se recurre al teorema de Kohn-Sham,<sup>[34]</sup> donde se define un sistema ficticio de referencia en donde los electrones no interactúan, pero con la misma densidad que el sistema real, y cuya energía puede calcularse de forma exacta tal que:

$$E_{e,r}[\rho] = T_{e,r}[\rho] + E_{ext,r}[\rho] \quad [11]$$

Puesto que la interacción entre electrones no se tiene en cuenta, la energía calculada para el sistema de referencia difiere de la exacta. Esta pequeña diferencia de energía entre el sistema real y el de referencia es lo que se denomina como energía de correlación-intercambio y se define como:

$$E_{xc}[\rho] = E_e[\rho] - E_{e,r}[\rho] \quad [12]$$

La mayor dificultad de la DFT es que no se conoce la forma exacta del funcional de correlación-intercambio, por lo que se han desarrollado varios funcionales que se parametrizan para que reproduzcan propiedades calculadas con métodos más exactos o experimentales.

La DFT original únicamente puede aplicarse para computar el estado fundamental de un sistema y sus propiedades. Para estudiar los estados electrónicos excitados del sistema es necesario analizar la evolución temporal de la densidad del sistema bajo un potencial externo dado que existe una relación directa entre el potencial externo dependiente del tiempo de un sistema y su densidad dependiente del tiempo, según afirma teorema de Runge-Gross.<sup>[35]</sup> La

introducción de una dependencia temporal en la densidad electrónica permite reformular las ecuaciones DFT de manera que se puedan calcular las energías de los estados electrónicos excitados, y no solo la del estado fundamental.

Como se ha mencionado anteriormente, los espectros de emisión se calculan con QM/MM para ambas proteínas de estudio. Los cálculos QM/MM dividen el sistema en dos regiones tratadas con diferentes niveles de cálculo. La región interna (I), compuesta por la parte del sistema de interés, se trata de manera cuántica mientras que la externa (E) de manera clásica. Para realizar este tipo de cálculos, se recurre al acoplamiento electrostático donde se considera un hamiltoniano efectivo ( $\hat{H}^{ef}$ ) que tiene en cuenta el hamiltoniano de la región interna en el vacío de la región interna ( $(\hat{H}^0(I))$ ), junto con un término que describe la interacción de las N cargas de la región externa ( $c^E$ ) con el potencial electrostático de la región interna ( $V^I$ ), por lo que la ecuación de Schrödinger independiente del tiempo (Ecuación [13]) puede escribirse como:

$$\hat{H}\Psi = E\Psi \quad [13]$$

$$\hat{H}^{ef}|\Psi_I\rangle = (\hat{H}^0(I) + \hat{H}^0(I/E))|\Psi_I\rangle = E^{QM}(I)|\Psi_I\rangle \quad [14]$$

donde el hamiltoniano de interacción  $\hat{H}^0(I/E)$  viene dado por:

$$\hat{H}^0(I/E) = \sum_i^N c_i^E V^I \quad [15]$$

Para construir las funciones de onda ( $\Psi$ ) descritas en la Ecuación [14] se recurre a la técnica habitual en química cuántica que consiste en expresar orbitales moleculares ( $\Phi_i$ ) como una combinación lineal de las funciones de base centradas en los átomos ( $\chi_j$ ) o CLOA (Combinación Lineal de Orbitales Atómicos) tal que:

$$\Phi_i = \sum_j k_j \chi_j \quad [16]$$

donde  $k_j$  son los parámetros variacionales que permiten expresar a los orbitales atómicos  $\chi_j$  como combinación lineal. Una vez hallados los orbitales moleculares, se aplica el determinante de Slater para generar la función de onda:

$$\Psi = ||\Phi_i|| \quad [17]$$

En el caso de la DFT, la densidad electrónica de la que dependen los funcionales se construye a partir de los orbitales que se han obtenido del cálculo QM/MM que tiene en cuenta la interacción con el entorno, descrito de manera clásica:

$$\rho = \sum_i |\Phi_i|^2 \quad [18]$$

---

## 7. Resultados y discusión

Los resultados presentados en este apartado comprenden el estudio de la energía libre de interacción entre el cromóforo y la proteína calculada mediante el método MMGBSA para 1500 geometrías seleccionadas de manera equiespaciada de la región de la dinámica. Seguidamente, se realiza un análisis de convergencia para comprobar que el número de geometrías introducidas en el cálculo es suficiente y el cálculo ya ha convergido. Además, se ha llevado a cabo un análisis estructural para clasificar las distintas geometrías de la dinámica en *clusters* o grupos de geometrías asociados según la similitud estructural que presenten, para las proteínas wt y mutada. De cada uno de estos clústeres se ha determinado su geometría representativa y se ha calculado la energía libre de interacción para, finalmente hacer el promedio de las energías libres de interacción de las geometrías representativas. El objetivo de este análisis es comprobar si la energía libre calculada para toda la trayectoria es similar a la calculada para solamente unas pocas geometrías representativas. De ser así, la realización del segundo protocolo sería preferible porque el tiempo computacional es menor. Además del cálculo de la energía libre de interacción total también se ha obtenido su descomposición por residuo. Finalmente, se han obtenido mediante cálculos QM/MM los espectros de emisión de ambas proteínas. Por ello, los resultados se discutirán en bloques para lograr una comparación y análisis más exhaustivos. Sin embargo, el primer paso es comprobar que las simulaciones de dinámica molecular están convertidas y, de ser así, escoger la región de la dinámica para la cual se llevarán a cabo los análisis.

### 7.1 Detalles computacionales

En este epígrafe se tratarán algunos detalles computacionales sobre los métodos y cálculos llevados a cabo. Se comentará con brevedad sobre la reparametrización del campo de fuerzas aplicado, las fases llevadas a cabo durante la dinámica molecular y algunos detalles del cálculo QM/MM.

#### 7.1.1 Parametrización del campo de fuerzas mediante el método de Seminario

Los campos de fuerza vienen generalmente predefinidos para el estado electrónico fundamental. Sin embargo, como se ha comentado anteriormente, el luciferil adenilato se oxida produciendo la oxiluciferina en estado excitado, que más tarde se relajará y emitirá luz en el visible, por lo que será necesario realizar la reparametrización del FF mediante la optimización de la geometría en estado electrónico excitado de la oxiluciferina y el cálculo de la matriz hessiana para obtener las constantes de fuerza en base al método de Seminario.<sup>[36]</sup> Para ello, se obtienen por un lado las distancias y ángulos de enlace de equilibrio con un cálculo de optimización de geometría en vacío a nivel DFT con el funcional B3LYP<sup>[10-12]</sup> y la base 6-31G\*<sup>[14, 17, 18]</sup> para el primer estado electrónico excitado con el software Gaussian09.<sup>[19]</sup> El funcional híbrido B3LYP<sup>[10-12]</sup> incluye una parte del término de intercambio calculada de forma exacta con el método de Hartree-Fock y el resto de la energía de intercambio-correlación de datos empíricos u obtenidos mediante métodos más exactos.<sup>[37]</sup> La 6-31G\*<sup>[14, 17, 18]</sup> es una base *split valence* de Pople en la que se usa una base simple zeta con 6 funciones gaussianas primitivas (PGTO) en

---

los orbitales del core, mientras que usa una base doble zeta en los orbitales de valencia, que están divididos en una parte interna y otra externa descritas por 3 y 1 PGTos, respectivamente. Adicionalmente, se añaden funciones de polarización (representadas por \*) con un número cuántico  $l$  más alto que los orbitales de valencia. Por otro lado, las constantes de fuerza  $k_b$  y  $k_a$  se obtienen de la diagonalización de la matriz Hessiana, mientras que los parámetros de torsión se obtienen del General Amber Force Field (GAFF).<sup>[28]</sup>

