



CÁLCULO DE PRIMEROS PRINCIPIOS DE
PROPIEDADES MECÁNICAS Y ELECTRÓNICAS DE
MOLÉCULAS ORGÁNICAS

Por

Ramsés Valente Salazar Aparicio

COMO REQUISITO PARCIAL PARA
OBTENER EL GRADO DE
MAESTRO EN CIENCIAS (ÓPTICA)

DIVISIÓN DE FOTÓNICA,
CENTRO DE INVESTIGACIONES EN ÓPTICA, A.C.
LEÓN, GUANAJUATO, MÉXICO.

2012

Supervisor: Dr. Raúl Alfonso Vázquez Nava

Sinodales: Dr. Norberto Arzate Plata (CIO)

Dr. Gregorio Hernández Cocoltzi (BUAP)

Contenido

Contenido	v
Lista de Figuras	vii
Resumen	x
Agradecimientos	xi
1. Introducción	1
1.1. Antecedentes	2
2. Teoría	7
2.1. Aproximación de Born-Oppenheimer	8
2.1.1. Análisis de la aproximación	9
2.2. Teoría de perturbaciones	12
2.2.1. Perturbación de sistemas de 2 niveles	12
2.2.2. Sistemas de muchos niveles	15
2.3. Métodos variacionales	17
2.4. Campos autoconsistentes (SCF)	19
2.5. Método autoconsistente de Hartree-Fock	20
2.5.1. Hartree-Fock restringido y no restringido	23
2.6. La selección de conjuntos de funciones de base	24
2.7. Teoría funcional de la densidad	28
2.7.1. Teoría de Thomas-Fermi	29
2.7.2. Derivación funcional	31

2.7.3. La ecuación Thomas-Fermi-Dirac	33
2.8. Simulación de Dinámica molecular	34
2.8.1. Funciones de campo de fuerza	34
3. Cálculo de primeros principios de moléculas orgánicas	37
3.1. Primeros conceptos de simulación molecular	37
3.2. Generación de parámetros	39
3.3. Simulación de la dinámica molecular	40
3.4. Cálculo del RMSD	41
4. Resultados	43
4.1. Proteína fotoactiva amarilla en un medio biológico oscuro e iluminado	44
4.2. Proteína fotoactiva amarilla actualizada	48
4.3. Nanotubo de nitruro de boro con moléculas de hidrógeno	54
5. Conclusiones	63
Referencias	65
A. Archivos PDB (Protein Data Bank)	69
B. Archivo de topología	75
C. Archivo de parámetros	81

Índice de figuras

2.1. Una típica curva de energía potencial de una molécula diatómica.	8
2.2. Enlaces mecánicos de interacciones partícula-partícula.	36
4.1. La proteína fotoactiva amarilla en su estado base (a) y su estado desplegado (b).	44
4.2. La proteína fotoactiva amarilla en su estado base en una esfera de agua antes de la simulación (a) y después de la simulación (b).	45
4.3. La proteína fotoactiva amarilla en su estado excitado en una esfera de agua antes de la simulación (a) y después de la simulación (b).	46
4.4. RMSD de la proteína calculado para el estado excitado (1XFQ) y su estado base (1XFN).	46
4.5. RMSD de los residuos para el estado base (izquierda) y su estado excitado (derecha).	47
4.6. Diferencia del RMSD para cada residuo entre los estados 1XFN y 1XFQ.	47

4.7. El residuo CYS original (izquierda) y el mismo unido al residuo HC4 (derecha). Hidrógenos en blanco, oxígenos en rojo, carbonos en azul claro, nitrógeno en azul marino y azufre en amarillo.	48
4.8. Segmento optimizado de la unión CYS-HC4.	49
4.9. Modos de vibración del segmento optimizado (fig. 4.8).	52
4.10. La proteína fotoactiva amarilla con ácido para-cumárico, en estado base (a) y estado excitado (b).	53
4.11. La proteína fotoactiva amarilla con ácido para-cumárico en su estado base en la esfera de agua, con ácido para-cumárico, antes de la simulación (a) y después de la simulación (b).	53
4.12. La proteína fotoactiva amarilla con ácido para-cumárico en su estado excitado después de la simulación, antes de la simulación (a) y después de la simulación (b).	54
4.13. RMSD de la proteína calculado para el estado excitado (1XFQ) y su estado base (1XFN).	55
4.14. RMSD de los residuos para el estado base (izquierda) y su estado excitado (derecha).	55
4.15. Diferencia del RMSD para cada residuo entre los estados 1XFN y 1XFQ.	56
4.16. Nanotubo de carbono con coordenadas optimizadas. Nitrógenos en azul marino, boros en color cobre e hidrógenos en blanco.	57
4.17. Nanotubo de carbono con una molécula de H_2 cercana.	58
4.18. Energía total del sistema durante cada paso de la optimización del nanotubo con una molécula de H_2	59

4.19. Nanotubo de carbono con dos moléculas de H_2 cercana.	59
4.20. Energía total del sistema durante cada paso de la optimización del nanotubo con dos moléculas de H_2	60
4.21. Nanotubo de carbono con tres moléculas de H_2 cercana.	60
4.22. Energía total del sistema durante cada paso de la optimización del nanotubo con tres moléculas de H_2	61
4.23. Nanotubo de carbono con cuatro moléculas de H_2 cercana.	61
4.24. Energía total del sistema durante cada paso de la optimización del nanotubo con cuatro moléculas de H_2	62

Resumen

La simulación de dinámica molecular es un campo de la investigación de la bioquímica, que se utiliza recientemente para describir y explorar fenómenos bioquímicos que nos rodean de un modo teórico. Este tipo de estudio, si bien no es experimental, tiene la capacidad de reproducir o predecir resultados experimentales en el mismo campo de la bioquímica y otros más, con una gran precisión.

En este trabajo se reportan los resultados obtenidos del modelado molecular de la proteína fotoactiva amarilla (PYP, por sus siglas en inglés) dentro de un medio biológico a temperatura ambiente, la simulación se desarrolló en dos diferentes casos, la diferencia está en la implementación de una estructura de ácido para-cumárico (HC4) a la estructura de cistamina (CYS) dentro de la proteína, la cual representa una mejoría en la respuesta óptica de la proteína [1]. La dinámica molecular fue desarrollada por cálculos mecánicos y electrostáticos por la complejidad de la proteína.

También se reportan los resultados obtenidos de la optimización de la estructura de un nanotubo de *BN* (nitruro de boro) teniendo presente una molécula de H_2 cercana [2]. La optimización se refiere a la relajación de la estructura hasta llegar a la geometría con menor energía, para observar el comportamiento natural de estos y encontrar la tendencia a este tipo de interacciones atómicas. La optimización se desarrolló con la teoría de Hartree-Fock [22] y de Thomas-Fermi-Dirac [25].

Las simulaciones computacionales se desarrollaron bajo el software libre: VMD Y MacMolPlt para visualización molecular, GAMESS para optimizaciones de origen cuántico y NAMD para la dinámica molecular mecánica.

AGRADECIMIENTOS

Agrezco a mis padres por dedicar su vida a enseñarme, educarme y amarme, sin ellos no me encontraría ahora culminando mis estudios de Maestría.

Al Centro de Investigaciones en Óptica y a CONACyT, por darme la oportunidad de continuar con mis estudios y aportes a la ciencia.

Al Dr. Raúl Alfonso Vázquez Nava, asesor de mi tesis, que tuvo la bondad y paciencia de compartir sus conocimientos y de ser guía para cumplir mi meta.

1 Introducción

En este capítulo se habla de los antecedentes del modelado molecular, así como de algunos ejemplos actuales de la aplicación del modelado en la ciencia, para familiarizarse con el objetivo de este trabajo de tesis.

El modelado y simulación de moléculas se ha utilizado, entre otros, en la investigación para describir, explicar y explorar fenómenos, procesos e ideas abstractas. Una gran virtud de un buen modelo es que sugiriendo preguntas más avanzadas, nos lleva más allá del fenómeno del cual empezamos a formular hipótesis que puedan ser examinadas experimentalmente.

El modelado de moléculas provee de imágenes a escala nanométrica a una resolución atómica e incluso electrónica, predice la interacción de materiales orgánicos e inorgánicos, y evalúa las estrategias para rediseñar biopolímeros para usos nanotecnológicos. En efecto, resulta que las simulaciones computacionales son una excelente herramienta para formar imágenes en la nanotecnología, cubriendo el ancho de la escala bastante bien y cubriendo la escala de tiempo hasta muchos nanosegundos para describir niveles atómicos. La veracidad de las simulaciones

por computadora han mejorado constantemente y, en algunos casos, ha alcanzado la exactitud de buenos métodos experimentales.

El uso de modelos moleculares para posibilitar la visualización de ideas complejas, procesos y sistemas en la química ha sido divulgado por un largo tiempo. La elección del tipo de modelo tiene un impacto en la imagen mental que se crea. Un problema que surge usando modelos es que se pone insuficiente énfasis en el hecho de que los modelos son simulaciones de la realidad basadas en la teoría. Aplicado a la química se logran modelos análogos para explicar conceptos nuevos y abstractos.

1.1. Antecedentes

La aplicación de modelos computacionales utilizando funciones analíticas de energía potencial en el campo de la mecánica clásica ha probado ser una herramienta muy útil para estudiar moléculas de interés bioquímico y orgánico. Estas aplicaciones de mecánica molecular han empleado minimización de la energía, dinámica molecular y métodos de Monte Carlo [3] para estudiar las superficies analíticas de la energía potencial. En principio, dichos métodos son capaces de dar una perspicacia de todo el espectro de interacciones no covalentes entre moléculas. Estos métodos se combinan con cálculos de estructura electrónica por mecánica cuántica, el modelado de enlaces covalentes cambia esencialmente todas las reacciones e interacciones moleculares.

El modelado molecular tiene un rol importante en la bionanotecnología, por

ejemplo:

Los nanotubos de carbono (CNT, por sus siglas en inglés) son microtubos huecos descubiertos en 1991 [4]. Con estrechos diámetros de escala nanométrica, los CNTs pueden crecer hasta varias micras, o hasta centímetros. Se hace referencia a un sencillo tubo como un nanotubo de carbono unitario (SWNT). A diferencia de materiales ordinarios, los SWNTs pueden ser metálicos o semiconductores dependiendo de su particular estructura [5].

Después del descubrimiento de los CNTs, los científicos e ingenieros empezaron a indagar en el prometedor potencial de las aplicaciones biológicas de los CNTs, en particular, como dispositivos de administración de medicamentos [6–8] y como biosensores [9, 10]. En el campo de la biomedicina se utilizan nanotubos de carbono, en donde el punto clave en la metodología es que el sistema electrónico de π -electrones en los nanotubos de carbono unipared (SWNT, por sus siglas en inglés) son altamente polarizables y fáciles de influenciar por factores del entorno molecular [11]. Esta influencia, que afecta también las propiedades de la absorción de la luz, convierte los SWNTs en sensores y contribuye a la energía de interacción entre moléculas interiorizadas o absorbidas por los SWNTs también. Un desarrollo sistemático es necesario para unir la física de los SWNTs con la biología molecular celular. Aquí es donde el modelado entra desarrollando un método de simulación suficientemente flexible para ser eficientemente empleado para SWNTs incrustados en entornos biológicos, y también suficientemente exacto para reproducir los efectos descritos por teorías físicas computacionalmente costosas.

Como investigación en nanotecnología se encuentran las herramientas de fabricación de circuitos integrados de dimensiones nanométricas, y tecnologías alternativas para secuenciación de ADN más barata y rápida está por aparecer. Por ejemplo, un dispositivo para secuencias de ADN de alto rendimiento podría ser construido alrededor de un poro de 2nm de diámetro en una membrana sintética delgada (de 2-5nm de grosor) [12–14]. La secuencia de una molécula de ADN puede ser leída por dicho dispositivo, en principio, con un detector semiconductor, integrado con un poro, que graba las señales eléctricas inducidas por el ADN pasando por el poro [15]. En este caso el detalle se encuentra en el grabado eléctrico de una molécula individual con el eventual objetivo de una secuenciación de ADN barata, rápida y precisa. En este caso, el modelado necesita adaptarse a nuevos materiales, Si_3N_4 tanto como SiO_2 . Más adelante, el modelado necesita proveer imágenes de los procesos de translocación y la electrostática de los sistemas poro-permeables que son esenciales para guiar experimentos y desarrollo tecnológico. Las imágenes obtenidas de los modeladores, probadas mediante predicciones como corrientes bloqueadoras o conductividad de los iones verificadas por la observación, son invaluable para los ingenieros.

Los nanodiscos son membranas de bicapas de lípidos de orden nanométrico que son estabilizados por proteínas anfipáticas. Estos proveen de una plataforma para incrustar, solver y estudiar proteínas de única membrana. Las proteínas de membrana pueden ser difícil de estudiar experimentalmente y son a menudo estudiadas bajo condiciones no nativas, como en altas concentraciones de detergentes y en micelas. Los efectos de estos entornos no nativos en la estructura y la función de las proteínas son conocidas. Es por lo tanto deseable poder estudiar las proteínas

de membrana, especialmente proteínas de membrana integral, en un entorno de lípido bicapa más nativo. Esto ha motivado a científicos para elaborar bicapas de lípidos discoidales de orden nanométrico estabilizados con proteínas anfipáticas helicoidales, llamados nanodiscos [16,17], que componen el nanoentorno deseado para las proteínas de membrana. la cuestión en este aspecto está en hacer disponible una membrana de bicapa de lípidos a nanoescala para su estudio o procesos biomédicos relacionados como unión de receptores o extracción de colesterol. Proteínas nativas de las células humanas necesitan ser rediseñadas para el propósito; adicionalmente las estructuras de equilibrio y las propiedades dinámicas de las proteínas y el colesterol en los nanodiscos necesita ser visualizado. El modelado biomolecular está envuelto en ambos aspectos, probando las propiedades de proteínas de Scaffold rediseñadas manteniendo la forma de nanodisco, e ilustrando la más probable estructura de los nanodiscos.

Los ejemplos anteriores son ejemplos de las oportunidades que el modelado biomolecular ofrece a la bionanotecnología moderna, y muestran que se puede emplear para prácticamente cualquier ámbito de la nanotecnología molecular.

En el presente trabajo de tesis se plantea la simulación de la dinámica molecular de algunas estructuras moleculares existentes en el ámbito científico.

En los posteriores capítulos se incluye la teoría precedente al modelado molecular, como conceptos de mecánica y química cuántica, junto con varios métodos de aproximación. Después se describe el procedimiento teórico que se utiliza en el trabajo de tesis, abarcando el planteamiento, la descripción de las herramientas de visualización, de generación de parámetros y la simulación de la dinámica

molecular y la interpretación de los resultados.

Al final se complementa con apéndices con una muestra de los tipos de archivos utilizados durante las simulaciones.

2 Teoría

En este capítulo presento los conceptos teóricos precedentes a la aproximación mecánica de la interacción interatómica molecular. Desde las aproximaciones cuánticas hasta los parámetros de la simulación de dinámica molecular.

En la sección 2.1 hablo de la aproximación de Born-Oppenheimer porque explica la interacción molecular desde la molécula más simple y después hablo de la teoría de perturbaciones en la sección 2.2.1, que trata de como resolver sistemas de muchos cuerpos como conjuntos de sistemas más pequeños. En la sección 2.3 hablo de métodos variacionales y el método auto-consistente de Hartree-Fock, que es el método numérico iterativo utilizado en este trabajo para resolver sistemas de muchos cuerpos. En la sección 2.6 hablo de los conjuntos de base, estos son las funciones que describen las orbitales de los átomos. Hay varias maneras de escribir analíticamente los orbitales, en esta sección se explican diferentes formas. Más adelante, en la sección 2.7 hablo de la teoría funcional de la densidad, esta es importante porque describe cómo se calculan las energías de las configuraciones atómicas en el método numérico, principalmente cuando existen conjuntos de

átomos unidos por enlaces físicos en lugar de químicos, llevando al cálculo de densidades electrónicas. Finalmente en la última sección se describe el cálculo de las fuerzas interatómicas en el modelo mecánico, para los casos en que el sistema de muchos cuerpos es muy grande para cálculos cuánticos.