### 7.1.2 Dinámica Molecular Clásica

Tras haber parametrizado el FF para el estado electrónico excitado, se lleva a cabo una simulación de dinámica molecular clásica (MD, por su siglas en inglés) tanto para la proteína natural como la mutada. Primero, se obtiene la información estructural de la proteína wt, esta proviene de datos experimentales sobre la estructura cristalina termoestable de la luciferasa de la luciérnaga japonesa (*Nipponoluciola cruciata*) complejada con oxiluciferina y AMP monoiónico y se encuentra disponible en la base de datos de proteínas (Protein Data Base, PDB ID: 2D1R).<sup>[39]</sup> Tras esto, se introduce el sistema proteína/cromóforo en una celda repleta de moléculas de agua a 12 Å con forma de octaedro truncado, que es más eficiente que una esfera, ya que con el mismo volumen, pueden asegurar condiciones periódicas de contorno. El sistema cuenta una carga total de -1, la cual se neutraliza con un contraión  $\text{Na}^+$  ya que los cálculos de MD se tienen que realizar en celdas neutras. A continuación, se realiza la minimización de todo el sistema iniciando con el método *steepest descent* y cambiando a *conjugate gradient* tras 5000 pasos, se aplica una rampa de calentamiento de 0 a 300 K a volumen constante para evitar expansiones anómalas del sistema y finalmente se corre la fase de producción del sistema a temperatura y presión constante mientras que el volumen se varía para poder controlar la densidad. En estos tres pasos, se aplica PBC y se recurre al uso del termostato de Langevin<sup>[26]</sup> y barostato de Berendsen<sup>[27]</sup> para el control de la temperatura y la presión, respectivamente.

Para describir el sistema, se realiza un cálculo con DFT a nivel B3LYP<sup>[13]</sup>/6-31G\*<sup>[14, 17, 18]</sup> para obtener las distancias y ángulos de la geometría de equilibrio del cromóforo. Además, aplicando el método Merz-Singh-Kollmann<sup>[29]</sup> se calculan las cargas del cromóforo y el AMP. Asimismo, se obtienen las constantes de fuerza  $k_a$  y  $k_b$  con el método de Seminario y se cogen los parámetros de torsión del GAFF<sup>[28]</sup> para describir al cromóforo. El resto de parámetros del AMP (ángulos, enlaces y torsión) vienen descritos en los parámetros del GAFF<sup>[28]</sup>. El agua fue descrita siguiendo el modelo TIP3P<sup>[30]</sup> y la luciferasa con un campo de fuerzas FF19SB.<sup>[42]</sup>

### 7.1.3 Cálculos QM/MM

En el último paso del estudio computacional del sistema luciferina/luciferasa, se obtiene el espectro de emisión de las proteínas wt y mutada teniendo en cuenta 10 estados excitados, con Gaussian09<sup>[19]</sup> con el fin de compararlos entre sí y con datos experimentales y estudiar la modulación de emisión de luz por parte de ambas proteínas.

Para el cálculo de los espectros de emisión se seleccionan 200 puntos de la trayectoria separados equidistantemente (pasos 2605-3998, correspondiente a 69,650 ns de la trayectoria

---

de 200 ns), y se calcula la energía y la fuerza del oscilador en cada uno de ellos, ya que este parámetro es directamente proporcional al coeficiente molar de extinción y a la intensidad de emisión. Finalmente, se superponen los espectros calculados para todas las geometrías aplicado una distribución gaussiana y se obtiene el espectro final de emisión. Para ello, se realiza un cálculo a nivel QM/MM dentro del marco del acoplamiento electrostático, esto es, se trata de manera cuántica la oxiluciferina (región interna) con Teoría del Funcional de la Densidad Dependiente del Tiempo (TD-DFT, por sus siglas en inglés) y el resto del sistema (región externa), de forma clásica introduciendo las cargas puntuales en el hamiltoniano. Para los cálculos cuánticos, se recurre al funcional CAM-B3LYP,<sup>[13]</sup> junto con la base de Pople 6-311G(2d,p)<sup>[14-16]</sup> donde los orbitales del core usan una base simple zeta con seis PGTO, mientras que los de valencia usan una base triple zeta con tres, uno y uno PGTOs. Además, se añaden dos funciones de polarización de tipo *d* para los átomos pesados (C, N, O y S en este caso) y un conjunto de tipo *p* para el hidrógeno.<sup>[38]</sup>

## 7.2 Equilibrado y convergencia de la dinámica

Una vez el campo de fuerzas ha sido parametrizado y el sistema ha sido montado, se lanzan los cálculos de dinámica molecular clásica. Los cálculos MD constan de tres etapas principales: minimización, termalización y producción. En la primera etapa, se minimiza la energía total del sistema a 0 K, es decir, se busca un mínimo en la superficie de energía potencial. Como se observa en la Figura 3A, la energía de ambos sistemas converge entorno a  $-2,6 \cdot 10^5$  kcal/mol. Tras esto, se pasa a la fase de termalización donde se aplica una rampa de calentamiento de 0 K a 300 K durante 1 ns (Figura 3B) aplicando el control térmico de Langevin dentro del conjunto canónico NVT para evitar expansiones anómalas del sistema. La última etapa corresponde a la fase de producción, donde se corre MD durante 200 ns dentro del conjunto isobárico-isotérmico NPT donde la temperatura se fija a 303,15 K con el termostato de Langevin<sup>[26]</sup> y la presión fija a 1 bar aplicando el control del barostato de Berendsen.<sup>[27]</sup> En esta última etapa, el volumen se varía para poder controlar la densidad, que como se observa en la Figura 3C (donde se representan solo los primeros 10000 ps), permanece constante en  $1\text{g/cm}^3$  durante toda la simulación, es decir, la densidad del sistema bioluminiscente se aproxima a la densidad del agua de la caja de solvatación donde está contenido. Asimismo, en la Figura 3D (solo los primeros 2000 ps representados) se muestra la descomposición de la energía total en las contribuciones de energía cinética y energía potencial durante la fase de producción. Durante esta fase, las posiciones y velocidades de los átomos del sistema se propagan, incrementando la contribución de energía cinética.

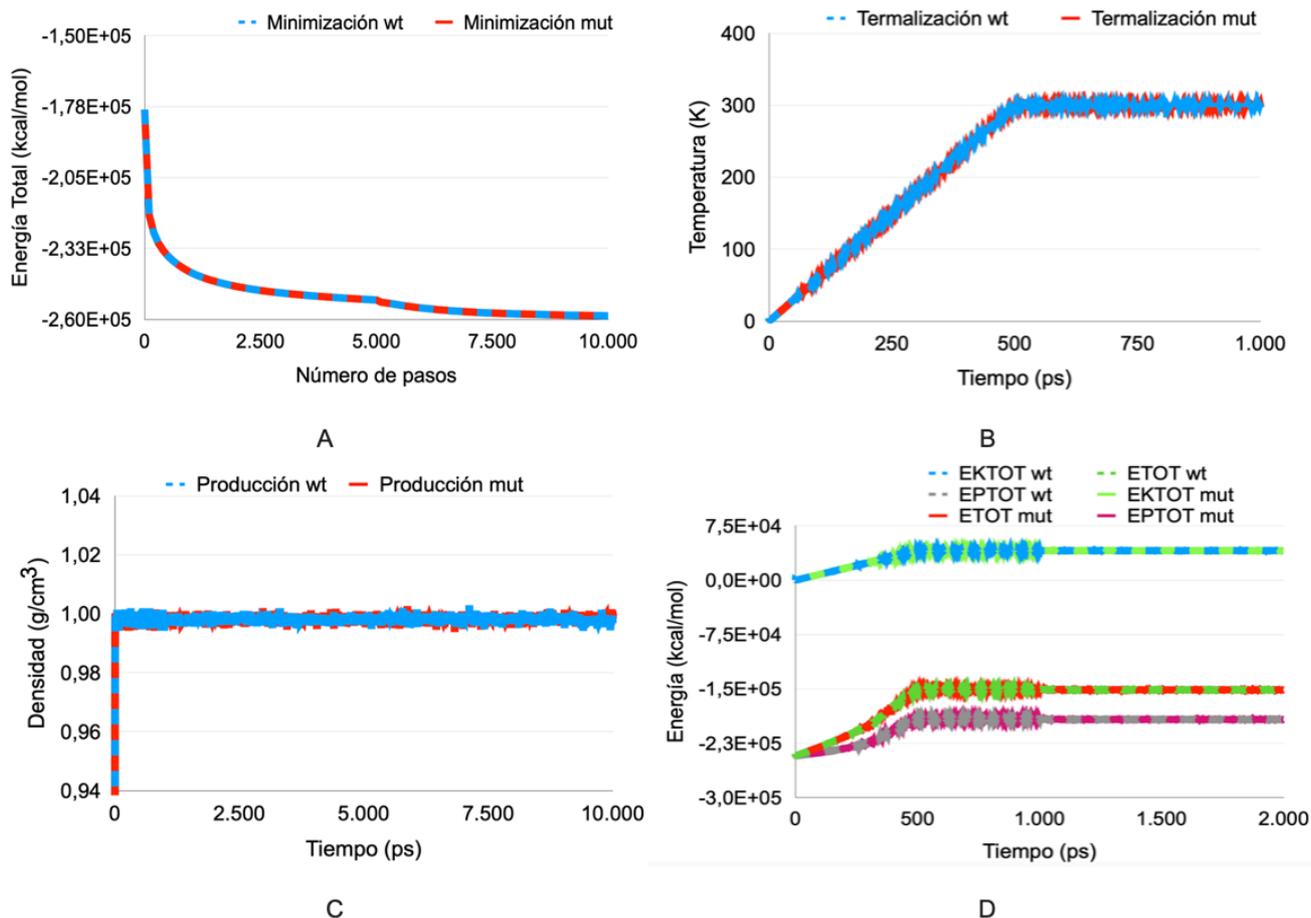


Figura 3. Variación de algunas propiedades del sistema durante el protocolo de simulación. (A) Variación de la energía total durante el proceso de minimización. (B) Variación temporal de la temperatura durante las etapas de calentamiento. Variación temporal de la (C) densidad y (D) las energías cinéticas (EKTOT), potencial (EPTOT) y total (ETOT) durante el proceso de producción. Se muestran los resultados para la proteína sin mutar o wild tape (wt) y para la proteína mutada (mut).

Una vez la dinámica molecular ha concluido, se estudia la estabilidad de las proteínas durante todo el proceso para hallar el tramo donde se estudiará la energía libre de interacción. Para ello, se recurre a la raíz del desplazamiento cuadrático medio (RMSD, por sus siglas en inglés). Esta es una medida estadística de precisión que es aplicada en análisis de estructuras macromoleculares y dinámicas para medir la distancia entre las geometrías obtenidas a lo largo de la dinámica y la geometría inicial. Para calcular la RMSD, se sigue la Ecuación [19], donde  $d_i$  es la distancia entre los átomos  $i$  a tiempo  $t$  y los átomos  $i$  en el momento inicial de la trayectoria y  $N$  es el número total de átomos.

$$RMSD = \sqrt{\frac{\sum_{i=1}^N d_i^2}{N}} \quad [19]$$

En la Figura 4 se representa la RMSD tanto de la proteína wt (en azul) como de la mutada (en rojo). Analizando esta gráfica, se observa que en el tramo comprendido entre 2500 y 4000 pasos ambas proteínas se equilibran y no se observan fluctuaciones importantes, por ende, el estudio

sobre la energía libre de interacción de cada proteína se realizará sobre este mismo espacio convergido de 1500 pasos en ambos casos. En sendas proteínas, la convergencia de la estructura ocurre en una distancia media entre átomos de alrededor de 3Å. Aunque es cierto que el sistema con la proteína wt presenta una RMSD ligeramente superior, también se observa que éste presenta mayores fluctuaciones que el sistema con la proteína mutada, especialmente alrededor de 2000 pasos. Por ende, la mutación parece disminuir la distancia entre las geometrías encontradas durante la trayectoria y la geometría inicial, pero sin llegar a marcar una diferencia significativa en temas de estabilización de la proteína.

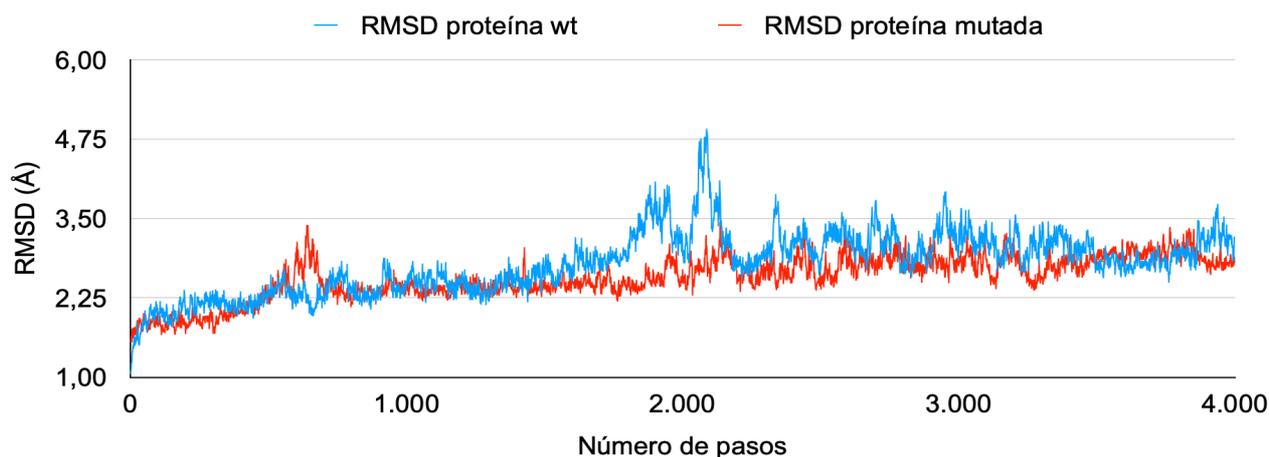


Figura 4. Representación gráfica de la RMSD de la proteína wt (azul) y la mutada (rojo).

### 7.3 Energía libre de interacción

Tras haber determinado el tramo donde las proteínas se equilibran, se realiza el estudio de la energía libre de interacción entre la oxiluciferina y la proteína separando el tramo estabilizado de la proteína (los 1500 últimos puntos) en distintos números de geometrías equiespaciadas y en agrupaciones con similitud estructural para calcular y comparar la energía libre de interacción obtenido por ambos métodos con el fin de hallar el más eficiente. Asimismo, se ha analizado la descomposición de la energía libre de interacción total por residuo para estudiar la influencia de la mutación del residuo con mayor contribución.

#### 7.3.1 Análisis de la trayectoria

Tras realizar los cálculos de MD y estudiar el rango de pasos donde las proteínas se encuentran estabilizadas con RMSD (Figura 4), es necesario conocer cuántos de esos 1500 pasos, del tramo de entre 2500 y 4000, son necesarios introducir en el cálculo para que este converja. Para ello, se realiza un análisis de convergencia de la energía libre de interacción entre el cromóforo y la proteína mediante el método MMGBSA<sup>[40]</sup> dividiendo el tramo de 1500 pasos en distintos intervalos separados equidistantemente con diferentes número de geometrías de la trayectoria que se introducen en el cálculo. Tras ello, se computa la energía libre de interacción total ( $\Delta G_T$ ), la cual se representa en la Figura 5.