2.1. Aproximación de Born-Oppenheimer

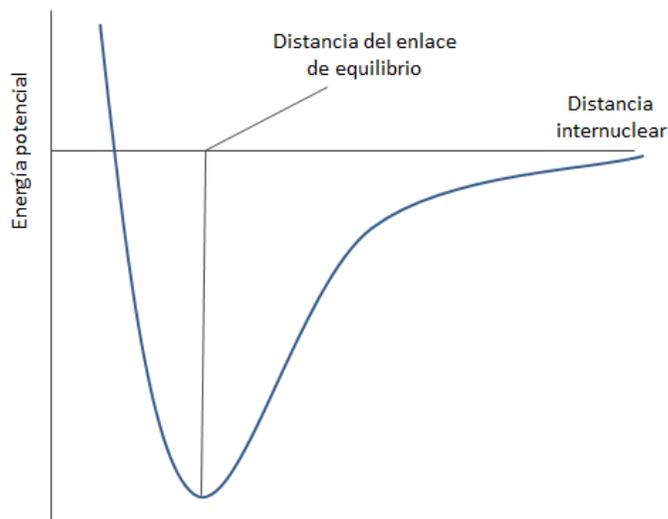


Figura 2.1: Una típica curva de energía potencial de una molécula diatómica.

Desafortunadamente, hasta la molécula más simple, H_2^+ , consiste de tres partículas, y la ecuación de Schrödinger no puede resolverse analíticamente; con el resto de las moléculas por lo tanto se complica todavía más. Para sobrepasar esta dificultad, adoptamos la aproximación de Born-Oppenheimer [23], que hace notar la

gran diferencia entre masas de los electrones y del núcleo. A causa de esta diferencia, los electrones pueden responder casi instantáneamente al desplazamiento del núcleo. Entonces, en lugar de tratar de resolver la ecuación de Schrödinger para todas las partículas simultáneamente, es posible suponer que el núcleo se encuentra fijo en su posición y resolver la ecuación de Schrödinger para los electrones en el potencial electrostático proveniente del núcleo en ese particular arreglo. Diferentes arreglos del núcleo deben entonces ser utilizados para cada electrón. El grupo de soluciones obtenido nos permite construir la curva de energía potencial molecular de una molécula diatómica (fig. 2.1) y en general, una superficie de energía potencial de moléculas poliatómicas. La aproximación de Born-Oppenheimer es muy confiable para estados electrónicos base, pero menos para estados excitados, porque estos presentan otras interacciones electrónicas más complejas, como.

2.1.1. Análisis de la aproximación

La aproximación involucrada en el procedimiento de Born-Oppenheimer puede apreciarse cuantitativamente. La aproximación más sencilla es considerar un análogo en una dimensión de la molécula de hidrógeno, confinando el movimiento en una sola dimensión, el eje x en este caso. El hamiltoniano H para este problema es

$$H = -\frac{\hbar^2}{2m_e} \frac{d^2}{dx^2} - \sum_j \frac{\hbar^2}{dX_j^2} + V(x, X_1, X_2), \quad (2.1)$$

donde x es la posición del electrón, X_j , con $j = 1, 2$, son las posiciones de los dos núcleos y V es el potencial eléctrico del electrón con respecto a los núcleos. La

ecuación de Schrödinger es

$$H\Upsilon(x, X_1, X_2) = \xi\Upsilon(x, X_1, X_2). \quad (2.2)$$

Se quiere obtener una solución de la forma

$$\Upsilon(x, X_1, X_2) = \Psi(x; X_1, X_2)\chi(X_1, X_2), \quad (2.3)$$

donde la notación $\Psi(x; X_1, X_2)$ significa que la función de onda para el electrón depende paraméricamente de los dos núcleos de manera que obtenemos una diferente función de onda $\Psi(x)$ para cada arreglo del núcleo. Cuando esta solución es sustituida en la ecuación (2.2) obtenemos

$$H\Psi\chi = -\frac{\hbar^2}{2m_e} \frac{\delta^2\Psi}{\delta x^2} \chi - \sum_j \frac{\hbar^2}{2m_j} \frac{\delta^2\chi}{\delta X_j^2} \Psi - \sum_j \frac{\hbar^2}{2m_j} \left(2 \frac{\delta\Psi}{\delta X_j} \frac{\delta\chi}{\delta X_j} + \frac{\delta^2\Psi}{\delta X_j^2} \chi \right) + V(x, X_1, X_2)\Psi\chi, \quad (2.4)$$

que puede escribirse como

$$H\Psi\chi = \xi'\Psi\chi - \sum_j \frac{\hbar^2}{2m_j} \left(2 \frac{\delta\Psi}{\delta X_j} \frac{\delta\chi}{\delta X_j} + \frac{\delta^2\Psi}{\delta X_j^2} \chi \right), \quad (2.5)$$

donde ξ' es el eigenvalor de la ecuación

$$-\sum_j \frac{\hbar^2}{2m_j} \frac{\delta^2\chi}{\delta X_j^2} + E(X_1, X_2)\chi = \xi'\chi, \quad (2.6)$$

y E es el eigenvalor obtenido de resolver

$$-\frac{\hbar^2}{2m_e} \frac{\delta^2\Psi}{\delta x^2} + V(x, X_1, X_2)\Psi = E(X_1, X_2)\Psi. \quad (2.7)$$

La ecuación de Schrödinger en la ecuación (2.5) casi tiene la forma que se requiere, una ecuación de eigenvalores, y tendría la forma correcta si el último término

cumple que

$$-\sum_j \frac{\hbar^2}{2m_j} \left(2 \frac{\delta \Psi}{\delta X_j} \frac{\delta \chi}{\delta X_j} + \frac{\delta^2 \Psi}{\delta X_j^2} \chi \right) = 0. \quad (2.8)$$

Si no es igual a cero, pero es muy pequeño comparado con $\xi' \Psi \chi$, tomando en cuenta las masa nucleares en el denominador, entonces la idea de la aproximación de Born-Oppenheimer es igualarlo a cero, en cuyo caso las soluciones a las ecuaciones (2.7) y (2.8) constituirían una solución al problema.

La ecuación (2.7) es la ecuación de Schrödinger para el electrón en un potencial $V(x, X_1, X_2)$ que depende de las posiciones de los dos núcleos. La solución es la función de onda Ψ , y el eigenvalor $E(X_1, X_2)$ es la contribución electrónica a la energía total de la molécula mas la energía potencial de la repulsión nuclear. La ecuación (2.6) es la ecuación de Schrödinger para la función de onda χ del núcleo cuando la energía potencial tiene la forma de la curva de la energía potencial molecular. Su eigenvalor es la energía total de la molécula para la aproximación de Born-Oppenheimer.

Por simplicidad, la ecuación (2.7) se escribe generalmente como

$$H\Psi = E\Psi, \quad \text{con } H = -\frac{\hbar^2}{2m_e} \nabla^2 + V, \quad (2.9)$$

donde V contiene las contribuciones de la energía potencial del electrón en el campo del núcleo estacionario mas la contribución de la interacción nuclear y E es la energía electrónica y de repulsión núcleo-núcleo totales para un arreglo nuclear estacionario.

2.2. Teoría de perturbaciones

Hay muy pocos problemas en los cuales la ecuación de Schrödinger puede ser resuelta de manera exacta, tan pronto como la forma del potencial involucrado se complica, como en los casos en que dos o más partículas interactúan entre sí, la ecuación ya no se puede resolver de manera exacta.

Existen maneras de llevar este proceso. La primera forma es tratar de indagar la forma de la función de onda del sistema; es útil ser guiado a la forma de la función de onda por un conocimiento previo de la distorsión del sistema inducido por los aspectos que intervienen en el potencial de las interacciones. Este procedimiento precede a la teoría de perturbación [23]. La teoría de perturbación es útil cuando estamos interesados en la respuesta de los átomos y moléculas a los campos eléctricos y magnéticos. Cuando estos campos cambian en función del tiempo tenemos que lidiar entonces con una teoría de perturbaciones dependiente del tiempo. Otra manera más de resolver estos casos es con procedimientos de campos autoconsistentes, los cuales son métodos iterativos para resolver la ecuación de Schrödinger para sistemas de muchas partículas.

2.2.1. Perturbación de sistemas de 2 niveles

Para este caso se utiliza un hamiltoniano dado por $H = H^{(0)} + H^{(1)}$, donde $H^{(1)}$ es una perturbación independiente del tiempo en primer lugar. Consideramos entonces un sistema que tiene sólo 2 eigenestados conocidos, dados por $|1\rangle$ y $|2\rangle$, las correspondientes funciones de onda son $\varphi_1^{(0)}$ y $\varphi_2^{(0)}$, respectivamente. Estas

funciones y estados forman una base ortonormal completa y corresponden a las energías E_1 y E_2

$$H^{(0)}\Psi_m^{(0)} = E_m\Psi_m^{(0)}, \quad m = 1, 2. \quad (2.10)$$

Entonces esperamos resolver la ecuación con la función de onda

$$\Psi = a_1\Psi_1^{(0)} + a_2\Psi_2^{(0)}, \quad (2.11)$$

donde a_1 y a_2 son constantes. Para encontrar las constantes a_m insertamos la combinación lineal en la ecuación de Schrödinger y obtener así:

$$a_1(H - E)|1\rangle + a_2(H - E)|2\rangle = 0 \quad (2.12)$$

Cuando esta ecuación es multiplicada por la izquierda por los bras $\langle 1|$ y $\langle 2|$ uno a la vez se obtiene el par de ecuaciones

$$\begin{aligned} a_1(H_{11} - E) + a_2H_{12} &= 0, \\ a_1H_{21} + a_2(H_{22} - E) &= 0, \end{aligned} \quad (2.13)$$

donde $H_{mn} = \langle m|H|n\rangle$. La condición para que existan soluciones no triviales de este par de ecuaciones es que el determinante de los coeficientes de las constantes a_1 y a_2 desaparezca. Esta condición se satisface con los siguientes valores de E :

$$E_{\pm} = \frac{1}{2}(H_{11} + H_{22}) \pm \frac{1}{2}\{(H_{11} - H_{22})^2 + 4H_{12}H_{21}\}^{1/2}. \quad (2.14)$$

En el caso especial, pero común en que la perturbación para el cual los elementos de la matriz diagonal sean cero ($H_{mm}^{(1)} = 0$), esta expresión se simplifica como

$$E_{\pm} = \frac{1}{2}(E_1 + E_2) \pm \frac{1}{2}\{(E_1 - E_2)^2 + 4\varepsilon^2\}^{1/2}, \quad (2.15)$$

donde $\varepsilon^2 = H_{12}^{(1)} H_{21}^{(1)}$. Como $H^{(1)}$ es hermitiano, se puede escribir $\varepsilon^2 = |H_{12}^{(1)}|^2$. Cuando la perturbación no está presente, $\varepsilon = 0$ y $E_+ = E_1$, $E_- = E_2$, y en cambio cuando $E_1 = E_2$, entonces $E_+ + E_- = 2\varepsilon$.

La ecuación (2.15) muestra que mientras más fuerte es la perturbación, más fuerte es la repulsión de los niveles. En resumen:

1. Cuando una perturbación es aplicada, los niveles inferiores disminuyen su energía y los niveles superiores aumentan.
2. Mientras más cercanos estén los estados no perturbados en energía, más notorio es el efecto de la perturbación.
3. Para perturbaciones más fuertes, más fuerte es el efecto en las energías de los niveles.

El efecto de la perturbación puede verse con más detalle considerando el caso de una perturbación que es pequeña comparada con la separación de los niveles de energía de modo que $\varepsilon^2 \ll (E_1 - E_2)^2$. Con esta condición la ecuación (2.15) puede ser expandida para obtener

$$E_{\pm} = \frac{1}{2}(E_1 + E_2) \pm \frac{1}{2}(E_1 - E_2)\left(1 + \frac{2\varepsilon^2}{\Delta E^2} + \dots\right), \quad (2.16)$$

donde $\Delta = E_2 - E_1$. Entonces para el segundo orden de ε se tiene

$$\begin{aligned} E_+ &\approx E_1 - \frac{\varepsilon^2}{\Delta E}, \\ E_- &\approx E_2 + \frac{\varepsilon^2}{\Delta E}. \end{aligned} \quad (2.17)$$

Estas dos soluciones convergen a una solución exacta cuando $(2\varepsilon/\Delta E)^2 \ll 1$.

Las funciones de onda se obtienen resolviendo la ecuación (2.13) para los coeficientes en turno $E = E_+$ (para obtener Ψ_+) y $E = E_-$ (para obtener Ψ_-). Una manera conveniente de escribir las soluciones es

$$\begin{aligned}\Psi_+ &= \Psi_1^{(0)} \cos \zeta + \Psi_2^{(0)} \sin \zeta, \\ \Psi_- &= \Psi_1^{(0)} \sin \zeta + \Psi_2^{(0)} \cos \zeta,\end{aligned}\tag{2.18}$$

con ζ un ángulo arbitrario, y entonces se encuentra que

$$\tan 2\zeta = \frac{2|H_{12}^{(1)}|}{E_1 - E_2}.\tag{2.19}$$

Para un sistema degenerado ($E_1 = E_2$), tenemos $\tan 2\zeta = \infty$, que corresponde a $\zeta = \pi/4$. En este caso las funciones de onda quedan como

$$\begin{aligned}\Psi_+ &\approx \frac{1}{\sqrt{2}}(\Psi_1^{(0)} + \Psi_2^{(0)}), \\ \Psi_- &\approx -\frac{1}{\sqrt{2}}(\Psi_1^{(0)} - \Psi_2^{(0)}).\end{aligned}\tag{2.20}$$

Después se nota con esto que en cada modelo hay influencia del otro modelo en un grado pequeño.

2.2.2. Sistemas de muchos niveles

Suponemos ahora que sabemos todas las eigenfunciones y eigenvalores de un sistema con hamiltoniano $H^{(0)}$ que difiere del sistema verdadero en un pequeño grado para facilitar el cálculo. Un ejemplo sería un oscilador no armónico o una

molécula en un campo eléctrico pequeño, el sistema sería entonces un oscilador armónico o una molécula en ausencia de un campo eléctrico, respectivamente. Entonces suponemos que hemos encontrado la solución de las ecuaciones

$$H^{(0)}|n\rangle = E_n|n\rangle, \quad (2.21)$$

con $n = 0, 1, 2, \dots$ y $|n\rangle$ un miembro de una base ortonormal. Suponemos entonces que estamos calculando la forma perturbada del estado $|0\rangle$ de la energía E_0 , pero este estado no es necesariamente el estado base del sistema.

El hamiltoniano del sistema perturbado se escribe como

$$H = H^{(0)} + \lambda H^{(1)} + \lambda^2 H^{(2)} + \dots \quad (2.22)$$

La única finalidad del parámetro λ es seguir la pista del orden de la perturbación, y nos ayudará a identificar todos los términos de primer orden en la energía, todos los de segundo orden, y así sucesivamente. Al final del cálculo se sustituirá $\lambda = 1$ pues ya habrá cumplido su propósito. De igual manera, las funciones de onda perturbadas del sistema se escriben como

$$\Psi = \Psi^{(0)} + \lambda \Psi^{(1)} + \lambda^2 \Psi^{(2)} + \dots, \quad (2.23)$$

lo cual muestra cuanto es corregida la función $\Psi^{(0)}$ por los términos de varios órdenes en la perturbación. La energía del estado perturbado también tiene términos de corrección de varios órdenes

$$E = E^{(0)} + \lambda E^{(1)} + \lambda^2 E^{(2)} + \dots \quad (2.24)$$

Nos referimos a $E_0^{(1)}$ como la corrección de primer orden de la energía, a $E_0^{(2)}$ como

la de segundo orden, y así sucesivamente. La ecuación que necesitamos resolver es

$$H\Psi = E\Psi. \quad (2.25)$$

Se insertan las ecuaciones precedentes en esta última, seguido de juntar términos que comparten la misma potencia de λ , entonces resulta

$$\begin{aligned} & \lambda^0 \{H^{(0)}\Psi_0 - E_0\Psi_0\} + \lambda^1 \{H^{(0)}\Psi_0^{(1)} + H^{(1)}\Psi_0 - E_0\Psi_0^{(1)} - E_0^{(1)}\Psi_0\} \\ & + \lambda^2 \{H^{(0)}\Psi_0^{(2)} + H^{(1)}\Psi_0^{(1)} + H^{(2)}\Psi_0 - E_0\Psi_0^{(2)} - E_0^{(1)}\Psi_0^{(1)} - E_0^{(2)}\Psi_0\} \\ & + \dots = 0. \end{aligned} \quad (2.26)$$

Como λ es un parámetro arbitrario, el coeficiente de cada potencia de λ debe ser igual a cero por separado, y entonces encontramos el siguiente sistema de ecuaciones

$$\begin{aligned} H^{(0)}\Psi_0 &= E_0\Psi_0 \\ \{H^{(0)} - E_0\}\Psi_0^{(1)} &= \{E_0^{(1)} - H^{(1)}\}\Psi_0 \\ \{H^{(0)} - E_0\}\Psi_0^{(2)} &= \{E_0^{(2)} - H^{(2)}\}\Psi_0 + \{E_0^{(1)} - H^{(1)}\}\Psi_0^{(1)}, \end{aligned} \quad (2.27)$$

y así sucesivamente.