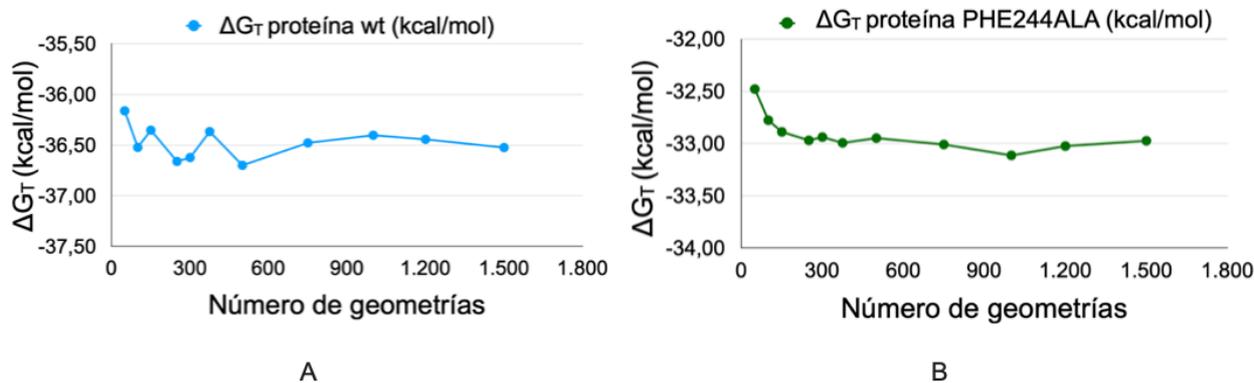


Figura 5. Estudio de convergencia del cálculo de energía libre de interacción total entre oxiluciferina y luciferasa de (A) la proteína wt y (B) proteína mutada.

Por un lado, como se observa en la Figura 5A, el cálculo de la energía libre de interacción total del cromóforo con la proteína wt no está convergido para un número pequeño de geometrías y se obtienen resultados oscilantes hasta la introducción de 500 geometrías. A partir de 750 geometrías, el cálculo continúa mostrando una pequeña variación de alrededor de 0,1 kcal/mol, por lo que se concluye que a partir de aquí el cálculo converge entorno a una  $\Delta G_T$  de  $-36 \pm 2$  kcal/mol.<sup>[40]</sup> Por otro lado, analizando la Figura 5B, se observa que la convergencia del cálculo de la energía libre de interacción total del cromóforo con la proteína mutada se alcanza a partir de 375 geometrías, donde el cálculo parece converger entorno a  $-33 \pm 3$  kcal/mol. Introduciendo un número menor a 375 geometrías, se aprecia que el cálculo no es estable y que por ende, la energía resultante sobreestima a la convergida.

En resumen, se observa que la energía libre de interacción de la oxiluciferina con la proteína PHE244ALA (Figura 5B) es de  $-33 \pm 3$  kcal/mol, mientras que la energía libre de interacción de la oxiluciferina con la proteína wt (Figura 5A) es de  $-36 \pm 2$  kcal/mol, por lo que la mutación produce una inestabilidad en el sistema de aproximadamente 3 kcal/mol. Esto se debe a que existen interacciones  $\pi$ - $\pi$  entre el fenilo de la cadena lateral de la fenilalanina y el benzotiazol de la oxiluciferina que se pierden al mutar este residuo a alanina, que presenta un metilo en la cadena lateral que no interacciona con la oxiluciferina. Asimismo, se observa que el cálculo de la energía libre de interacción de la oxiluciferina con la proteína mutada (Figura 5B) parece estabilizarse y converger antes que en el caso del cálculo de la energía libre de interacción del cromóforo con la proteína no mutada (Figura 5A). El primer cálculo converge a partir de 375 geometrías mientras que el segundo a partir de 750. Por ende, se opta por realizar el resto de cálculos incluyendo 1500 pasos ya que se ha demostrado que en este punto los cálculos de energía libre de interacción de la oxiluciferina con las proteínas han convergido, pero futuros estudios podrían realizarse en menor tiempo introduciendo 750 geometrías en el cálculo.

Además de estudiar la energía libre de interacción total entre cromóforo y proteína, con el método MMGBSA<sup>[40]</sup> es posible conocer la descomposición de esta energía en diferentes contribuciones en los aminoácidos más cercanos al cromóforo (Figura 6). Estas contribuciones incluyen las interacciones de van der Waals entre dipolos, las interacciones electrostáticas descritas por el potencial de Coulomb (ambas ocurren entre la proteína y la oxiluciferina) y las

interacciones de solvatación polar y no polar con el disolvente. Se utiliza CPPTRAJ<sup>[41]</sup> para analizar los últimos 1500 pasos de la trayectoria con el fin de encontrar los 15 aminoácidos o residuos más cercanos a la oxiluciferina durante toda la simulación y que se encuentran dentro de una cota del 3% del tiempo de la simulación como resultado de realizar el análisis MMGBSA<sup>[40]</sup> sobre 1500 geometrías.

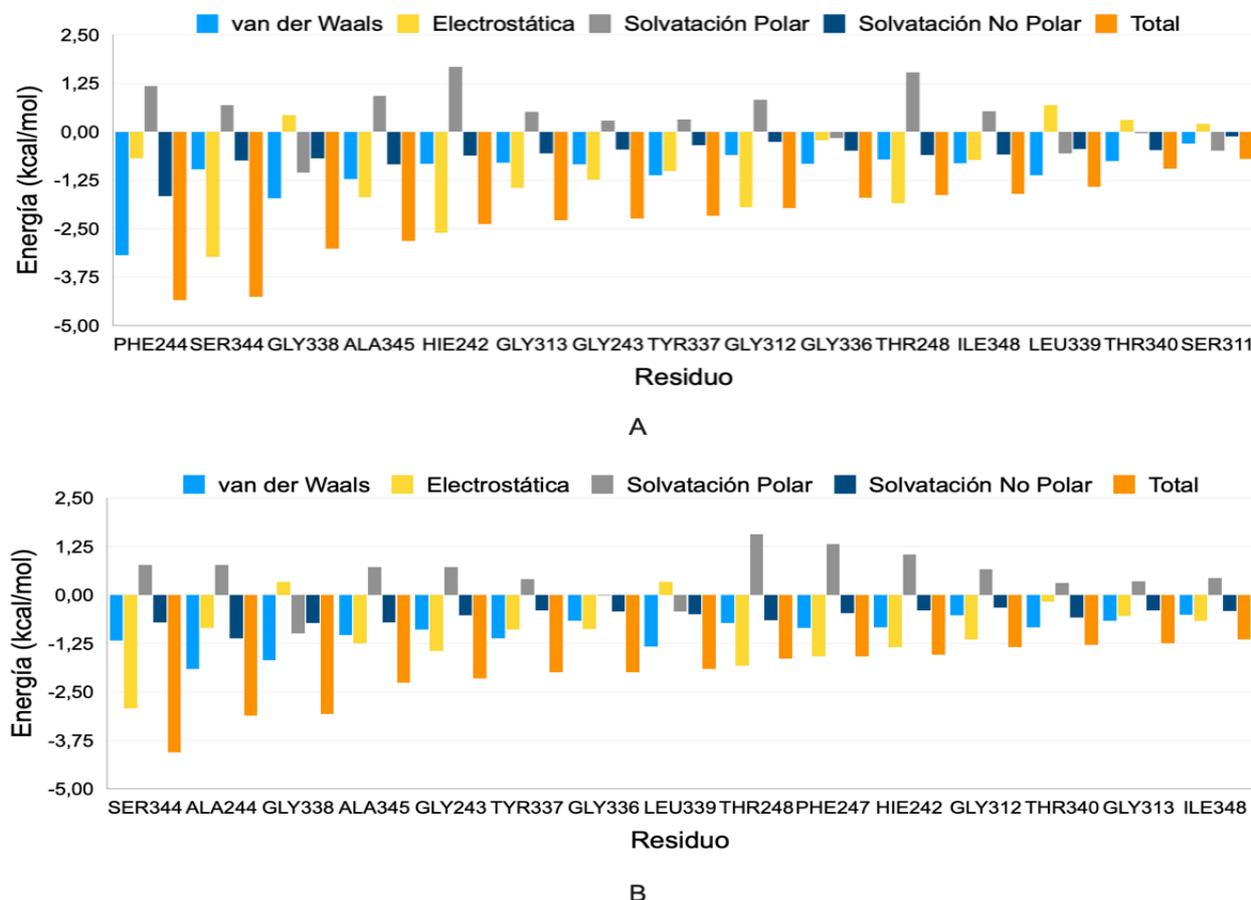


Figura 6. Análisis de la descomposición de la energía de interacción teniendo en cuenta los residuos más cercanos al cromóforo. (A) Interacción del cromóforo con la proteína wt e (B) Interacción del cromóforo con la proteína mutada.

Como se observa en las Figura 6A y B, se encuentran una gran variedad de aminoácidos proteicos con cadenas laterales variadas: aromáticas, presentes en fenilalanina (PHE) y tirosina (TYR); polares sin carga e hidrofílicas, en serina (SER) y treonina (THR); apolares alifáticas e hidrofóbicas, en glicina (GLY), alanina (ALA), isoleucina (ILE) y leucina (LEU) y básicas en histidina (HIE). Asimismo, se descomponen sus energías en diferentes contribuciones. Analizando las interacciones entre oxiluciferina y proteína representadas en las Figura 6A y B, se observa los residuos a en la posición 244 (PHE244 en Figura 6A y ALA244 en Figura 6B) presentan la contribución mayor en cuanto a fuerzas de van der Waals se refiere (-3,20 kcal/mol la PHE244 y -1,91 kcal/mol la ALA244). Sin embargo, ambos muestran las interacciones electrostáticas más bajas de todos los residuos (-0,68 kcal/mol la PHE244 y -0,85 kcal/mol la ALA244). Cabe destacar que la serina en la posición 344 (SER344) presenta la mayor contribución electrostática atractiva de entre todos los aminoácidos (-3,23 kcal/mol en la Figura 6A y -2,93 kcal/mol en la Figura 6B) mientras que el AMP presenta la más repulsiva en ambos

---

sistemas (18,83 kcal/mol en el sistema con proteína wt y 18,40 kcal/mol en el sistema con la proteína mutada).

Por otro lado, la solvatación en contacto con un disolvente polar como al agua se calcula como diferencia entre la interacción del complejo (oxiluciferina y proteína en conjunto) con el agua y la interacción de cada uno de los monómeros (oxiluciferina y proteína) por separado con el agua. Al estar separados, existe una interacción directa entre los monómeros y el disolvente que resulta en una gran contribución negativa. Sin embargo, al tratarlos de manera conjunta, pueden existir regiones donde la proteína y la oxiluciferina reaccionen entre sí en lugar de con el agua, resultando en una contribución conjunta menor. Por lo tanto, al realizar la diferencia entre las contribuciones de la interacción del complejo con el agua y las interacciones de cada monómero con el agua, el resultado final resulta en una contribución positiva, como se observa en la mayoría de los casos, acentuándose en HIE242 (1,67 kcal/mol) y THR248 (1,53 kcal/mol) en el caso de contar con la proteína wt y en THR248 (1,56 kcal/mol) y PHE247 (1,31 kcal/mol) al contar con la proteína mutada, mientras que en ambos sistemas se encuentra que ocurre todo lo contrario con AMP (presenta una energía de -18,55 kcal/mol en el sistema con la proteína wt y -18,45 kcal/mol en el sistema con la proteína mutada). Por otro lado, la solvatación con disolventes no polares se describe como la energía necesaria para crear una cavidad en el disolvente para introducir al soluto y se computa de manera análoga a la solvatación polar, tratándose como la diferencia entre la energía necesaria para crear una cavidad cuando el complejo está presente y cuando están los monómeros por separado. De manera global, se encuentra que se necesita más energía para crear una cavidad cuando se tratan la oxiluciferina y la proteína por separado que en conjunto, especialmente en los residuos en la posición 244, la fenilalanina (-1,66 kcal/mol) y el resultante de su mutación, alanina (-1,12 kcal/mol). Finalmente, se encuentra que los residuos con mayores contribuciones globales al tener en cuenta las interacciones de la proteína wt son los de PHE244 (-4,34 kcal/mol), SER344 (-4,26 kcal/mol) y GLY338 (-3,02 kcal/mol) mientras que en caso de contar con las interacciones con la proteína mutada son SER344 (-4,05 kcal/mol), ALA244 (-3,11 kcal/mol) y y GLY338 (-3,07 kcal/mol). Para realizar la mutación de la proteína wt a proteína mutada, se decide optar por mutar el residuo más contribuyente de entre todos los aminoácidos que se encuentran en la Figura 6A, por lo que se producirá una transformación de fenilalanina a alanina, esto es, la mutación PHE244ALA.

En general, la transformación de fenilalanina en alanina produce una disminución de la contribución de todas las energías de todos los residuos, resultando en una inestabilidad en la proteína de alrededor de 3 kcal/mol, como se ha mencionado anteriormente. Además, se observa que casi todos los residuos más cercanos al cromóforo coinciden tanto en el análisis de la proteína wt como mutada, a excepción del residuo de SER311 el cual aparece en la Figura 6A pero no en la 6B, al contrario que PHE247 que pasa a ser uno de los residuos más cercanos al cromóforo al mutar la proteína.

### 7.3.2 Análisis de los clústeres

El objetivo de computar el valor de  $\Delta G_T$  mediante análisis de clúster, separando las geometrías de la trayectoria en bloques o asociaciones bajo el criterio de similitud estructural, en lugar de tratarlas todas de forma conjunta separándolas en intervalos equidistantes sin ningún criterio como en el apartado anterior, es averiguar si ambos métodos de tratamiento de datos proporcionan valores estadísticamente similares. En caso de ser así, la realización del segundo protocolo es preferible porque el tiempo computacional es menor. Asimismo, el método de agrupamiento de clústeres permite observar las geometrías obtenidas durante la dinámica de una manera mucho más visual al agruparlas en bloques de características similares. Este se basa en la idea de ordenar un conjunto de objetos formando asociaciones o grupos de manera que sean lo más homogéneos dentro de sí mismos y heterogéneos entre sí. Para ello, se seleccionan todas las geometrías correspondientes de los 1500 pasos seleccionados y se dividen en 10 clústeres según las similitudes estructurales que presenten. Como ejemplo visual, se representan en las Figura 7 la superposición de las geometrías correspondientes al primer clúster y la estructura representativa de este grupo, respectivamente.