2.3. Métodos variacionales

Sea \hat{H} el hamiltoniano del sistema a estudiar y φ_k su eigenfunción que corresponde al eigenvalor E_k ,

$$\hat{H}\varphi_k = E_k\varphi_k \quad (2.28)$$

El índice k se escoge de tal manera que la energía crece a la vez ($E_k > E_{k'}$ para $k > k'$); el estado base se denota con $k = 0$ y le corresponde la energía E_0 . Sea ahora Φ una función arbitraria normalizada a la unidad; expresándola en términos de la base φ_k obtenemos

$$\Phi = \sum_k a_k \varphi_k, \quad (2.29)$$

en donde, como Φ está normalizado, se cumple que

$$\sum_k |a_k|^2 = 1. \quad (2.30)$$

Del valor esperado de \hat{H} calculado con la función arbitraria Φ se obtiene, sustituyendo la ecuación (2.29), que

$$\begin{aligned} \langle \Phi | \hat{H} | \Phi \rangle &= \sum_{m,n} a_m^* a_n \int \varphi_m^* \hat{H} \varphi_n dx = \sum_{m,n} a_m^* a_n E_n \delta_{mn} \\ &= \sum_n |a_n|^2 E_n \geq E_0 \sum_n |a_n|^2, \end{aligned} \quad (2.31)$$

donde se tomó en cuenta que $E_n \geq E_0$; entonces, utilizando la expresión (2.30), se obtiene que

$$\langle \Phi | \hat{H} | \Phi \rangle \geq E_0. \quad (2.32)$$

Esta última expresión quiere decir que la eigenfunción que corresponde al estado base es la función (normalizada) que minimiza el valor esperado del hamiltoniano. La calidad del resultado del método variacional depende de la función de prueba Φ , así que hay que elegirla cuidadosamente.

2.4. Campos autoconsistentes (SCF)

Los mejores orbitales atómicos para describir las posiciones electrónicas son encontrados por soluciones numéricas de la ecuación de Schrödinger. La suposición detrás de la técnica es que cualquier electrón se mueve en un potencial, el cual es un promedio causado por los demás electrones y el núcleo. Entonces la ecuación de Schrödinger es integrada numéricamente para ese electrón y en ese potencial esférico promedio, tomando en cuenta el hecho de que la carga total dentro de la esfera, definida por el resto de los electrones, varía así como cambia la distancia del electrón al núcleo. Esta aproximación supone que las funciones de onda de los demás electrones ya son conocidas de manera que el potencial esférico promedio se pueda calcular. La ecuación de Schrödinger para el electrón, entonces, está resuelta, y el procedimiento se repite para el resto de los electrones del átomo. Al final del primer conjunto de cálculos tenemos un conjunto de funciones de onda mejoradas para todos los electrones. Estas funciones de onda mejoradas son usadas entonces para calcular el potencial esférico promedio, y el ciclo computacional se repite. El ciclo se repite hasta que el grupo de funciones de onda mejoradas no tenga cambios significativos con respecto a las funciones de onda al principio del ciclo. Las funciones se dice que son autoconsistentes, y son aceptadas como buenas aproximaciones a la verdadera función de onda de muchos electrones [24].

2.5. Método autoconsistente de Hartree-Fock

Este método, desarrollado por D. R. Hartree desde 1928 y por V. A. Fock hasta 1930 [22], es utilizado para resolver la ecuación de Schrödinger para sistemas atómicos de más de un electrón. El concepto surge de la idea de que cada electrón interactúa con los campos eléctricos producidos por el resto de los electrones y el núcleo del átomo, que a su vez afecta a los mismos electrones. Éste es un método muy complejo y sólo se desarrolla en forma numérica.

Para empezar a desarrollar el método consideremos un átomo con N electrones. El hamiltoniano del k -ésimo electrón independiente es $\hat{H}_0(r_k)$, pero considerando la repulsión coulombiana del resto de los electrones, el hamiltoniano resulta

$$H = \sum_k \hat{H}_0(\mathbf{r}_k) + \sum_{k < l} V(\mathbf{r}_k, \mathbf{r}_l) = \sum_k \hat{H}_0(\mathbf{r}_k) + \frac{1}{2} \sum_{l \neq k} V(\mathbf{r}_k, \mathbf{r}_l). \quad (2.33)$$

Se propone una función que es solución del hamiltoniano

$$\varphi(\mathbf{r}_1, \dots, \mathbf{r}_N) = \varphi_1(\mathbf{r}_1)\varphi_2(\mathbf{r}_2) \dots \varphi_N(\mathbf{r}_N). \quad (2.34)$$

El valor esperado del hamiltoniano queda

$$\langle \hat{H} \rangle = \sum_k \int \varphi_k^* \hat{H}_0 \varphi_k d^3 \mathbf{r}_k + \frac{1}{2} \sum_k \int \varphi_k^* \varphi_l^* V(\mathbf{r}_k, \mathbf{r}_l) \varphi_k \varphi_l d^3 \mathbf{r}_k d^3 \mathbf{r}_l. \quad (2.35)$$

Esta cantidad debe corresponder a un mínimo, y debe, además, ser estacionaria frente a cambios arbitrarios de la función de prueba. Variando las φ_k^* y dejando las φ_k fijas, se obtiene

$$\delta \langle \hat{H} \rangle = \sum_k \int \delta \varphi_k^* \left[\hat{H}_0 + \sum_{l \neq k} \langle l | V(k, l) | l \rangle \right] \varphi_k d^3 \mathbf{r}_k, \quad (2.36)$$

pero las $\delta\varphi_k^*$ no son independientes, pues de la condición de que la norma de cada función de onda no se altere se sigue que

$$\int \delta\varphi_k^* \varphi_k d^3\mathbf{r}_k = 0. \quad (2.37)$$

Para tomar en cuenta estas restricciones al calcular $\delta\langle\hat{H}\rangle$ se emplea el método de los multiplicadores indeterminados de Lagrange [22]. Entonces se multiplica la ecuación (2.37) por la constante indeterminada E_k y se resta la suma de estas expresiones a la ecuación (2.36); se obtiene

$$\sum_k \int \delta\varphi_k^* \left[\hat{H}_0 + V_k - E_k \right] \varphi_k d^3\mathbf{r}_k = 0, \quad (2.38)$$

en donde se ha introducido la abreviación

$$V_k = \sum_{l \neq k} \langle l | V(k, l) | l \rangle = \sum_{l \neq k} \int \varphi_l^* V(\mathbf{r}_k, \mathbf{r}_l) \varphi_l d^3\mathbf{r}_l; \quad (2.39)$$

$V_k(\mathbf{r}_k)$ representa el potencial promedio en el punto \mathbf{r}_k ocupado por el electrón k , producido por el resto de los electrones. Al contrario de la ecuación (2.36), en la expresión (2.38), las variaciones arbitrarias $\delta\varphi_k^*$ pueden considerarse independientes, entonces sólo se cumple si se satisface el siguiente sistema de ecuaciones

$$(\hat{H}_0 + V_k)\varphi_k = E_k\varphi_k. \quad (2.40)$$

Este último grupo de ecuaciones consiste en la ecuación de Schrödinger para cada electrón que se mueve en el campo producido por el núcleo y el potencial promedio V_k producido por el resto de los electrones. Sin embargo, el método expuesto no ha tomado en consideración el principio de exclusión, esto se corrige al

sustituir la función de prueba (2.34) por su correspondiente función antisimétrica, incluyendo la dependencia espinorial; la función de prueba quedaría

$$\Psi(\xi_1, \dots, \xi_N) = \sum_{P_k} \hat{P}_k a_{P_k} \varphi_1(\xi_1) \varphi_2(\xi_2) \dots \varphi_N(\xi_N). \quad (2.41)$$

El cálculo es complicado, pero puede realizarse explícitamente. A final se obtiene un sistema de ecuaciones análogo a las ecuaciones (2.40), llamado ecuaciones de Fock-Dirac

$$(\hat{H}_0 + V_k)\varphi_k - \sum_{l \neq k} V_{kl}\varphi_l = E_k\varphi_k, \quad (2.42)$$

donde

$$V_{kl} = \int \varphi_k^* V(k, l) \varphi_l d^3 \mathbf{r}_l = \langle k | V(k, l) | l \rangle. \quad (2.43)$$

El procedimiento numérico que se utiliza para resolver estas ecuaciones consiste en encontrar una solución autoconsistente de la siguiente manera: se hace una selección inicial $\varphi_k^{(0)}$ de las funciones φ_k , con las que se calcula una primera aproximación a los potenciales V_k (y V_{kl} , en su caso), los cuales se introducen en el sistema de ecuaciones que se resuelve para determinar una nueva aproximación $\varphi_k^{(1)}$ de las funciones de onda. Con estos nuevos valores se recalculan los potenciales efectivos y se repite el proceso para obtener una siguiente aproximación $\varphi_k^{(2)}$, y así sucesivamente.

El procedimiento continúa hasta que las funciones de onda que se obtienen coinciden con las que se introdujeron en el paso anterior, dentro de los errores permisibles. Finalmente, con las soluciones obtenidas se calcula la energía con la

fórmula

$$E = \sum_k E_k - \frac{1}{2} \sum_k \sum_{l \neq k} [\langle kl | \hat{V}(k, l) | kl \rangle - \langle kl | \hat{V}(k, l) | lk \rangle], \quad (2.44)$$

la cual se obtiene del producto escalar de la ecuación (2.42) con φ_k y sumando sobre todas las k . Las E_k representan una aproximación a la energía de ionización asociada a cada electrón.

2.5.1. Hartree-Fock restringido y no restringido

Es costumbre en cálculos de campos autoconsistentes en sistemas cerrados de átomos (para los cuales el número de electrones n es siempre impar), suponer que las componentes espaciales de los orbitales son idénticos para cada par de electrones. Entonces hay $\frac{1}{2}n$ orbitales de la forma $\varphi_1(\xi_1)$. Este caso se trata de una función de onda de Hartree-Fock restringido (RHF).

Dos procedimientos son comúnmente utilizados para sistemas abiertos de átomos. En el formalismo de orbital abierto restringido, todos los átomos excepto los que ocupan orbitales abiertos, son forzados a ocupar orbitales doblemente ocupados. Por ejemplo, en la función de onda abierta restringida para el litio los dos primeros orbitales tienen la misma función de onda espacial. Sin embargo, la función de onda restringida impone una fuerte restricción en la solución. En el formalismo de Hartree-Fock no restringido (UHF) y abierto, los electrones del primer orbital no están limitados a la misma función de onda espacial. En cambio, en este caso para el litio, los tres orbitales espaciales son diferentes. Haciendo a un lado la limitante de ocupar orbitales por pares, el formalismo UHF da una menor energía variacional que en el formalismo RHF. Una desventaja de la aproximación

UHF es que mientras que la función de onda RHF es una eigenfunción de S^2 (los momentos angulares del conjunto de partículas), la función UHF no lo es; esto significa que el momento angular total no es una cantidad bien definida en una función de onda UHF.

En la práctica, el valor de expectación de S^2 de una función de onda no restringida es calculada y comparada con el valor verdadero de $S(S+1)\hbar^2$ del estado base. Si la discrepancia no es significativa, el método UHF ha dado entonces una función de onda molecular razonable. La función de onda UHF es comúnmente utilizada como una primera aproximación de la verdadera función de onda aún si la discrepancia es significativa.

2.6. La selección de conjuntos de funciones de base

En principio, un conjunto completo de funciones base deben ser utilizados para representar orbitales exactamente, y el uso de un número infinito de funciones resultaría en una energía de Hartree-Fock equivalente al límite de Hartree-Fock, este límite no es exactamente la energía del estado base de la molécula porque este aún ignora los efectos de la correlación de electrones. Como un conjunto infinito de funciones de base no es computacionalmente factible, un conjunto finito de base es siempre utilizado, y el error debido a la insuficiencia del conjunto de base es llamado error de truncamiento del conjunto de base. La diferencia entre el límite de Hartree-Fock y la menor energía computada en un cálculo de Hartree-Fock es una medida del error de truncamiento del conjunto de base. Una consideración computacional clave entonces sería mantener bajo el número de funciones base

(para minimizar el número de integrales de dos-electrones a evaluar), elegir las inteligentemente (para minimizar el esfuerzo computacional para el cálculo de cada integral), pero no obstante, conseguir un pequeño error de truncamiento de conjunto de base [18, 21].

Una opción de funciones de base para utilizar son los orbitales tipo Slater (STO), como en la ecuación

$$\Phi_{nlm_l}(r, \theta, \phi) = N r^{n_{eff}-1} e^{-Z_{eff}\rho/n_{eff}} Y_{lm_l}(\theta, \phi), \quad (2.45)$$

donde n , l y m_l son los orbitales cuánticos del núcleo del átomo con número atómico Z , N es una constante de normalización, Y_{lm_l} es un armónico esférico y $\rho = r/a_0$. En la práctica, sólo un pequeño número de todas las posibles funciones son utilizadas. Para cálculos SCF atómicos, las funciones de base de STO se centran en el núcleo del átomo, para especies diatómicas y poliatómicas, son centradas en cada uno de los átomos. Sin embargo, para cálculos de Hartree-Fock SCF en moléculas con tres o más átomos, el cálculo de muchas integrales de dos electrones no son prácticas.

La introducción de los orbitales tipo Gaussianos (GTO), por S.F.Boys, jugó un rol importante en los cálculos computacionales de primeros principios. Las funciones Gaussianas en coordenadas cartesianas son de la forma

$$\theta_{ijk}(\mathbf{r}_1 - \mathbf{r}_c) = (x_1 - x_c)^i (y_1 - y_c)^j (z_1 - z_c)^k e^{-\alpha|\mathbf{r}_1 - \mathbf{r}_c|^2}, \quad (2.46)$$

donde $\mathbf{r}_c = (x_c, y_c, z_c)$ son coordenadas cartesianas del centro de la función Gaussiana en \mathbf{r}_c ; $\mathbf{r}_1 = (x_1, y_1, z_1)$ son las coordenadas cartesianas de un electrón en \mathbf{r}_1 ; i , j y k son números enteros no negativos; y α es un exponente positivo.

Cuando $i = j = k = 0$, la Gaussiana cartesiana es una Gaussiana tipo s , cuando $i + j + k = 1$, es una Gaussiana tipo p ; cuando $i + j + k = 2$ es una Gaussiana tipo d , y así sucesivamente. Hay seis tipos de Gaussianas tipo d . Si se prefiere, se puede utilizar una combinación lineal de estas Gaussianas.

La ventaja principal de las GTOs es que el producto de dos funciones Gaussianas en centros diferentes es equivalente a una función Gaussiana centrada en un punto entre los dos centros. Entonces, las integrales de dos electrones en tres y cuatro diferentes centros atómicos se pueden reducir a integrales en dos diferentes centros, que son más fáciles de computar.

El tipo más sencillo de conjunto de base es un conjunto de base mínimo en el cual una función es utilizada para representar cada uno de los orbitales de la teoría de valencia elemental.

Una mejora significativa se obtiene al implementar conjuntos de base doble-zeta (DZ), en donde cada conjunto de base en el mínimo requerido es reemplazado por dos funciones de base. Comparado con un conjunto de base mínimo, el número de funciones de base se duplica. Un conjunto de base DZ para H_2O , por ejemplo, involucraría 14 funciones. En un conjunto de base triple-zeta (TZ), tres funciones base son usadas para representar cada una de las orbitales involucradas en la teoría de valencia elemental.

Un conjunto de base de valencia partida (SV, por sus siglas en inglés) es una combinación entre la deficiencia de un conjunto de base mínimo y la demanda computacional de conjuntos de base DZ y TZ. Cada orbital de valencia atómica

es representado por dos funciones base mientras que cada orbital atómico interno es representado por una función base simple. Por ejemplo, para un cálculo SCF atómico en un carbono usando funciones Gaussianas, hay una función representando el orbital $1s$, dos representando el orbital $2s$, y dos más para cada uno de los tres orbitales $2p$.

Los conjuntos de base descritos hasta ahora ignoran posibles contribuciones de funciones que representan orbitales para los cuales el valor del número cuántico l es más grande que el máximo valor considerado en la teoría de valencia elemental. Como sea, cuando se forman enlaces en las moléculas, los orbitales atómicos se distorsionan (o se polarizan) por los átomos adyacentes. Esta distorsión puede ser tomada en cuenta incluyendo funciones base que representan orbitales con valores altos de l . La inclusión de estas funciones de polarización a un conjunto de base DZ resulta en lo que se llama conjunto de base doble-zeta mas polarización (DZP).