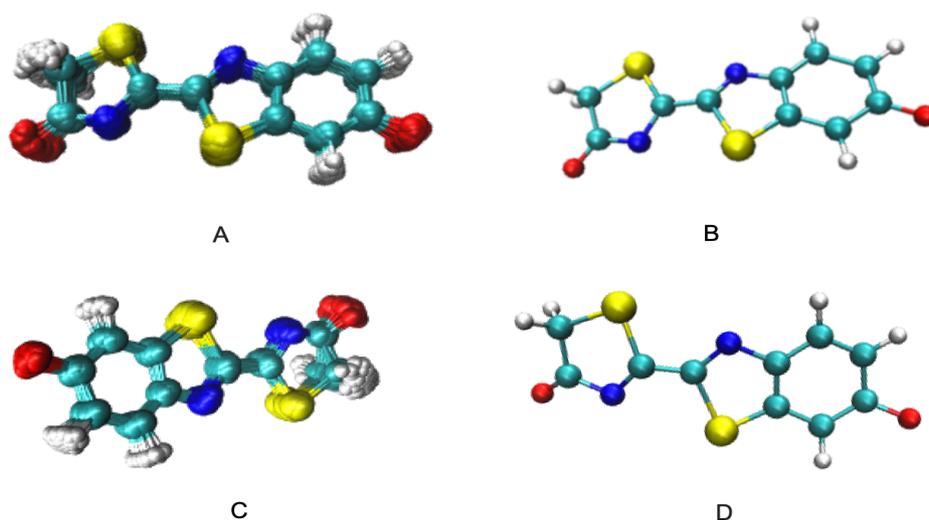


Figura 7. Representación gráfica de la superposición de todas las geometrías de oxiluciferina que conforman el primer clúster cuando la proteína es (A) natural y (C) mutada. Geometría representativa del primer clúster cuando la proteína es (B) natural y (D) mutada.

Por un lado, comparando las Figuras 7A y C, donde se representan el primer clúster de la secuencia de 10 clústeres estudiada, se observa que en ambos casos, las geometrías no parecen variar mucho y se observa que sendos clústeres son visualmente homogéneos y los átomos presentan poco movimiento. Por otro lado, al comparar las estructuras representativas del primer clúster de cada sistema (Figuras 7B y 7D), se observa que ambas son muy similares entre sí, aunque cabe mencionar que en el caso de la proteína PHE244ALA (Figura 7D) se produce un elongamiento del enlace carbono-azufre y un acortamiento del enlace carbono-nitrógeno en el anillo de tiazol con respecto a estas mismas distancias interatómicas en el caso de la proteína wt.

La obtención de los clústeres y su clasificación se realizan con CPPTRAJ<sup>[41]</sup> según el número de pasos que compone cada grupo y finalmente, para cada estructura representativa de estos 10 grupos, se realiza un cálculo de energía libre de interacción siguiendo el método MMGBSA.<sup>[40]</sup> El valor de la energía de interacción se computa como el promedio de las energías de cada geometría representativa ponderada por el número de geometrías del clúster al que representa dicha geometría mediante la Ecuación [20] donde  $\Delta G_T^i$  es la energía libre de interacción de cada clúster y Freq la frecuencia de aparición o peso de cada clúster  $i$  (número de geometrías del clúster dividido entre el número de geometrías totales de la dinámica). Idealmente, la energía libre de interacción así calculada debe ser igual o similar al valor calculado teniendo en cuenta cientos de geometrías equiespaciadas.

$$\Delta G_T = \sum_{i=0}^q \Delta G_T^i * Freq^i \quad [20]$$

En la Tabla 1 se recogen los datos del cálculo de la energía libre de interacción total con 10 clústeres calculados para ambos sistemas.

Tabla 1. Cálculo de la energía libre de interacción total mediante análisis de clúster .

Cluster	$\Delta G_T^i$ proteína wt (kcal/mol)	Freq	$\Delta G_T^i$ proteína PHE244ALA(kcal/mol)	Freq
1	-35,6	0,24	-31,3	0,19
2	-34,1	0,17	-35,1	0,18
3	-35,9	0,15	-32,8	0,16
4	-36,3	0,12	-35,9	0,15
5	-38,3	0,10	-31,7	0,11
6	-36,2	0,09	-33,5	0,10
7	-37,7	0,07	-30,7	0,06
8	-35,7	0,04	-34,2	0,03
9	-38,5	0,02	-34,1	0,01
10	-39,2	0,01	-37,0	0,01
	$\Delta G_T$ (kcal/mol)	-36,0	$\Delta G_T$ (kcal/mol)	-33,3

Dado que se trata de un promedio ponderado y algunos datos presentan más peso que otros, se calcula la desviación estándar ponderada siguiendo la Ecuación [21] para expresar el resultado correctamente. Finalmente, se obtiene un resultado de  $-36 \pm 1$  kcal/mol para el sistema con la proteína wt y un resultado de  $-33 \pm 2$  kcal/mol para el sistema con la proteína mutada, los cuales son estadísticamente igual al resultado del cálculo con 1500 geometrías ( $-36 \pm 2$  kcal/mol para el caso de la proteína wt y  $-33 \pm 3$  kcal/mol para la proteína PHE244ALA), aunque este método presenta una menor desviación estándar. Por ende, se comprueba que el método de análisis de la energía libre de interacción mediante clústeres es la opción más beneficiosa desde el punto de vista de coste computacional, ya que se obtienen resultados estadísticamente

iguales al calcular la energía libre de interacción con un número mínimo de 750 geometrías para el caso de la proteína wt y 375 para la proteína mutada que con 10 geometrías representativas. Esta reducción del número de geometrías que se introducen en el cálculo hace que el tiempo de cálculo se reduzca de horas a minutos.

$$SD_w = \sqrt{\frac{\sum_{i=1}^n Freq_i (\Delta G_i - \bar{\Delta G}_{Freq})^2}{\frac{M-1}{M} \sum_{i=1}^n Freq_i}} \quad [21]$$

Asimismo, y en concordancia con la frecuencia representada en la Tabla 1, se puede representar y analizar gráficamente la población de cada uno de estos grupos a lo largo de la trayectoria (Figuras 8A y C), al igual que la distancia entre las clústeres durante la trayectoria y la oxiluciferina inicial (Figuras 8B y D).