Una aproximación de la construcción de un conjunto de funciones Gaussianas es hacer un ajuste de N funciones Gaussianas primitivas a un conjunto de STOs que han sido optimizados en un cálculo atómico de SCF. La expansión de un STO en términos de N funciones Gaussianas primitivas es designada $STO - NG$. Una opción común es $N = 3$, dando un conjunto de funciones Gaussianas, cuyo caso sería $STO - 3G$. Alternativamente, un cálculo atómico SCF puede ser llevado a cabo usando una base relativamente larga de funciones Gaussianas primitivas. Los exponentes optimizados y coeficientes de funciones Gaussianas primitivas se pueden utilizar para obtener conjuntos base Gaussianos contractos para usar en

cálculos moleculares. En el modelo de contracción $(4s)/[2s]$, cuatro funciones Gaussianas primitivas tipo s son utilizadas para construir dos funciones de conjuntos de base para el átomo de hidrógeno. En el modelo de contracción $(9s5p)/[3s2p]$, nueve Gaussianas primitivas tipo s y cinco tipo p son contraídas en tres y dos funciones base, respectivamente.

Hay otros modelos de contracción, como el conjunto de base $3-21G$, en donde una función Gaussianas contraída y tres primitivas son usadas para representar cada órbita atómica interna. Cada orbital de valencia es representado por dos funciones, una Gaussianas contraída de dos primitivas y una primitiva sencilla. Las primitivas se optimizan primero en un previo cálculo SCF de los átomos, y los conjuntos contraídos son utilizados entonces en el cálculo molecular. El conjunto base $6-31G^*$ añade funciones de polarización en la forma de seis funciones tipo d para cada átomo diferente de H a la base $6-31G$. Otro asterisco, una función de polarización adicional: $6-31G^{**}$ indica la inclusión a $6-31G^*$ de un conjunto de tres funciones de polarización tipo p para cada átomo H .

2.7. Teoría funcional de la densidad

Dentro de la teoría de muchos cuerpos existe una variable fundamental a estudiar, es la densidad electrónica completa. Thomas y Fermi derivaron una ecuación diferencial para la densidad sin recurrir a los orbitales de un electrón; aunque la aproximación de Thomas-Fermi original para la energía cinética de los electrones no era capaz de sustentar estados enlazados. Como sea, esta teoría dio las bases para la posterior creación de la teoría funcional de la densidad (DFT, por sus siglas

en inglés), la cual ha sido la manera de elegir en cálculos de estructura electrónica en la física de materia condensada recientemente.

2.7.1. Teoría de Thomas-Fermi

Thomas y Fermi, de manera independiente, dieron una metodología para calcular la energía de un sistema electrónico exclusivamente en términos de la densidad electrónica. Propusieron una expresión para la energía electrónica total donde las contribuciones cinética, de intercambio y de correlación fueron tomadas de la nube de electrones homogénea. La idea era construir las mismas cantidades para el sistema inhomogéneo como

$$E_\alpha[\rho] = \int \rho(\mathbf{r}) \epsilon_\alpha[\rho(\mathbf{r})] d(\mathbf{r}), \quad (2.47)$$

en donde $\epsilon_\alpha[\rho(\mathbf{r})]$ es la densidad de energía de contribución α (cinética, de intercambio y de correlación), calculada localmente para el valor asumido para la densidad en cada punto en el espacio. Los corchetes indican una dependencia funcional de la energía y la densidad de la energía en la densidad electrónica.

Para una nube de electrones homogénea, la densidad electrónica está relacionada con la energía de Fermi (ϵ_F) de la forma

$$\rho = \frac{1}{3\pi^2} \left(\frac{2m}{\hbar^2} \right)^{3/2} \epsilon_F^{3/2}, \quad (2.48)$$

y la energía cinética es $T = 3\rho\epsilon_F/5$, de modo que la densidad de la energía potencial es

$$t[\rho] = \frac{3}{5} \frac{\hbar^2}{2m} (3\pi^2)^{2/3} \rho^{2/3}. \quad (2.49)$$

Entonces la aproximación de densidad local (LDA) de la energía cinética se escribe como

$$T_{TF} = C_k \int \rho(\mathbf{r})^{5/3} d\mathbf{r}, \quad (2.50)$$

con $C_k = 3(3\pi^2)^{2/3}/10 = 2.871\text{Ha}$ (la unidad atómica de energía, equivalente a dos veces la energía de ionización del átomo de hidrógeno, 27.21eV). Se hace un intercambio poco después, introduciendo la expresión de Slater para la nube de gas electrónico homogéneo

$$E_X[\rho] = -C_X \int \rho(\mathbf{r})^{4/3} d\mathbf{r}, \quad (2.51)$$

con $C_X = 3(3/\pi)^{1/3}/4 = 0.739\text{Ha}$. En este momento, la teoría es llamada ahora de Thomas-Fermi-Dirac (TFD). La correlación se puede incluir usando cualquier aproximación local a la nube de electrones homogénea, por ejemplo la propuesta por Wigner (1938)

$$E_C[\rho] = -0,056 \int \frac{\rho(\mathbf{r})^{4/3}}{0,079 + \rho(\mathbf{r})^{1/3}} d\mathbf{r}, \quad (2.52)$$

donde todas las constantes numéricas son dadas en unidades atómicas. Reemplazando las expresiones anteriores en la expresión general para la energía de un sistema electrónico inhomogéneo [25] se obtiene la expresión de la energía de TFD

$$E_{TFD}[\rho] = C_k \int \rho(\mathbf{r})^{5/3} d\mathbf{r} + \int \rho(\mathbf{r}) v_{ext} d\mathbf{r} + \frac{1}{2} \int \int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' - C_X \int \rho(\mathbf{r})^{4/3} d\mathbf{r} + E_C[\rho], \quad (2.53)$$

donde v_{ext} es la interacción con los campos externos.

Se puede ver que la única dependencia de E_{TFD} en las variables electrónicas es a través de la densidad electrónica. En ese sentido se dice que es una funcional

de la densidad. Asumiendo intuitivamente algún principio variacional, podemos buscar la densidad $\rho(\mathbf{r})$ que minimiza $E_{TFD}[\rho]$, sujeta a la restricción de que la integración total de la densidad de carga es igual al número de átomos: $\int \rho(\mathbf{r}) d\mathbf{r} = N$. Como la variación no es con respecto a un parámetro pero sí a una función, por ejemplo $\rho(\mathbf{r})$, la minimización asume la forma de la búsqueda de una función en espacio tridimensional que hace a la energía estacionaria con respecto a cualquier tipo de variación de densidad.

2.7.2. Derivación funcional

La idea de la derivación funcional es una generalización del concepto de derivación parcial de una función de varias variables al caso de un continuo de variables; en este caso el valor de la densidad en cada punto del espacio. Si queremos minimizar una función de varias variables debemos buscar puntos estacionarios, por ejemplo los puntos donde las derivadas parciales desaparecen y la Hessiana es una matriz positiva definida. La generalización de un continuo implica que debemos buscar densidades tales que la derivada de la funcional de la energía con respecto al valor de la densidad en cada punto en el espacio desaparece. Este procedimiento es llamado derivación funcional.

Suponemos que $F[f]$ es una funcional de $f(x)$, y modificamos la función f añadiendo una pequeña contribución $f(x) \rightarrow f(x) + \xi g(x)$. El cambio en la funcional F está dada por

$$\Delta F = F[f + \xi g] - F[f]. \quad (2.54)$$

Si hacemos discreta la variable x en una rejilla de puntos x_i con $i = 1, \dots, n$,

entonces la función f es representada por sus valores en la rejilla $f_i = f(x_i)$. Esto genera otro conjunto de valores $h_i = f_i + \xi g_i$. Con esta discretización la funcional $F[f]$ se puede pensar como una función de n variables $f_i (i = 1, \dots, n)$. Por lo tanto, un pequeño cambio en F se puede escribir como

$$\Delta F = \sum_{i=1}^n \left(\frac{\delta F}{\delta f_i} \right) \Delta f_i = \xi \sum_{i=1}^n \left(\frac{\delta F}{\delta f_i} \right) g_i, \quad (2.55)$$

en donde para $\xi \rightarrow 0$ se convierte en

$$\lim_{\xi \rightarrow 0} \frac{\Delta F}{\xi} = \sum_{i=1}^n \left(\frac{\delta f}{\delta f_i} \right) g_i = \nabla F \cdot g. \quad (2.56)$$

Regresando al continuo implica que deberíamos considerar puntos en la rejilla infinitamente cercanos, de manera que la derivación es llevada a cabo con respecto al valor de la función en el punto x , por ejemplo $f(x)$. Para poder hacer una distinción con respecto a la derivación parcial, esta derivada es indicada por $\delta F[f]/\delta f(x)$ y nombrada como derivada funcional. En este límite la sumatoria sobre los puntos de la rejilla se convierte en una integral y el cambio en la funcional F termina como

$$\lim_{\xi \rightarrow 0} \frac{\Delta F}{\xi} = \int_{-\infty}^{\infty} \left(\frac{\delta f}{\delta f(x)} \right) g(x) dx. \quad (2.57)$$

Un grupo de reglas de derivación funcional análogas a las de derivación usual pueden ser obtenidas. Por ejemplo

$$\frac{\delta}{\delta f(x)} \left(\int g[f] dx \right) = \frac{\delta g[f]}{\delta f(x)}, \quad (2.58)$$

$$\frac{\delta}{\delta f(x)} \left(\int g[f] f(x) dx \right) = \frac{\delta g[f]}{\delta f(x)} f(x) + g[f(x)], \quad (2.59)$$

$$\frac{\delta}{\delta f(x)} \left(\frac{1}{2} \int \int g(x, x') f(x) f(x') dx dx' \right) = \int g(x, x') f(x') dx', \quad (2.60)$$

donde la función $f(x)$ está definida como la densidad $\rho(\mathbf{r})$. La ecuación (2.59) es útil para calcular la derivada funcional de elementos de la aproximación de densidad local como la cinética, la de intercambio y la de correlación en la aproximación de Thomas-Fermi-Dirac, y la ecuación (2.60) es útil para integrales tipo Coulombiana cuando se identifica $g(x, x')$ con el potencial Coulombiano $1/|\mathbf{r} - \mathbf{r}'|$.

2.7.3. La ecuación Thomas-Fermi-Dirac

Ahora se aplican las reglas de derivación funcional a la funcional de Thomas-Fermi (ecuación (2.53)), sujeta a la restricción de un número fijo de átomos. Esta restricción se introduce vía multiplicadores de Lagrange. De la siguiente ecuación (2.61) se obtiene la ecuación (2.62).

$$\frac{\delta}{\delta\rho(\mathbf{r})}(E_{TFD}[\rho] - \mu \int \rho(\mathbf{r})d\mathbf{r}) = 0, \quad (2.61)$$

$$\mu = \frac{3}{5}C_k\rho(\mathbf{r})^{2/3} + v_{ext}(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}d\mathbf{r}' - \frac{4}{3}C_X\rho(\mathbf{r})^{1/3} + \mu_c[\rho(\mathbf{r})] \quad (2.62)$$

Con μ como el potencial químico y $\mu_c[\rho] = \delta E_c[\rho]/\delta\rho(\mathbf{r})$ el potencial de correlación. Esta última ecuación puede ser invertida para obtener la densidad como una única función del potencial externo. De hecho, esta forma no es fácil de invertir directamente por el término integral y las potencias no enteras de la densidad. En la práctica se resuelve utilizando métodos iterativos.

2.8. Simulación de Dinámica molecular

En las simulaciones mecánicas de dinámica molecular (MD), los átomos de un polímero se mueven de acuerdo con las ecuaciones de movimiento de Newton

$$m_\alpha \ddot{\mathbf{r}}_\alpha = -\frac{\partial}{\partial \mathbf{r}_\alpha} U_{total}(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N), \quad \alpha = 1, 2, \dots, N, \quad (2.63)$$

donde m_α es la masa del átomo α , \mathbf{r}_α es su posición, y U_{total} es la energía potencial total que depende de todas las posiciones atómicas, y por consiguiente, une el movimiento de los átomos. La energía potencial es la parte más importante de la simulación porque debe representar fehacientemente la interacción entre átomos, y debe tener la forma de una función matemática sencilla para que se pueda calcular rápidamente.

2.8.1. Funciones de campo de fuerza

Para una simulación de MD de todos los átomos, se asume que todo átomo experimenta una fuerza especificada por un modelo de campo de fuerza que toma en cuenta la interacción del átomo con el resto del sistema. Hoy en día, estos campos de fuerza representan un compromiso entre exactitud y eficiencia computacional [18]. Para esto se emplea una función de energía potencial común que tiene las siguientes contribuciones

$$U_{total} = U_{enlace} + U_{ángulo} + U_{dihedral} + U_{vdW} + U_{Coulomb}. \quad (2.64)$$

Los primeros tres términos describen, en otras palabras, las interacciones de estiramiento, enlazamiento y de torsión respectivamente,

$$U_{enlace} = \sum_{enlaces\ i} k_i^{enlace} (b_i - b_{0i})^2, \quad (2.65)$$

$$U_{ángulo} = \sum_{ángulos\ i} k_i^{ángulo} (\theta_i - \theta_{0i})^2, \quad (2.66)$$

$$U_{dihedral} = \sum_{dihedrales\ i} k_i^{dihedral} [1 + \cos(n_i \chi_i - \sigma_i)], \quad (2.67)$$

donde los enlaces incluyen cada enlace covalente del sistema, los ángulos son los comprendidos entre cada par de enlaces covalentes que comparten un mismo electrón en el vértice, y los dihedrales describen átomos pares separados por exactamente tres enlaces covalentes con el enlace central sujeto a un ángulo de torsión ϕ (fig. 2.2).

Los dos últimos términos en la ecuación (2.64) describen las interacciones de pares de átomos no enlazados

$$U_{vdW} = \sum_i \sum_{j>i} 4\varepsilon_{ij} \left[\left(\frac{r^{min,ij}}{r_{ij}} \right)^{12} - \left(\frac{r^{min,ij}}{r_{ij}} \right)^6 \right], \quad (2.68)$$

$$U_{Coulomb} = \sum_i \sum_{j>i} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}, \quad (2.69)$$

que corresponden a las fuerzas de van der Waals y a las interacciones electrostáticas, respectivamente.

Para cada partícula, en un dado contexto de enlaces, los parámetros k_i^{enlace} , b_{0i} , etc., para las interacciones dadas en las ecuaciones (2.65) a (2.67) son proporcionadas en un archivo de parámetros para la simulación. La obtención de estos

parámetros se lleva a cabo generalmente por una combinación de técnicas empíricas y cálculos mecánicos cuánticos; el campo de fuerza entonces es probado para probar su confiabilidad reproduciendo las propiedades estructurales, dinámicas y termodinámicas de pequeñas moléculas que han sido bien caracterizadas experimentalmente, así como su confiabilidad reproduciendo propiedades en masa.

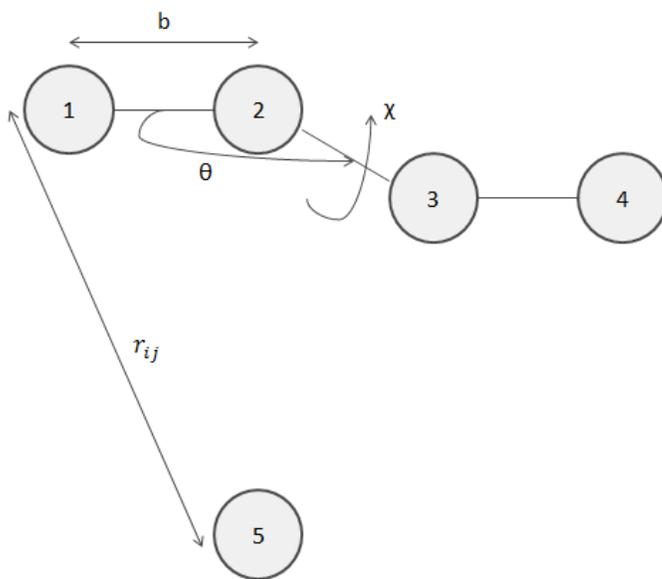


Figura 2.2: Enlaces mecánicos de interacciones partícula-partícula.

3 Cálculo de primeros principios de moléculas orgánicas

En el presente capítulo se presenta el desarrollo computacional de este trabajo, desde la recopilación de los elementos necesarios para poder realizar exitosamente la simulación de la dinámica molecular hasta la interpretación de los resultados arrojados por estos procesos, haciendo mención de cada uno de los paquetes computacionales utilizados.