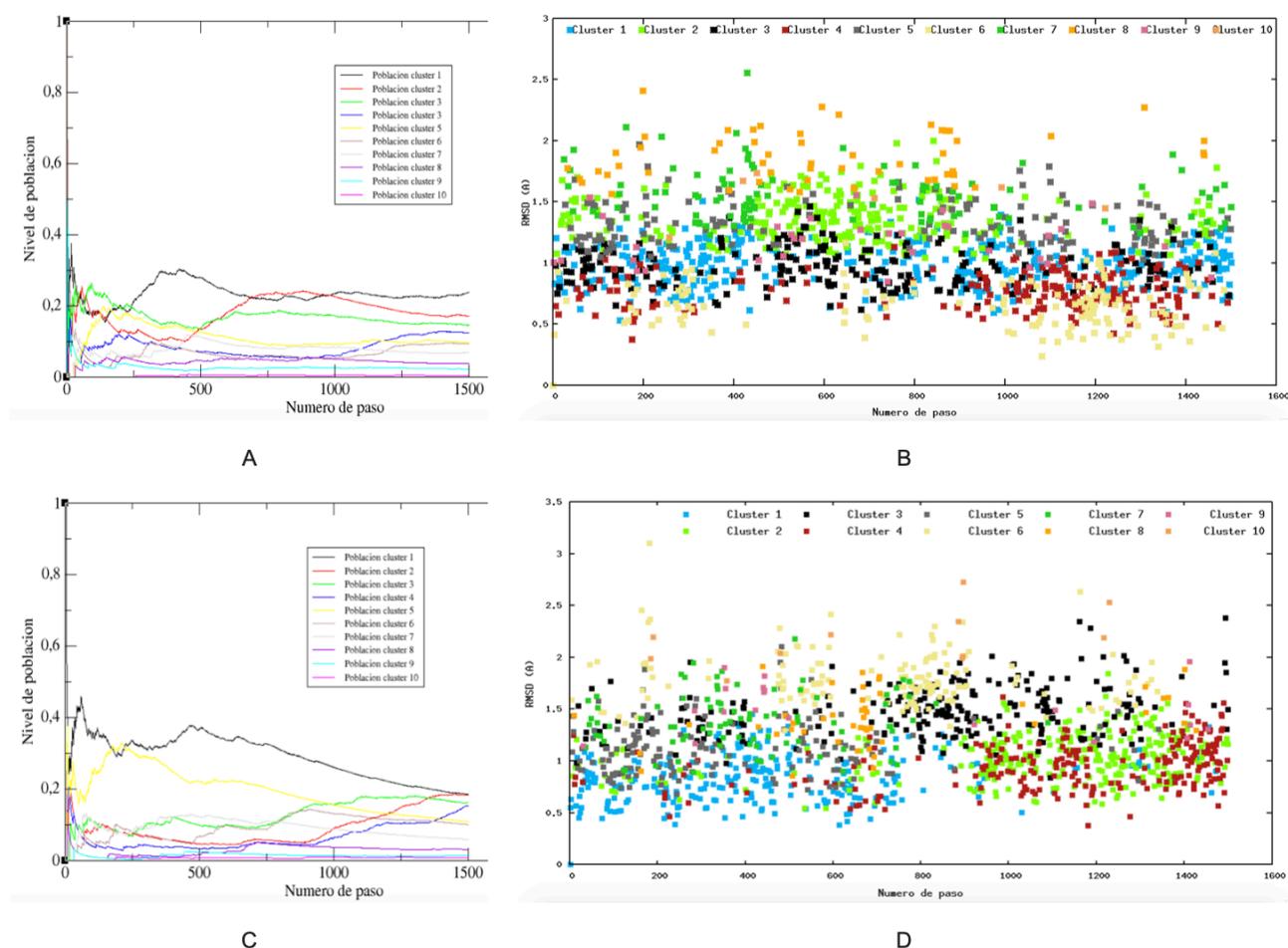


Figura 8. Representación gráfica de la población de cada clúster con respecto al número de paso para los sistemas con (A) proteína wt y (C) proteína mutada junto con la RMSD de la oxiluciferina con respecto a cada clúster para los sistemas con (B) proteína wt y (D) proteína mutada. Cada paso corresponde a 50 ps, por lo que se representan los últimos 75000 ps (o 1500 pasos) de la trayectoria.

En las Figura 8A y C, se observa que el primer clúster (línea negra) es el más poblado de todos en ambos sistemas, y por ende, el que mayor frecuencia presenta. Analizando la Figuras

8B y D, se puede realizar una asignación de las geometrías de cada paso a los clústeres representativos por ejemplo, se observa que el primer clúster tiene una mayor población entre los pasos 400 y 1000 en el caso del sistema con la proteína wt (Figura 8B), mientras que en el caso del sistema con la proteína mutada (Figura 8D), está mayormente compuesto por las primeras 750 geometrías de la trayectoria. Realizando un análisis similar para el resto de grupos, se observa que la población de cada clúster disminuye progresivamente hasta llegar al clúster 10 donde su población es prácticamente nula.

#### 7.4 Espectros de emisión

Además de estudiar la energía libre de interacción total y su descomposición teniendo en cuenta los residuos más cercanos, se calculará el espectro de emisión UV-Vis del complejo oxiluciferina/luciferasa para comprender su bioluminiscencia y en última instancia conocer su aplicabilidad biomédica en el campo de la bioimagen, aunque este no es el objetivo del trabajo, es un primer paso para a largo plazo, con la realización de más cálculos y experimentos, conseguir proponer mejores moléculas y proteínas para las aplicaciones biomédicas. Para ello, se realiza el cálculo de los espectros de emisión mediante cálculos QM/MM para relacionar la energía de los estados excitados con las coordenadas iniciales de los núcleos y que el coste del cálculo sea computacionalmente asequible.

Según se refleja en la Figura 9, se obtiene un pico de emisión a 2,36 eV para la proteína wt y un pico a 2,28 eV para la proteína mutada. Ambos picos de emisión corresponden a la zona del visible (1,5-3,0 eV) y emiten luz verde. Sin embargo, estos picos de emisión no caen dentro de la ventana de transparencia biomédica (zona del rojo e IR cercano), por lo que más adelante en futuros trabajos se estudiará si la mutación realizada producirá una modulación de la luz de emisión tal que esta acabe cayendo en este rango.

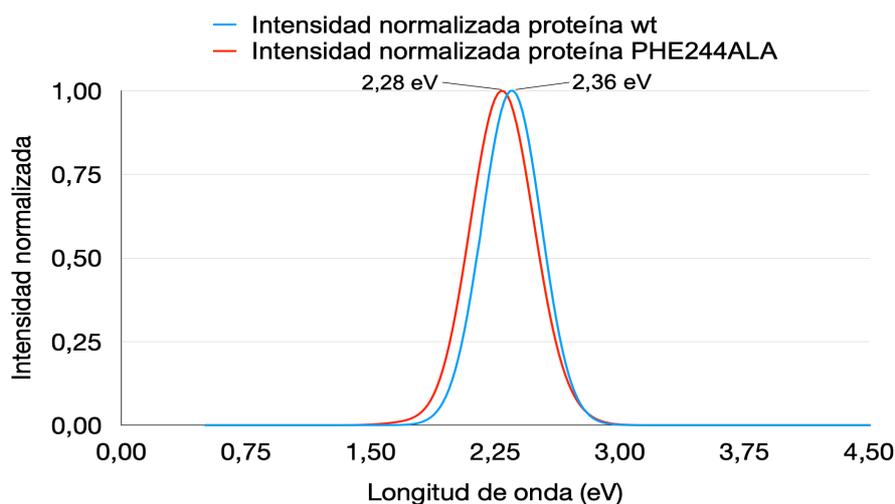


Figura 9. Espectros de emisión de la proteína wt (azul) y mutada (rojo).

Comparando los picos de emisión de ambas proteínas y teniendo en cuenta que para realizar los cálculos QM/MM se introducen las cargas puntuales en el hamiltoniano. Se observa, que aunque de manera general se haya producido una disminución de las contribuciones de los aminoácidos al realizar la mutación como se refleja en la Figura 6. Estudiando detenidamente las

---

interacciones electrostáticas de PHE244 y del resultante de su mutación, ALA244, representadas en las Figuras 6A y 6B, respectivamente, se observa un pequeño aumento de la contribución. Mientras que la contribución a las interacciones electrostáticas de la PHE244 es de -0,69 kcal/mol, la de ALA244 es de -0,81 kcal/mol. Por lo que la mutación de fenilalanina por alanina resulta en una mayor contribución en cuanto a interacciones entre cargas eléctricas, que finalmente se traduce en un pico de emisión desplazado hacia el rojo, menos energético, con respecto a la proteína sin mutar. Es decir, la mutación PHE244ALA acaba produciendo un desplazamiento batocrómico hacia longitudes de onda más largas.

Asimismo, se comprueba que el pico de emisión de la proteína wt es coherente al compararlo con el valor experimental obtenido de la bibliografía procedente de la base de datos de proteínas (PDB).<sup>[39]</sup> Experimentalmente, se obtiene un pico de emisión a 2,21 eV de la luciferasa natural mientras que computacionalmente aparece a 2,36 eV, lo que es una muy buena aproximación considerando que la diferencia está dentro del error del método.

## 8. Conclusiones

En este trabajo se ha llevado a cabo el estudio mediante dinámica molecular clásica y cálculos QM/MM del sistema bioluminiscente oxiluciferina/luciferasa con dos diferentes proteínas, una natural y otra mutada. Primero, se ha realizado un análisis de la energía libre de interacción entre el cromóforo y la proteína siguiendo el método MMGBSA y utilizando dos protocolos distintos, seleccionando cientos de geometrías equiespaciadas de la trayectoria y seleccionando solamente diez geometrías después de un análisis de clúster. Asimismo, se ha analizado la descomposición por residuos de la energía libre de interacción de la oxiluciferina con ambas proteínas. Finalmente, se han simulado los espectros de emisión de ambas proteínas recurriendo a cálculos QM/MM usando la TD-DFT con el funcional CAM-B3LYP y la base 6-311G(2d,p). A partir de estos cálculos se han alcanzado las siguientes conclusiones:

1. Tras realizar las simulaciones de dinámica molecular para la proteína wt y estudiar la diferencia entre las posiciones de los átomos durante la trayectoria y en la geometría inicial con RMSD, se observa que la estructura converge en el tramo de 1500 pasos comprendido entre 2500 y 4000 pasos, el cual corresponde a los últimos 75 ns de la trayectoria de 200 ns. Por lo que se deduce que simulación es suficientemente larga para poder ser analizada.
2. Tras calcular la energía libre de interacción total entre la oxiluciferina y la proteínas wt seleccionando 1500 geometrías de la trayectoria y mediante análisis de clúster teniendo en cuenta solo 10 geometrías representativas, se observa que se obtienen resultados estadísticamente iguales, por lo que se concluye que el análisis mediante clúster es un protocolo de estudio alternativo muy adecuado puesto que reduce en gran medida el coste computacional sin afectar a la calidad de los resultados.
3. Una vez obtenida la descomposición por residuo de la energía libre de interacción entre el cromóforo y la proteína wt, se determina que el residuo PHE244 es el que presenta mayor contribución de entre todos los residuos más cercanos al cromóforo. Esto es especialmente

---

cierto atendiendo a las interacciones de van der Waals entre la proteína y la oxiluciferina y en las interacciones de solvatación apolar con el disolvente. Por tanto, el residuo PHE244 se elige para realizar la mutación (mutación PHE244ALA), y llevar a cabo simulaciones de dinámica molecular.

4. El análisis de la convergencia de la estructura de la proteína mutada es similar al de la proteína wt. Además, a partir de la comparación entre los dos protocolos para el cálculo de la energía libre de interacción, también se concluye que el uso de solo 10 geometrías representativas es suficiente para obtener buenos resultados. Ambos protocolos muestran que la mutación produce un cambio en la energía libre de interacción de en torno a 3 kcal/mol, posiblemente debido a la naturaleza química de la cadena lateral del aminoácido mutado.
5. El cálculos de los espectros de emisión revela que la mutación PHE244ALA produce un ligero desplazamiento hacia longitudes de onda más largas de la emisión de la luciferina. Aun así, el desplazamiento encontrado no es importante y no modula la longitud de onda de emisión lo suficiente como para que se emita luz roja, que es la usada en aplicaciones biomédicas de bioimagen de tejidos.

Este proyecto de investigación es un estudio preliminar de la influencia de la mutación PHE244ALA de la enzima luciferasa en la bioluminiscencia de la oxiluciferina. Sin embargo, hay un gran número de mutaciones que podrían afectar a la fotofísica del sistema y, por tanto, podrían realizarse nuevos estudios experimentales y teóricos modificando más residuos que presenten gran contribución a la energía libre de interacción, como la SER344 o la GLY338. Esto permitiría encontrar mutaciones que mejoren las propiedades bioluminiscentes del sistema que puedan ser utilizadas en nuevas aplicaciones biomédicas.

## 6. Bibliografía

- [1] Vieira, J., et al., (2012). *J. Photochem. Photobiol. B: Biol.*, 117, 33-39.
- [2] Fleiss, A., et al (2019). *Curr. Genet.*, 65(4), 877-882.
- [3] Kaskova, Z. M., et al., (2016). *Chem. Soc. Rev.* 45(21), 6048-6077.
- [4] Kaplan, H. B., et al., (1985). *J. Bacteriol.*, 163(3), 1210-1214.
- [5] Stevani, et al., (2013). *J. Photochem. Photobiol.*, 89(6), 1318-1326.
- [6] Al-Handawi, et al., (2022). *Chem. Revi.*, 122(16), 13207-13234.
- [7] Garcia-Iriepa, C., et al., (2018). *J. Chem. Theory Comput.*, 14(4), 2117-2126.
- [8] Robert, A. G., et al., (2004). *Nat. Rev. Cancer*, 4(11), 891-899.
- [9] Kim, J. W., et al., (2006). *Cancer Res.*, 66(18), 8927-8930.
- [10] Becke, A.D., (1993). *J. Chem. Phys.*, 98, 5648–5652.
- [11] Stephens, P. J. et al.,(1994). *J. Chem. Phys.*, 98, 11623– 11627.
- [12] Lee, C., et al., (1988), *Phys. Rev. B*, 1988, 37, 785–789.
- [13] Yanai, T., et al., (2004). *Chem. Phys. Lett.*, 393(1-3), 51-57.
- [14] Francl, M. M., et al., (1982). *J. Chem. Phys.*, 77, 3654–3665.
- [15] Krishnan, R., et al., (1980). *J. Chem. Phys.*, 72, 650–654.
- [16] McLean, A. D., et al., (1980). *J. Chem. Phys.*, 72, 5639–5648.

- 
- [17] Hehre, W. J., et al., (1969). *J. Chem. Phys.*, 51 (6): 2657–2664
- [18] Hehre, W. J., (1970). *J. Chem. Phys.*, 52 (5): 2769–2773.
- [19] Gaussian 09, Frisch, M. J., et al., Gaussian, Inc., Wallingford CT, 2016
- [20] Hayes, J. (2012). *Molecular dynamics: studies of synthetic and biological macromolecules*, 171-190.
- [21] Rastelli, G. (2010a). *J. Comput. Chem.* 31, 797–810.
- [22] Kollman, P. (1993). *Chem. Rev.* 93, 2395–2417.
- [23] Kollman, P.A., et al., (2000). *Acc. Chem. Res.* 33, 889–897.
- [24] Greenidge, P.A., et al., (2013). *J. Chem. Inf. Model.* 53, 201–209.
- [25] Hou, T., et al., (2011a). *J. Comput. Chem.* 32, 866–877.
- [26] Schneider, T., et al., (1978), *Phys. Rev. B*, 17, 1302.
- [27] Berendsen, H. J. C. et al., (1984), *J. Chem. Phys.*, 81, 3684.
- [28] Wang, J., et al., (2004), *J. Comput. Chem.*, 2004, 25, 1157–1174.
- [29] Singh, U.C., et al., (1984). *J. Comput. Chem.*, 5, 129-145.
- [30] Jorgensen, W. L., et al., (1983), *J. Chem. Phys.*, 1983, 79, 926–35.
- [31] Davidchack, R. L., et al., (2009). *J. Chem. Phys.*, 130(23), 234101
- [32] Sutmann, G. (2002). *Quantum simulations of complex many-body systems: from theory to algorithms*, 10, 211-254.
- [33] Hohenberg, P., et al., (1964), *J. Phys. Rev.*, 136, B864.
- [34] Kohn, W., et al., (1965) *J. Phys. Rev.*, 140, A1133.
- [35] Runge, E., et al., (1984). *Phys. Rev. Lett.*, 52(12), 997.
- [36] Seminario, J. M., (1996), *Int. J. Quantum Chem.*, 60, 1271–1277.
- [37] Tsuneda, T., et al., (2014), *Density Functional Theory in Quantum Chemistry*, Japan, Springer, 102.
- [38] Jensen, F., *Introduction to Computational Chemistry*, Second., 2007.
- [39] Nakatsu, T., et al., (2006). *Nature*, 440(7082), 372-376.
- [40] Miller III, B. R., et al., (2012). *J. Chem. Theory Comput.*, 8(9), 3314-3321.
- [41] Daniel R. Roe, et al., (2013). *J. Chem. Theory Comput.*, 9(7), 3084-3095.
- [42] Tian, C., et al., (2020). *J. Chem. Theory Comput.*, 16, 1, 528-552.