3.1. Primeros conceptos de simulación molecular

Para estudiar cualquier molécula es necesaria la información teórica referente a esta, como las posiciones, cargas y enlaces de cada átomo que interviene en la configuración de la molécula. Esta información debe ser proporcionada con anterioridad, en efecto existen bases de datos que contienen esta información de una variedad de moléculas que ya han sido medidas experimentalmente [29].

Se parte de un archivo llamado Base de Datos de Proteína, o bien pdb por sus

siglas en inglés, que contiene la información de cada átomo de la proteína, como el nombre, tipo, posición, ocupancia y factor de temperatura del átomo así como el residuo y segmento al que pertenece dentro de la proteína. Este archivo cumple como requisito mínimo para poder visualizar una molécula o proteína, muy útil para distinguir las condiciones iniciales y finales de la proteína.

El siguiente paquete de datos consta de la topología de la molécula, en este archivo se especifica la geometría de cada residuo que está o puede estar en la proteína, donde un residuo se entiende como una parte reconocible dentro de una cadena molecular más grande, especificando también los diferentes tipos de enlaces entre átomos y el tipo, carga, grupo y peso molecular de cada átomo contenido en los residuos. Además hace falta un archivo con los parámetros involucrados en cada tipo de enlace y para cada tipo de átomo, para poder obtener la energía de la configuración atómica, como se explica en la sección 2.8. Los parámetros se pueden obtener experimentalmente o por procedimientos computacionales también, como se verá posteriormente.

La información necesaria para lograr las simulaciones de dinámica molecular existen actualmente en bases de datos que se encuentran en internet de manera gratuita, así como los paquetes computacionales, tales como los que se utilizan en este trabajo de tesis.

Para observar las proteínas se utiliza el paquete Visual Molecular Dynamics (VMD), este puede visualizar la información contenida en los archivos pdb y de topología de las moléculas, especificando tipo de bosquejo y color de las moléculas y restricciones en la visualización. También se utiliza en este trabajo el paquete

gratuito MacMolPlt, que cumple primordialmente con las mismas funciones.

3.2. Generación de parámetros

Puede darse el caso que, para un nuevo residuo, no se conozcan los parámetros mecánicos necesarios para obtener la energía de la configuración atómica (Sección 2.8), y no se pueda obtener esta información por métodos experimentales, la opción recae en el método computacional, en este caso, se utiliza química de primeros principios, en donde se resuelve la ecuación de Schrödinger asociada al Hamiltoniano molecular, sin utilizar métodos empíricos ni semi-empíricos en las ecuaciones.

Para este procedimiento se utiliza el paquete GAMESS “General Atomic and Molecular Electronic Structure System” [19], un paquete computacional de química cuántica con cálculos de primeros principios. Con GAMESS se procede a resolver un sistema de átomos con la teoría de Hartree-Fock (sección 2.5), se sugiere la forma de la función de onda y se obtiene el potencial y la energía del sistema. Con estos resultados se propone una nueva configuración atómica y se recalcula la energía, hasta obtener la configuración óptima del sistema. Este proceso puede llevar días en terminarse, dependiendo de la cantidad de átomos en la estructura atómica, a partir de 30 por ejemplo.

La geometría optimizada de la estructura atómica provee los parámetros necesarios para la simulación de la dinámica molecular, con ayuda de los paquetes de

visualización mencionados, se interpretan los resultados de los procesos del paquete GAMESS, obteniendo los modos normales de vibración y los enlaces óptimos de cada átomo.

3.3. Simulación de la dinámica molecular

Primero se tiene que disolver la proteína, por ejemplo en agua, para asemejar el entorno celular y poder observar la interacción con un medio biológico, este paso también se desarrolla con el paquete VMD, ya sea en una esfera o una caja de agua.

El siguiente paso es más importante y complejo, se trata de la minimización y equilibrio de la proteína, estos procesos difieren por la naturaleza en que implementan el campo de fuerza de la dinámica molecular. La minimización de la energía busca que la energía total de la molécula sea un mínimo local, por ejemplo, relajar la molécula calculando varias veces la energía variando las posiciones de los átomos. El equilibrio involucra dinámica molecular, en donde la segunda Ley de Newton es resuelta para cada átomo del sistema para obtener su trayectoria. La obtención del equilibrio depende de qué tan bien las velocidades, presiones, etc. son distribuidas en todo el sistema en un período de tiempo dado.

Como se puede notar, la minimización de energía y equilibrio de la proteína, además de ser calculados con primeros principios (sección 2.4) con el paquete GAMESS, también se obtiene con cálculos de origen mecánico y electrostático (sección 2.8), el segundo se utiliza en casos cuando la estructura contiene más

de un residuo de preferencia, ésto se debe a que la carga computacional para los cálculos mecánicos no es tan grande como en el cálculo de primeros principios. Aunque el cálculo de primeros principios es más preciso, el cálculo mecánico da resultados aproximados y aceptables.

La minimización de energía y equilibrio para grandes estructuras moleculares se lleva a cabo con el paquete NAMD [20], que arroja como resultados, entre otros, las posiciones atómicas, las velocidades, las presiones y la energía del sistema en varios momentos de la simulación.

3.4. Cálculo del RMSD

Un objetivo importante de la simulación es obtener el “Root Mean Square Deviation” (RMSD) de la proteína, que equivale a caracterizar la cantidad en la que, dada una molécula, se desvía de una posición definida en el espacio; en particular del esqueleto de la proteína, es decir la proteína sin los hidrógenos, pues estos son muy volátiles a comparación del resto de los demás átomos de la proteína, y el RMSD se vería afectado en un gran parte por estos.

$$RMSD_{\alpha}(t_j) = \sqrt{\frac{\sum_{\alpha=1}^{N_{\alpha}} (\mathbf{r}_{\alpha}(t_j) - \langle \mathbf{r}_{\alpha} \rangle)^2}{N_{\alpha}}}, \quad (3.1)$$

$$\langle \mathbf{r}_{\alpha} \rangle = \frac{1}{N_t} \sum_{j=1}^{N_t} \mathbf{r}_{\alpha}(t_j) \quad (3.2)$$

El RMSD se obtiene mediante la expresión de la ecuación 3.1, en donde N_{α} es el número de átomos que están siendo comparados, N_t es el número de pasos en

los que se están comparando los átomos, \mathbf{r}_α es la posición del átomo α a un tiempo t_j y $\langle \mathbf{r}_\alpha \rangle$ es el valor promedio de la posición del átomo α del cual se comparan sus posiciones $\mathbf{r}_\alpha(t_j)$ (ecuación 3.2).

Con el RMSD de la proteína se puede comparar el movimiento de los residuos de los que está compuesta, o también átomo a átomo, y concluir cuáles de estos experimentan más interacción que otros en el momento de disolver la proteína. Estos cálculos son desarrollados con el paquete VMD, y la simulación de la que se toman los datos se obtiene con el paquete NAMD.

4 Resultados

En el presente capítulo se muestran los resultados obtenidos al simular la dinámica molecular de la proteína fotoactiva amarilla (PYP) en un medio biológico, estando en su estado base y también parcialmente desplegada por la interacción con la luz [1]. Se analizó para la proteína en dos fases: con la geometría original utilizada en estudios de interacción de proteínas con luz azul y con difracción de rayos X [26–28] por ejemplo, y con la geometría modificada, añadiendo elementos a un residuo, en específico, uniendo ácido paracumárico (HC4) a la cistamina (CYS) presente en la proteína, como se descubrió recientemente [29].

Después se analizó en este trabajo, la interacción atómica de un nanotubo de *BN* (nitruro de boro) con moléculas de H_2 [2], optimizando la geometría atómica del conjunto para obtener la posición final de las moléculas de H_2 con respecto al nanotubo.

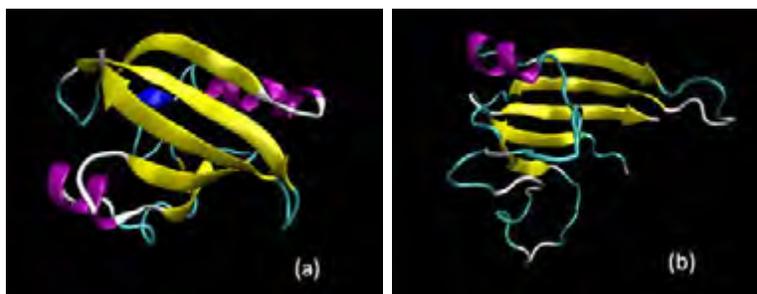


Figura 4.1: La proteína fotoactiva amarilla en su estado base (a) y su estado desplegado (b).

4.1. Proteína fotoactiva amarilla en un medio biológico oscuro e iluminado

La estructura de la proteína se obtuvo de la base de datos electrónica disponible en www.pdb.org, para un medio oscuro (1XFN) y un medio iluminado (1XFQ). Las estructuras se muestran en la figura 4.1 de manera representativa.

Se simuló la dinámica molecular de ambas estructuras dentro de una esfera de agua (fig. 4.2 y 4.3) a una temperatura de 310K, y con una duración de 1ns a lo largo de 500,000 pasos de 2fs. La proteína consta de 100 residuos, numerados del 26 al 125, en las figuras 4.2 y 4.3 se observa el estado final de la proteína después de la simulación, a primera vista no se observa que los átomos se hayan cambiado mucho sus posiciones, sin embargo, en el momento de la simulación se nota con facilidad el movimiento de todos los átomos.

En seguida se calculó el RMSD de la proteína desde su estado inicial hasta el final de la simulación (fig. 4.4), pero sin tomar en cuenta el movimiento de las moléculas de agua. La gráfica obtenida muestra el “movimiento de la proteína” a

4.1. Proteína fotoactiva amarilla en un medio biológico oscuro e iluminado 45

lo largo de la simulación, de esta concluimos cuánto tiempo le lleva relajarse, o dicho de otro modo, cuando llega a una posición de equilibrio. En la gráfica se observa que efectivamente al final, en los últimos 1,000 pasos, el RMSD se mostró menos variante, es decir, que en el último lapso de 0.2ns la proteína se encontraba relajada.

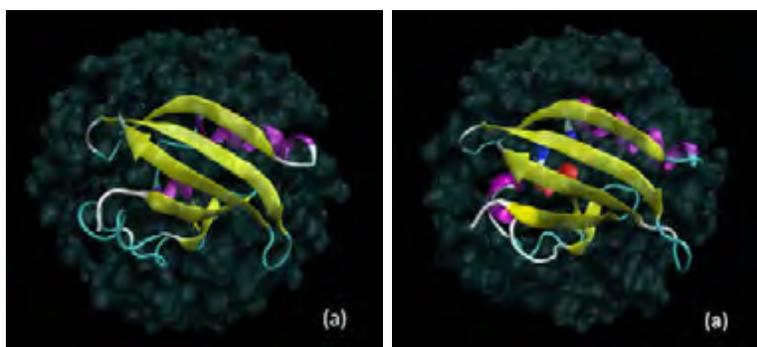


Figura 4.2: La proteína fotoactiva amarilla en su estado base en una esfera de agua antes de la simulación (a) y después de la simulación (b).

Nuevamente se calculó el RMSD de la proteína, pero esta ocasión para cada residuo por separado, para observar el movimiento interno de la proteína de un modo más particular. Se analizó para este caso el RMSD durante 0.2ns al final de la simulación, ya que es cuando la proteína encontró la posición de equilibrio. El resultado se muestra en la figura 4.5 para ambas estructuras, 1XFN y 1XFQ, donde se observa el RMSD calculado para cada residuo durante el periodo dicho, y enseguida se muestra en la figura 4.6 la diferencia de RMSD de cada residuo entre los estados 1XFN y 1XFQ. En la última se observa una diferencia más notoria de los residuos 45 al 70, lo que indica un mayor movimiento de parte de éstas y por lo tanto una mayor interacción con el medio, mientras que el resto presenta menos

fluctuaciones.

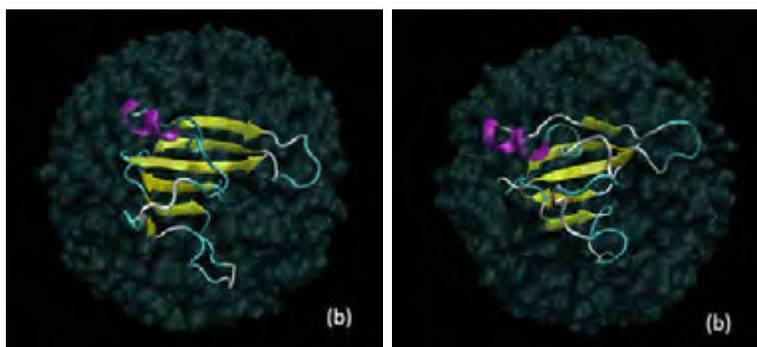


Figura 4.3: La proteína fotoactiva amarilla en su estado excitado en una esfera de agua antes de la simulación (a) y después de la simulación (b).

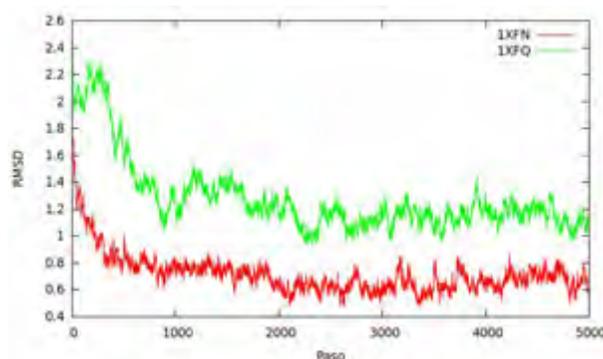


Figura 4.4: RMSD de la proteína calculado para el estado excitado (1XFQ) y su estado base (1XFN).

Estos resultados sin embargo, no son definitivos, existen algunos aspectos de la simulación que provocan discontinuidades. Esto se debe a que al paquete de moléculas, en un principio se le asocia una temperatura, y por lo tanto una energía cinética que lleva a asignar velocidades a cada átomo, estas velocidades no se pueden obtener previamente, ni se les puede asignar manualmente, pues requiere

4.1. Proteína fotoactiva amarilla en un medio biológico oscuro e iluminado 47

mucho trabajo; por consiguiente se generan de manera que en conjunto se logre obtener la temperatura requerida. La generación de las velocidades es aleatoria computacionalmente, y entonces es diferente para cada simulación. Existe la opción de utilizar la misma “semilla aleatoria” para reproducir un resultado, pero el trabajar en diferentes procesadores a la vez no garantiza que el resultado sea el mismo.

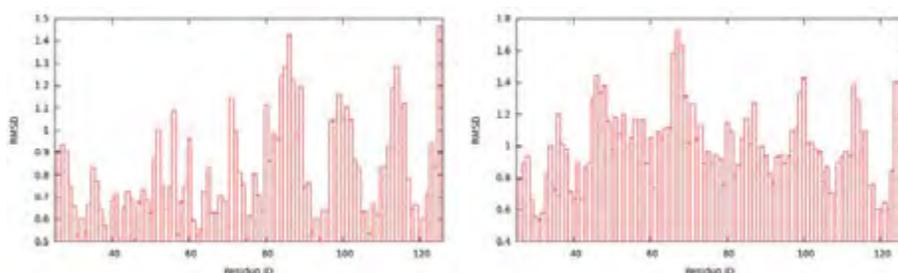


Figura 4.5: RMSD de los residuos para el estado base (izquierda) y su estado excitado (derecha).

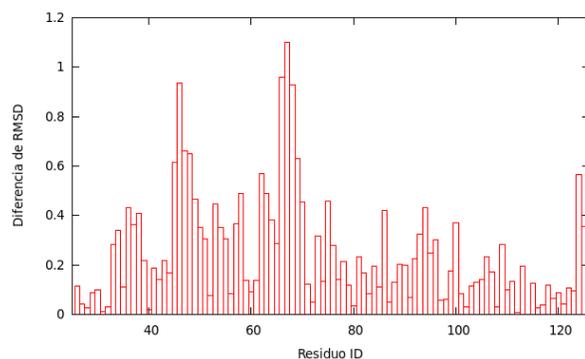


Figura 4.6: Diferencia del RMSD para cada residuo entre los estados 1XFN y 1XFQ.

4.2. Proteína fotoactiva amarilla actualizada

En esta sección se analiza la proteína fotoactiva amarilla de nuevo, pero con una variación, que consta de la actualización del residuo CYS del que consta la proteína, añadiendo el residuo HC4 a la estructura (nótese en el apéndice B).

La cumarina (CYS) contenida en la base de datos electrónica está incompleta para la proteína PYP por la falta del ácido para-cumárico (HC4), en este trabajo se complementa agregándole los átomos faltantes, que corresponden al residuo HC4. En la figura 4.10 se muestra el residuo antes y después de la implementación del resto de los átomos. El principal cambio es la desaparición de un hidrógeno para poder unir el azufre del residuo CYS al carbono del residuo HC4. La actualización del archivo PDB no representa mucha dificultad, ya que sólo implica agregar líneas y cambiar índices; lo mismo sucede con el archivo de topología, se implementan los enlaces necesarios para los nuevos átomos y se definen las cargas de los mismos.

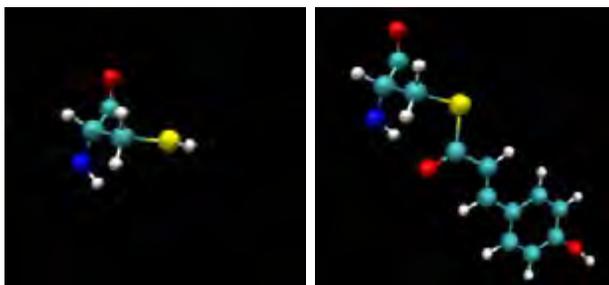


Figura 4.7: El residuo CYS original (izquierda) y el mismo unido al residuo HC4 (derecha). Hidrógenos en blanco, oxígenos en rojo, carbonos en azul claro, nitrógeno en azul marino y azufre en amarillo.

Lo que falta es actualizar el archivo de parámetros, optimizando la estructura

con GAMESS [19] y obteniendo las constantes mecánicas haciendo una comparación con los modos vibracionales de los átomos. Este proceso puede durar hasta un día en terminar en una computadora rápida. Los parámetros conocidos para los átomos utilizados en este trabajo, y otros más, se encuentran en la base de datos de www.pdb.org, pero no se conocen todos los necesarios para la unión CYS-HC4. El archivo de parámetros conocido es muy largo para mostrarlo aquí y se muestra una parte del archivo en el apéndice C.

Al optimizar el residuo se obtienen las distancias y ángulos necesarios para el archivo de parámetros. En las tablas siguientes se presentan los datos optimizados para cada tipo de enlace de los átomos presentes en la modificación del residuo, la inclusión de los demás átomos (fig. 4.7 y 4.8) en la simulación representaría más tiempo de procesamiento y se obtendrían valores ya conocidos previamente en el archivo de parámetros original. Los nombres de los átomos no son los mismos que los dados en el archivo de parámetros, son alias dados para diferenciar átomos que son del mismo tipo (fig. 4.8).

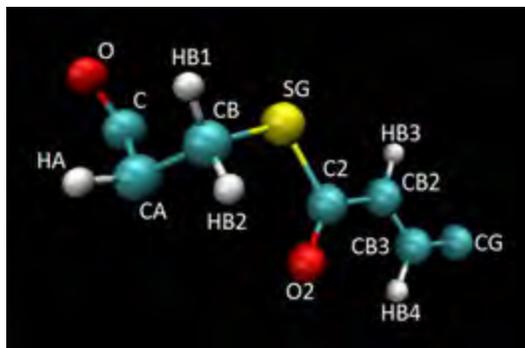


Figura 4.8: Segmento optimizado de la unión CYS-HC4.

Átomo	Átomo	Distancia calculada	Distancia conocida
CB2	CB3	1.33	1.53
CB2	HB3	1.07	1.11
C2	CB2	1.47	1.49
C2	O2	1.21	1.23
C2	SG	1.33	-
CB	SG	1.90	1.81
CB	HB1	1.08	1.11
CB	HB2	1.01	1.11
CB	CA	1.52	1.53

Tabla 4.1: Valores optimizados de las distancias de los enlaces entre los átomos de la unión CYS-HC4.

En la tabla 4.1 se observa que las distancias de los enlaces son similares a las distancias que se tienen en el archivo de parámetros, y sólo se encuentra un enlace que no está en la teoría, este se añadió al archivo de parámetros para poder realizar los cálculos.

En la tabla 4.2 se comparan los ángulos entre los enlaces de los átomos, en este caso los cambios fueron algo diferentes, pero apropiados para nuestros fines, obteniendo 3 nuevos ángulos que no se tienen entre los parámetros conocidos.

Por último, en la tabla 4.3 se muestran los dihedrales o ángulos de torsión de los átomos, en este caso la contribución fue más grande, con seis ángulos dihedrales, estos serán utilizados en el archivo de parámetros para posteriores cálculos. Cabe recalcar que la optimización realizada se efectuó sólo con los átomos mencionados en las tablas 4.1 a la 4.3. El resto de los átomos no son necesarios para este análisis.

De la misma optimización se obtienen los modos de vibración de todos los enlaces y ángulos (fig. 4.9), haciendo comparaciones para cada constante k_i con

Átomo	Átomo	Átomo	Ángulo calculado	Ángulo conocido
CB3	CB2	C2	121.30	108.00
CB2	C2	O2	124.45	121.00
CB2	C2	SG	113.67	-
O2	C2	SG	121.89	-
C2	SG	CB	100.21	-
SG	CB	HB1	105.82	113.20
SG	CB	HB2	103.70	111.30
HB1	CB	CA	112.49	110.10
HB2	CB	CA	112.16	110.10

Tabla 4.2: Valores optimizados de los ángulos entre los enlaces de los átomos en la unión CYS-HC4.

Átomo	Átomo	Átomo	Átomo	Dihedral calculado	Dihedral conocido
CB3	CB2	C2	O	179.77	0.00
CB3	CB2	C2	SG	179.51	-
HB3	CB2	C2	O2	179.82	180.00
HB3	CB2	C2	SG	179.56	-
CB2	C2	SG	CB	178.22	-
O2	C2	SG	CB	177.97	-
C2	SG	CB	HB1	38.64	-
C2	SG	CB	HB2	154.10	-
C2	SG	CB	CA	95.87	-

Tabla 4.3: Valores optimizados de los ángulos de torsión entre los enlaces de los átomos en la unión CYS-HC4.

por la relación $k = m\omega^2$ por ejemplo, asignando una masa adecuada en cada caso.

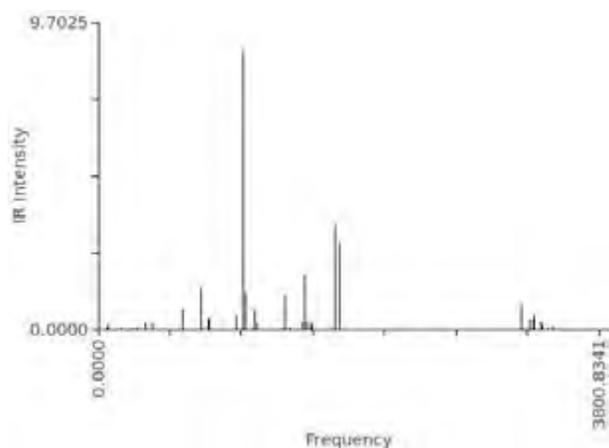


Figura 4.9: Modos de vibración del segmento optimizado (fig. 4.8).

Una vez modificados el archivo PDB, el archivo de topología y el archivo de parámetros, el análisis de la proteína se sigue igual que en la sección anterior, desde disolver la proteína en agua hasta el cálculo del RMSD, de la misma manera para una buena comparación. En las figuras 4.10, 4.11 y 4.12 se muestra la proteína completa, con la unión CYS-HC4 contenida en ella, en el espacio, disuelta en agua y después de la simulación, respectivamente, y para los estados base (1XFN) y excitado (1XFQ).

El cálculo del RMSD se hizo de la misma manera que para el residuo CYS sin presencia del ácido para-cumárico, desde el inicio de la simulación hasta el final. El resultado (fig. 4.13) muestra un comportamiento similar al de la sección anterior (fig. 4.4), con un RMSD final de entre 0.6 y 1.2, esta vez con el RMSD más estable en los últimos 500 pasos.

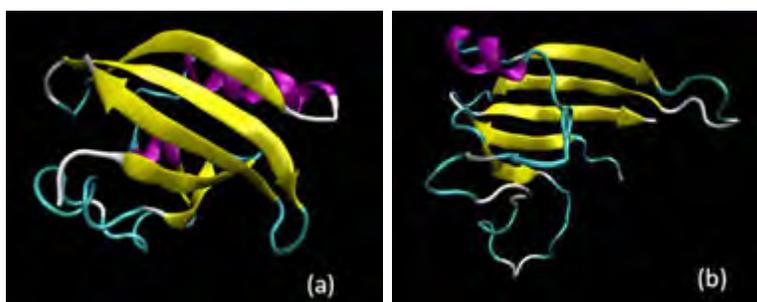


Figura 4.10: La proteína fotoactiva amarilla con ácido para-cumárico, en estado base (a) y estado excitado (b).

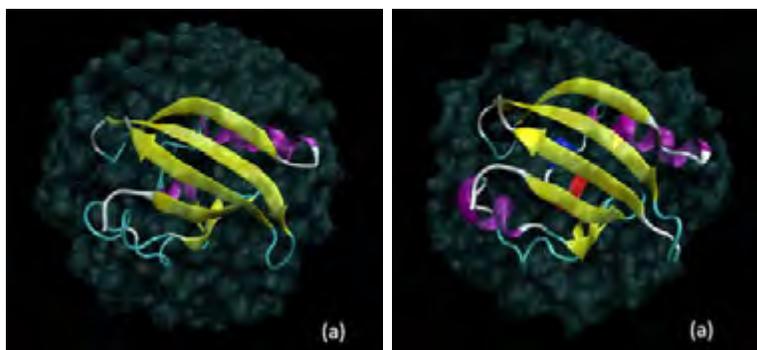


Figura 4.11: La proteína fotoactiva amarilla con ácido para-cumárico en su estado base en la esfera de agua, con ácido para-cumárico, antes de la simulación (a) y después de la simulación (b).

El cálculo del RMSD por residuo en los últimos 500 pasos se muestra en las figuras 4.14 y 4.15, tal como en la sección anterior, para realizar una buena comparación. Se notan ligeros cambios en los resultados, esta vez la zona con más cambio de RMSD fue entre los residuos 43 al 69, y la diferencia de RMSD fue menos grande; ésto significa que la implementación del ácido para-cumárico produjo más estabilidad a la proteína, sin alterar la zona más fotosensible de la proteína.

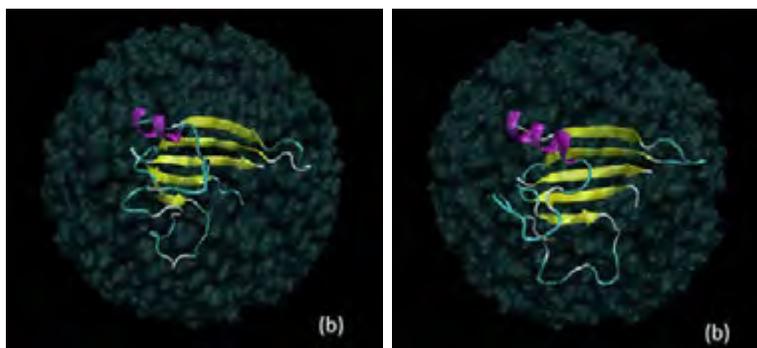


Figura 4.12: La proteína fotoactiva amarilla con ácido para-cumárico en su estado excitado después de la simulación, antes de la simulación (a) y después de la simulación (b).

4.3. Nanotubo de nitruro de boro con moléculas de hidrógeno

En este trabajo de tesis se analizó también la dinámica molecular de un nanotubo de *BN* (nitruro de boro) y su interacción con moléculas de H_2 . Este tipo de procedimiento se ha analizado con anterioridad [30, 31], sin embargo la optimización que se utiliza en este trabajo puede ser más adecuada y precisa. Cada una de las siguientes optimizaciones se desarrollaron con el paquete GAMESS, y la visualización de los resultados se hizo con el paquete VMD y con el paquete MacMolPlt, todos software libre.

Para empezar, se obtuvo la configuración óptima del nanotubo sólo en el espacio. Se complementaron los bordes del nanotubo con hidrógenos para no dejar enlaces libres, y la longitud de este fuera la suficiente para poder procesar adecuadamente (fig. 4.16), en total el nanotubo tiene 48 átomos, incluyendo los hidrógenos

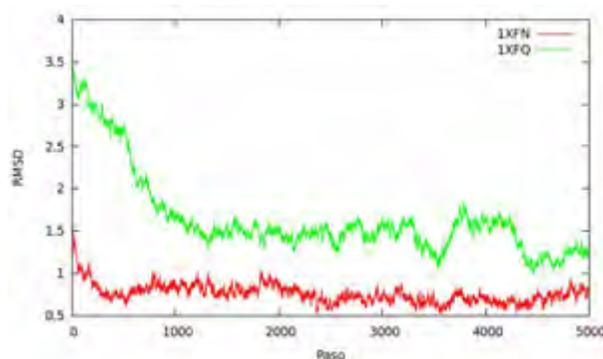


Figura 4.13: RMSD de la proteína calculado para el estado excitado (1XFQ) y su estado base (1XFN).

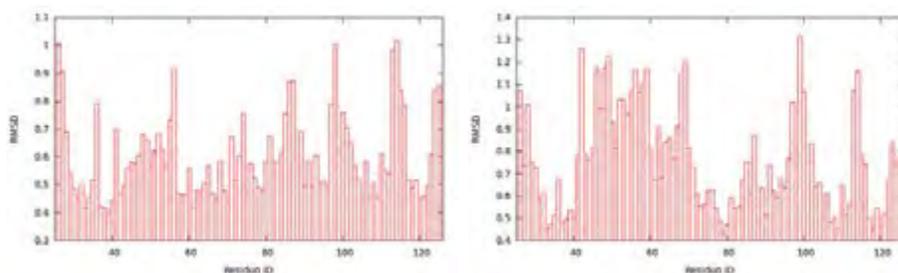


Figura 4.14: RMSD de los residuos para el estado base (izquierda) y su estado excitado (derecha).

de los bordes y mide 7\AA aproximadamente, la incrementación de más átomos al nanotubo representaría demasiada demanda de recursos computacionales. El nanotubo optimizado se obtuvo rápidamente, bastaron unos minutos para terminar el proceso, y el resultado no tuvo cambios considerables. Esto significa que las coordenadas dadas inicialmente son bastante confiables [2].

El siguiente paso fue colocar una molécula de H_2 cerca del nanotubo. Ésta se posicionó con una orientación paralela al nanotubo, y a 1.6\AA de distancia. Esta

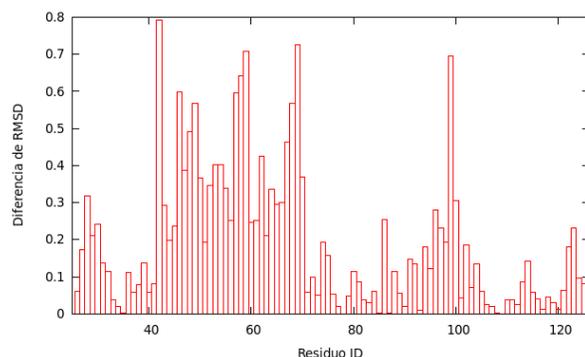


Figura 4.15: Diferencia del RMSD para cada residuo entre los estados 1XFN y 1XFQ.

vez el proceso fue muy tardado, con una capacidad muy alta de procesamiento disponible (12 horas con 36 procesadores trabajando a la vez). Cada una de las optimizaciones desarrolladas en esta sección se hicieron con la misma cantidad de nodos y procesadores. La estructura optimizada arrojó una molécula distanciada del nanotubo a 1\AA más de distancia, y con la orientación cambiada (fig. 4.17). Este resultado parece indicar que el nanotubo provoca un distanciamiento de las moléculas de H_2 que se encuentren cerca de él. El cambio en la energía total se muestra en la figura 4.18. Para confirmar el razonamiento se desarrollaron posteriores pruebas con más moléculas de H_2 .

La siguiente prueba constó de 2 moléculas alrededor del nanotubo, el proceso duró también 12 horas, y el resultado mostró de nuevo un desplazamiento de las moléculas, muy similar al caso anterior, como se observa en la figura 4.19: la molécula posicionada a 1.4\AA se alejó 1.1\AA y cambió su inclinación y la molécula posicionada a 2.5\AA permaneció en su lugar, indicando que ya se encuentra en una posición óptima. En la figura 4.20 se muestra el decaimiento de la energía total

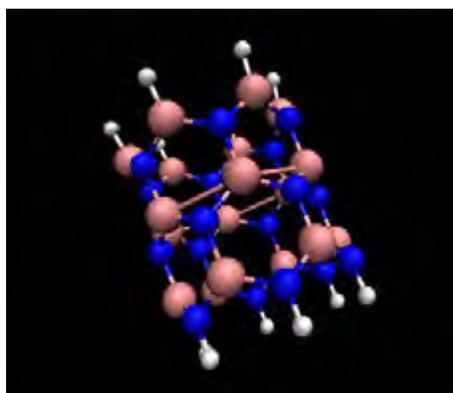


Figura 4.16: Nanotubo de carbono con coordenadas optimizadas. Nitrógenos en azul marino, boros en color cobre e hidrógenos en blanco.

durante la simulación.

Para el siguiente caso se colocaron 3 moléculas de H_2 alrededor del nanotubo, con más de medio día de procesamiento, el resultado continuó reafirmando el distanciamiento entre el nanotubo y las moléculas de H_2 . En la figura 4.21 se muestran los cambios: la molécula posicionada a 0.43Å y la otra a 1.02Å se alejaron hasta quedar a 2.5Å , igual que en los casos anteriores y coincidiendo con la otra molécula que ya se encuentra a 2.5Å y que no se movió de posición, también se observa que los enlaces hidrógeno-hidrógeno se “homogeneizaron” para coincidir a una distancia de 0.75Å . En la figura 4.22 se muestra la variación en la energía total a lo largo de la simulación.

Finalmente, se analizó el caso con 4 moléculas de hidrógeno, el proceso tardó 26 horas en completarse, la implementación de más elementos a la estructura podría exceder la cantidad de recursos computacionales que tenemos disponibles. En la figura 4.23 se ven las diferencias entre las posiciones. Se reafirmó la optimización

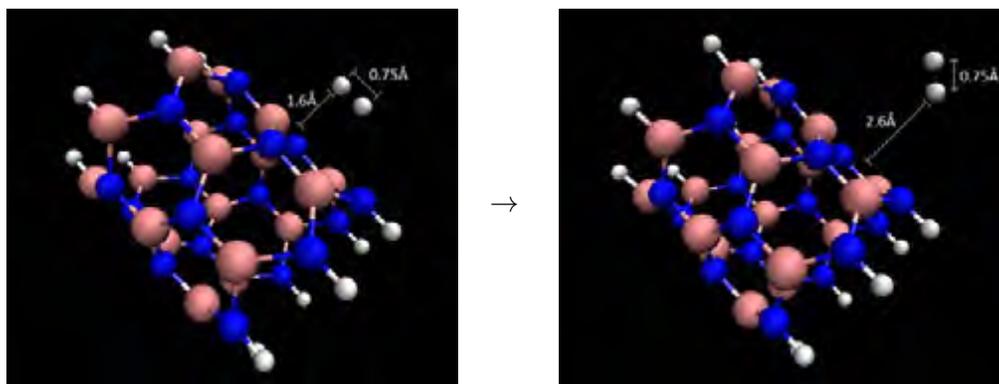


Figura 4.17: Nanotubo de carbono con una molécula de H_2 cercana.

de los enlaces hidrógeno-hidrógeno a la distancia de 0.75\AA , y el distanciamiento de cada uno de las moléculas a 2.5\AA . Para este último caso se eligieron distancias muy diferentes, precisamente para corroborar esta conclusión. En la figura 4.24 se muestra la gráfica de la energía total contra el número de paso de la simulación.

Hay que mencionar que estos procesos llevados a cabo con una computadora común, de 4 procesadores por ejemplo, tardarían alrededor de 100 horas, en lugar de las 12 horas de estas optimizaciones, y se vería propensa a diversos factores negativos para la finalización del trabajo, como una falla de la energía eléctrica por ejemplo. Otra consideración de estas simulaciones es que el paquete GAMESS desarrolla estas operaciones en un ambiente a temperatura constante, de modo que no simula reacciones químicas relacionadas con fluctuaciones en la temperatura, sólo casos ideales en que la temperatura no varía.

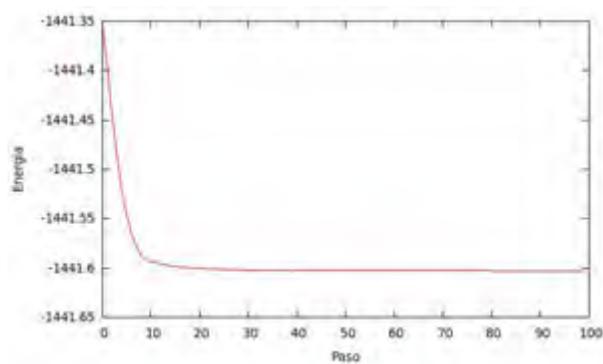


Figura 4.18: Energía total del sistema durante cada paso de la optimización del nanotubo con una molécula de H_2 .

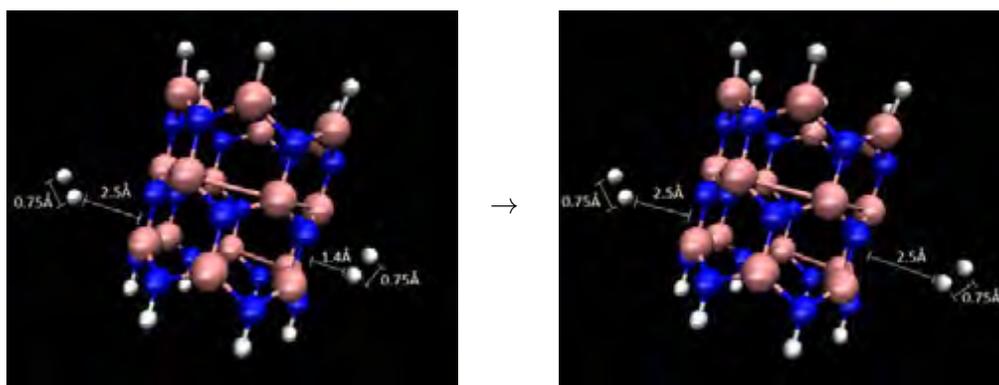


Figura 4.19: Nanotubo de carbono con dos moléculas de H_2 cercana.

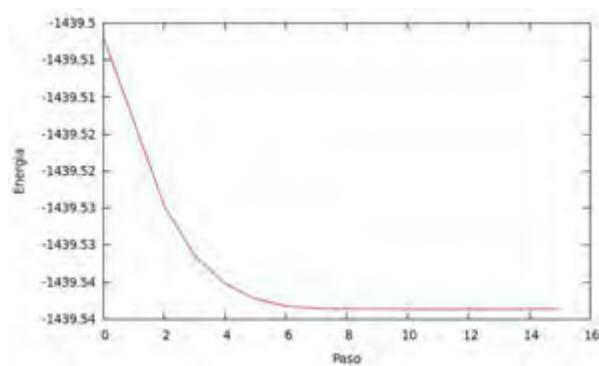


Figura 4.20: Energía total del sistema durante cada paso de la optimización del nanotubo con dos moléculas de H_2 .

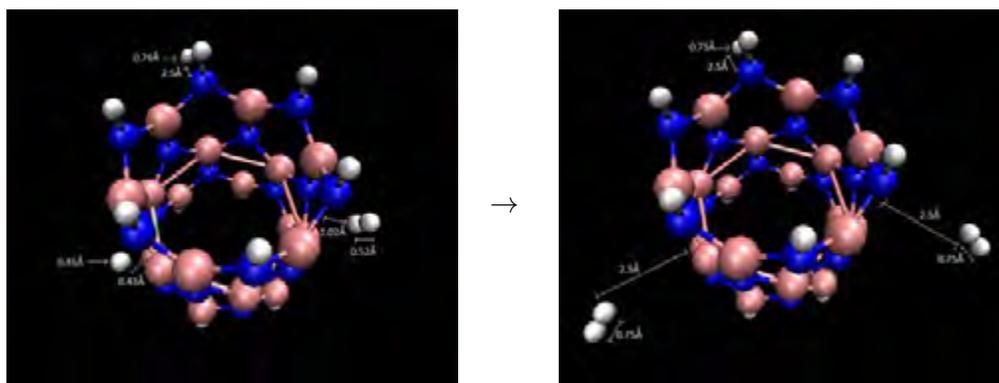


Figura 4.21: Nanotubo de carbono con tres moléculas de H_2 cercana.

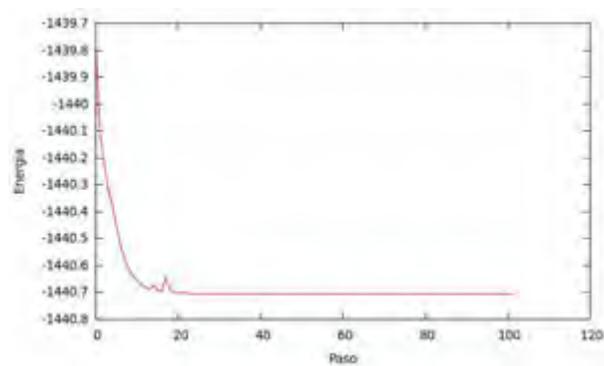


Figura 4.22: Energía total del sistema durante cada paso de la optimización del nanotubo con tres moléculas de H_2 .

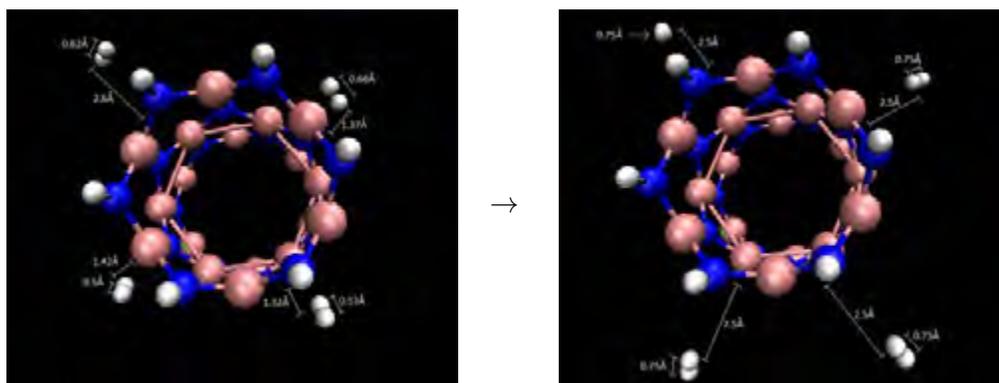


Figura 4.23: Nanotubo de carbono con cuatro moléculas de H_2 cercana.

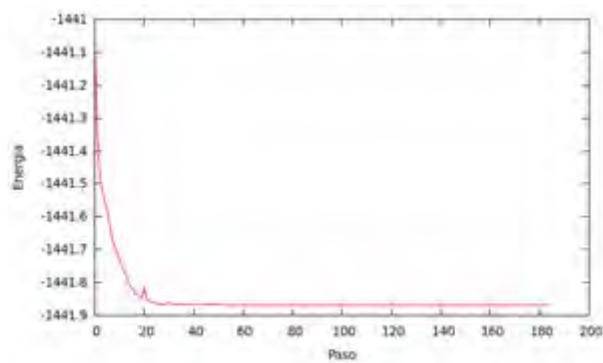


Figura 4.24: Energía total del sistema durante cada paso de la optimización del nanotubo con cuatro moléculas de H_2 .

5 Conclusiones

Desde un principio no fue sencillo imaginar cómo estudiar los efectos microscópicos de la naturaleza, mucho menos los nanométricos, sin embargo, con el paso del tiempo nos las hemos arreglado para poder observarlos y medirlos a un nivel muy razonable gracias a los avances tecnológicos. Hoy en día tenemos la herramienta computacional, que nos permite también observar y medir e incluso predecir la interacción molecular de una infinidad de estructuras atómicas presentes en la naturaleza con una precisión nanométrica.

Esta contribución tecnológica tiene grandes expectativas, principalmente por el hecho de poder realizar los cálculos con mayor comodidad que en el trabajo experimental y posteriormente con una mayor precisión. En este trabajo de tesis se realizaron pruebas teóricas de algunos casos de estructuras atómicas, ambas presentes en el ámbito científico.

En el primer caso se simuló la protreína fotoactiva amarilla, en su estado base y en su estado excitado o iluminado; se simuló para condiciones comunes para

un entorno biológico, solvatando la proteína en agua a una temperatura ambiente promedio, tal como se haría en algunos casos experimentales. Los resultados obtenidos reflejan un comportamiento tal como era esperado, ya que el comportamiento de esta proteína ya se ha observado experimentalmente (debido a este comportamiento se la llama así a la proteína). La implementación del ácido paracumárico faltante en la estructura comprueba el comportamiento de la proteína, obteniendo resultados similares a cuando no se encuentra en la estructura. En otro aspecto, se obtienen parámetros mecánicos faltantes para hacer un adecuado análisis teórico de la proteína fotoactiva amarilla y posiblemente de más proteínas y moléculas no mencionadas.

En el segundo caso se simuló el comportamiento de un nanotubo de BN interactuando con moléculas de H_2 . En este caso se llegó a la conclusión de que el nanotubo sirve para repeler escasamente moléculas de hidrógeno, llevándolas hasta una distancia fija ideal de 2.5\AA en promedio, y también definiendo la distancia óptima del enlace hidrógeno-hidrógeno como de 0.75\AA para todas las moléculas de H_2 . Cada una de las optimizaciones también confirmó la estructura optimizada del segmento del nanotubo manteniéndola casi intacta. El resultado provee también de la configuración óptima del nanotubo para posteriores trabajos teóricos en que el mismo nanotubo esté involucrado.

Referencias

- [1] E. Castro-Camus, B. M. Jonhston, *Conformational changes of photoactive yellow protein monitored by terahertz spectroscopy*. Chemicals Physics Letters **455**, 289-292. (2008).

- [2] Javad Behestian, Hamed Soleymanabadi, Mohammad Kamfiroozi, Ali Ahmadi, *The H₂ dissociation on the BN, AlN, BP and AlP nanotubes: a comparative study*. Springer-Verlag. (2011).

- [3] Deyu Lu, Aleksei Aksimentiev, Amy Y Shih, Eduardo Cruz-Chu, Peter L Freddolino, Anton Arkhipov and Klaus Schulten, *The role of molecular modeling in bionanotechnology*. Beckman Institute for Advanced Science and Technology, University of Lillinois at Urbana-Champaign. (2006).

- [4] Iijima S., *Helical microtubules of graphitic carbon*. Nature **354**, 56-8. (1991).

- [5] Saito R., Dresselhaus G., Dresselhaus M., *Physical properties of carbon nanotubes*. London: Imperial College Press. (1998).

- [6] Kam N. W. S., Jessop T. C., Wender P. A., Dai H., *Nanotube molecular transporters: internalization of carbon nanotube-protein conjugates into mammalian cells*. J. Am. Chem. Soc. **126**, 6850-1. (2004).
- [7] Cherukuri P., Bachilo S. M., Litovsky S. H., Weisman R. B., *Near-infrared fluorescence microscopy of single-walled carbon nanotubes in phagocytic cells*. J. A. Chem. Soc. **126**, 15638-9. (2004).
- [8] Kam N. W. S., O'Connell M., Wisdom J. A., Dai H., *Carbon nanotubes as multifunctional biological transporters and near-infrared agents for selective cancer cell destruction*. Proc. Natl. Acad. Sci. USA. **102**, 11600-9. (2005).
- [9] Barone P. W., Baik S., Heller D. A., Strano M. S., *Near-infrared optical sensors based on singled-walled carbon nanotubes*. Nat. Mater. **4**, 86-92. (2005).
- [10] Staii C., Johnson A. T. Jr., *DNA-decorated carbon nanotubes for chemical sensing*. Nano Lett. **5** 1774-8. (2005).
- [11] Iijima S., *Helical microtubules of graphitic carbon*. Nature **354**, 56-8. (1991).
- [12] Li J., Gershow M., Stein D., Brandin E., Golovchenko J. A., *DNA molecules and configurations in a solid-state nanopore microscope*. Nature Matter **2**, 611-5. (2003).
- [13] Heng J. B., Ho C., Kim T., Timp R., Aksimentiev A., Grinkova Y. B., Sligar S., Schulten K., Timp G., *Sizing DNA using a nanometer-diameter pore*. Biophys. J. **87**, 2905-11. (2004).

- [14] Chang H., Kosari F., Andreadakis G., Alam M. A., Vasmatazis G., Bashir R., *DNA-mediated fluctuations in ionic current through silicon oxide nanopore channels*. Nano Lett. **4**, 1551-6. (2004).
- [15] Heng J. B., Aksimentiev A., Ho C., Dimitrov V., Sorsch T., Miner J., Mansfield W., Schulten K., Timp G., *Beyond the gene chip*. Bell Labs Tech J. **10**, 5-22. (2005).
- [16] Sligar S. G., *Finding a single-molecule solution for membrane proteins*. Biochem. Biophys. Res. Commun. **312**, 115-9. (2003).
- [17] Bayburt T. H., Grinkova Y. V., Sligar S. G. *Self-assembly of discoidal phospholipid bilayer nanoparticles with membrane scaffold proteins*. Nano Lett. **2**, 853-6. (2002).
- [18] Becker, MacKerell, Roux, Watanabe, *Computational Biochemistry and Biophysics*. Taylor & Francis. (2001).
- [19] Journal of Computational Chemistry, Vol. 14, No. 11, 1347-1362 *General Atomic and Molecular Modeling Structure System*. John Wiley & Sons. (1993).
- [20] James C. Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel, Laxmikant Kale, Klaus Schulten, *Scalable Molecular Dynamics with NAMD*. Wiley Periodicals. (2005).
- [21] Cotton F. Albert, *Chemical Applications of Group Theory, Third Edition*. Wiley India. (2008).

- [22] Luis de la Pena, *Introduccion a la Mecanica Cuantica*. Fondo de Cultura Economica. (2006).
- [23] Ira N. Levine, *Quimica Cuantica*. Editorial AC, Liros Cientificos y Tecnicos. (2008).
- [24] Slater J. C., *The self-consistent field and the structure of atoms*. Phys. Rev. **32**, 339-348. (1930).
- [25] Jorge Kohanoff, *Electronic Structure Calculations for Solids and Molecules, Theory and Computational Methods*. Cambridge. (2006).
- [26] G. Rubinstenn, G.W. Vuister. F. A. A. Mulder, P. E. Dux, R. Boelens, K. J. Hellingwerf. R. Kaptein. Nat. Struct. Biol. **5**, 568. (1998).
- [27] ED. Getzoff, K. N. Gutwin, U. K. Genick, Nat. Struct. Biol. **10**, 663. (2003).
- [28] U. K. Genick, S. M. Soltis, P. Kuhn, I. L. Canestrelli, E. D. Getzoff, Nature **392**, 206. (1998).
- [29] Bernard, C., Houben, K., Derix, N. M., Marks, D., van der Horst, M. A., Hellingwerf, K. J., Boelens, R., Kaptein, R., van Nuland, N. A., Journal: STRUCTURE **13**, 953-962 (2005).
- [30] Jhi S-H, Kwon Y-K, *Hydrogen absorption in boron nitride nanotubes: a path to room-temperature hydrogen storage*. Phys Rev. B **69**, 245407-2454010. (2004).
- [31] Wu X, Yang JL, Zeng XC, *Absorption of hydrogen molecules on the platinum-doped boron nitride nanotubes*. J. Chem. Phys. **125**, 44704-44709. (2004).

A Archivos PDB (Protein Data Bank)

En este apéndice se muestra una porción del archivo PDB de la proteína fotoactiva amarilla, ya que es muy extensa para mostrar en esta tesis.

Los campos mostrados a continuación son, de izquierda a derecha, tipo de dato, ID del átomo, nombre del átomo, nombre del residuo, ID del residuo, coordenadas x, y, z, ocupancia, factor de temperatura (beta), nombre del segmento y número de línea. Este archivo es el requisito mínimo para visualizar la molécula con el paquete VMD.

Archivo 1XFN.pdb (proteína fotoactiva amarilla relajada):

```
ATOM 1 N LEU A 26 -8.692 -8.096 11.303 1.00 0.00 U N
ATOM 2 HT1 LEU A 26 -8.318 -8.880 11.800 1.00 0.00 U H
...
...
...
ATOM 639 C PRO A 68 -4.657 4.460 -9.489 1.00 0.00 U C
ATOM 640 O PRO A 68 -3.894 5.261 -10.035 1.00 0.00 U O
```

```
ATOM 641 N CYS A 69 -5.356 4.770 -8.397 1.00 0.00 U N
ATOM 642 HN CYS A 69 -5.937 4.128 -7.901 1.00 0.00 U H
ATOM 643 CA CYS A 69 -5.315 6.117 -7.836 1.00 0.00 U C
ATOM 644 HA CYS A 69 -5.290 6.809 -8.664 1.00 0.00 U H
ATOM 645 CB CYS A 69 -6.562 6.413 -6.986 1.00 0.00 U C
ATOM 646 HB1 CYS A 69 -7.282 5.836 -7.373 1.00 0.00 U H
ATOM 647 HB2 CYS A 69 -6.777 7.469 -7.039 1.00 0.00 U H
ATOM 648 SG CYS A 69 -6.433 5.974 -5.213 1.00 0.00 U S
ATOM 649 HG1 CYS A 69 -7.629 6.432 -4.841 1.00 0.00 U H
ATOM 650 C CYS A 69 -4.046 6.314 -7.009 1.00 0.00 U C
ATOM 651 O CYS A 69 -3.586 7.445 -6.846 1.00 0.00 U O
ATOM 652 N THR A 70 -3.324 5.219 -6.813 1.00 0.00 U N
ATOM 653 HN THR A 70 -3.532 4.321 -7.199 1.00 0.00 U H
...
...
...
ATOM 1555 HG22 VAL A 125 -4.845 9.012 10.542 1.00 0.00 U H
ATOM 1556 HG23 VAL A 125 -5.057 10.131 9.195 1.00 0.00 U H
END
```

Archivo 1XFN.pdb con la modificación del residuo CYS:

```
...
...
...
```

ATOM 639 C PRO A 68 -4.893 4.641 -9.694 1.00 0.00 U C
ATOM 640 O PRO A 68 -4.669 5.522 -10.524 1.00 0.00 U O
ATOM 641 N CYY A 69 -5.221 4.920 -8.439 1.00 0.00 U N
ATOM 642 HN CYY A 69 -5.449 4.193 -7.821 1.00 0.00 U H
ATOM 643 CA CYY A 69 -5.246 6.299 -7.977 1.00 0.00 U C
ATOM 644 HA CYY A 69 -5.326 6.926 -8.854 1.00 0.00 U H
ATOM 645 CB CYY A 69 -6.452 6.582 -7.062 1.00 0.00 U C
ATOM 646 HB1 CYY A 69 -6.543 7.649 -6.928 1.00 0.00 U H
ATOM 647 HB2 CYY A 69 -7.347 6.215 -7.544 1.00 0.00 U H
ATOM 648 SG CYY A 69 -6.378 5.833 -5.396 1.00 0.00 U S
ATOM 649 C CYY A 69 -3.937 6.643 -7.267 1.00 0.00 U C
ATOM 650 O CYY A 69 -3.567 7.814 -7.195 1.00 0.00 U O
ATOM 651 C2 CYY A 69 -7.100 4.190 -5.699 1.00 0.00 U C
ATOM 652 O2 CYY A 69 -7.324 3.768 -6.834 1.00 0.00 U O
ATOM 653 CG CYY A 69 -8.212 1.305 -3.426 1.00 0.00 U C
ATOM 654 CB2 CYY A 69 -7.393 3.377 -4.440 1.00 0.00 U C
ATOM 655 HB3 CYY A 69 -7.182 3.792 -3.462 1.00 0.00 U H
ATOM 656 CB3 CYY A 69 -7.905 2.171 -4.544 1.00 0.00 U C
ATOM 657 HB4 CYY A 69 -8.062 1.892 -5.577 1.00 0.00 U H
ATOM 658 CD1 CYY A 69 -7.994 1.708 -2.115 1.00 0.00 U C
ATOM 659 HD1 CYY A 69 -7.583 2.690 -1.928 1.00 0.00 U H
ATOM 660 CE1 CYY A 69 -8.293 0.873 -1.051 1.00 0.00 U C
ATOM 661 HE1 CYY A 69 -8.116 1.202 -0.039 1.00 0.00 U H
ATOM 662 CZ CYY A 69 -8.817 -0.376 -1.293 1.00 0.00 U C

```
ATOM 663 OH CYY A 69 -9.125 -1.210 -0.242 1.00 0.00 U O
ATOM 664 HH CYY A 69 -10.085 -1.080 0.007 1.00 0.00 U H
ATOM 665 CD2 CYY A 69 -8.742 0.039 -3.639 1.00 0.00 U C
ATOM 666 HD2 CYY A 69 -8.771 -0.347 -4.646 1.00 0.00 U H
ATOM 667 CE2 CYY A 69 -9.042 -0.797 -2.580 1.00 0.00 U C
ATOM 668 HE2 CYY A 69 -9.449 -1.780 -2.765 1.00 0.00 U H
ATOM 669 N THR A 70 -3.093 5.630 -7.095 1.00 0.00 U N
ATOM 670 HN THR A 70 -3.180 4.816 -7.669 1.00 0.00 U H
...
...
...
ATOM 1572 HG22 VAL A 125 -8.982 9.576 9.490 1.00 0.00 U H
ATOM 1573 HG23 VAL A 125 -7.522 8.748 8.950 1.00 0.00 U H
END
```

En el campo del residuo 69 se observa la implementación de 17 átomos, y la eliminación de un hidrógeno.

El residuo HC4 originalmente se muestra a continuación:

```
HETATM 1557 C1 HC4 A 169 -6.697 4.172 -5.252 1.00 0.71 C
HETATM 1558 O1 HC4 A 169 -6.833 3.549 -6.305 1.00 0.73 O
HETATM 1559 C2 HC4 A 169 -6.892 3.529 -3.880 1.00 0.72 C
HETATM 1560 C3 HC4 A 169 -7.095 2.230 -3.787 1.00 0.71 C
HETATM 1561 C1'HC4 A 169 -7.264 1.509 -2.546 1.00 0.75 C
HETATM 1562 C2'HC4 A 169 -7.494 0.138 -2.544 1.00 0.80 C
```

HETATM 1563 C3'HC4 A 169 -7.626 -0.562 -1.360 1.00 0.86 C
HETATM 1564 C4'HC4 A 169 -7.529 0.104 -0.160 1.00 0.89 C
HETATM 1565 C5'HC4 A 169 -7.299 1.461 -0.136 1.00 0.90 C
HETATM 1566 C6'HC4 A 169 -7.164 2.152 -1.323 1.00 0.82 C
HETATM 1567 O4'HC4 A 169 -7.661 -0.589 1.024 1.00 0.99 O
HETATM 1568 H2 HC4 A 169 -6.763 4.120 -2.977 1.00 0.78 H
HETATM 1569 H3 HC4 A 169 -7.186 1.768 -4.760 1.00 0.74 H
HETATM 1570 H2'HC4 A 169 -7.382 -0.405 -3.469 1.00 0.86 H
HETATM 1571 H3'HC4 A 169 -7.803 -1.626 -1.377 1.00 0.93 H
HETATM 1572 H5'HC4 A 169 -7.224 1.983 0.805 1.00 1.01 H
HETATM 1573 H6'HC4 A 169 -6.985 3.216 -1.306 1.00 0.87 H

El resto del archivo, y el archivo 1XFQ.pdb (proteína fotoactiva amarilla excitada) se pueden obtener de la base de datos disponible en <http://www.rcsb.org/pdb/home/home.do>.

B Archivo de topología

En este apéndice se muestra la topología del residuo CYS para ambos casos (original y actualizada). Nuevamente, el archivo se muestra incompleto.

El archivo de topología debe definir el tipo, la masa y la carga de cada átomo en cada residuo. Las masas de cada átomo se definen normalmente al principio del archivo. Después se definen, para cada residuo, la lista de átomos del residuo, incluyendo nombre, tipo y carga del átomo. Finalmente, se estipulan los enlaces del residuo, como los enlaces simples, los ángulos y los dihedrales.

Topología del residuo CYS:

```
RESI CYS 0.00
GROUP
ATOM N NH1 -0.47
ATOM HN H 0.31
ATOM CA CT1 0.07
ATOM HA HB 0.09
GROUP
```

ATOM CB CT2 -0.11
ATOM HB1 HA 0.09
ATOM HB2 HA 0.09
ATOM SG S -0.23
ATOM HG1 HS 0.16
GROUP
ATOM C C 0.51
ATOM O O -0.51
BOND CB CA SG CB N HN N CA
BOND C CA C +N CA HA CB HB1
BOND CB HB2 SG HG1
DOUBLE O C
IMPR N -C CA HN C CA +N O
DONOR HN N
DONOR HG1 SG
ACCEPTOR O C
IC -C CA *N HN 1.3479 123.9300 180.0000 114.7700 0.9982
IC -C N CA C 1.3479 123.9300 180.0000 105.8900 1.5202
IC N CA C +N 1.4533 105.8900 180.0000 118.3000 1.3498
IC +N CA *C O 1.3498 118.3000 180.0000 120.5900 1.2306
IC CA C +N +CA 1.5202 118.3000 180.0000 124.5000 1.4548
IC N C *CA CB 1.4533 105.8900 121.7900 111.9800 1.5584
IC N C *CA HA 1.4533 105.8900 -116.3400 107.7100 1.0837
IC N CA CB SG 1.4533 111.5600 180.0000 113.8700 1.8359

```
IC SG CA *CB HB1 1.8359 113.8700 119.9100 107.2400 1.1134
IC SG CA *CB HB2 1.8359 113.8700 -125.3200 109.8200 1.1124
IC CA CB SG HG1 1.5584 113.8700 176.9600 97.1500 1.3341
```

Se puede observar el esquema del residuo, está comentado y no afecta a la estructura del archivo. La modificación de la topología es la siguiente:

```
RESI CYY 0.00
GROUP
ATOM N NH1 -0.47
ATOM HN H 0.31
ATOM CA CT1 0.07
ATOM HA HB 0.09
GROUP
ATOM CB CT2 -0.11
ATOM HB1 HA 0.09
ATOM HB2 HA 0.09
ATOM SG S -0.07
GROUP
ATOM C C 0.51
ATOM O O -0.51
GROUP
ATOM C2 C 0.51
ATOM O2 O -0.51
```

GROUP

ATOM CG CA 0.00

GROUP

ATOM CB2 CT2 0.09

ATOM HB3 HA -0.09

GROUP

ATOM CB3 CT2 0.09

ATOM HB4 HA -0.09

GROUP

ATOM CD1 CA -0.115

ATOM HD1 HP 0.115

GROUP

ATOM CE1 CA -0.115

ATOM HE1 HP 0.115

GROUP

ATOM CZ CA 0.11

ATOM OH OH1 -0.54

ATOM HH H 0.43

GROUP

ATOM CD2 CA -0.115

ATOM HD2 HP 0.115

GROUP

ATOM CE2 CA -0.115

ATOM HE2 HP 0.115

BOND CB CA SG CB N HN N CA
BOND C CA C +N CA HA CB HB1
BOND CB HB2 SG C2 C2 CB2
BOND CB2 HB3 CB3 HB4 CB3 CG
BOND CG CD2 CD2 HD2 CE2 HE2 CE2 CZ
BOND CD1 HD1 CD1 CE1 CE1 HE1
BOND CZ OH OH HH
DOUBLE O C O2 C2 CB2 CB3
DOUBLE CD1 CG CE1 CZ CE2 CD2

El resto de la topología se puede obtener de la base de datos disponible en
<http://www.rcsb.org/pdb/home/home.do>

C Archivo de parámetros

En este apéndice se muestra el contenido del archivo de parámetros. Este archivo contiene todas las constantes numéricas necesarias para evaluar las fuerzas y las energías, dada una estructura atómica, tales como los enlaces simples, los ángulos y los dihedrales.

Cada tipo de enlace incluye el nombre de los átomos que lo conforman, la constante de elasticidad y la distancia/ángulo de equilibrio. A continuación se muestran unos ejemplos:

```
BONDS CPH1 CPH1 410.000 1.3600 ! ALLOW ARO
! histidine, adm jr., 6/27/90
CPM CPA 360.000 1.3716 ! ALLOW HEM
! Heme (6-liganded): porphyrin macrocycle (KK 05/13/91)
CPT CA 305.000 1.3680 ! ALLOW ARO
! adm jr., 12/30/91, for jwk
CPT CPT 360.000 1.4000 ! ALLOW ARO
!adm jr., 12/30/91, for jwk
```

ANGLES

CE2 CE1 CT2 48.00 126.00 !

! for 1-butene; from propene, yin/adm jr., 12/95

CE2 CE1 CT3 47.00 125.20 !

! for propene, yin/adm jr., 12/95

CP1 N C 60.000 117.0000 ! ALLOW PRO

! 6-31g* AcProNH2, ProNH2, 6-31g*//3-21g AcProNHCH3 RLD 4/23/93

CP2 CP1 C 52.000 112.3000 ! ALLOW PRO

! 6-31g* AcProNH2, ProNH2, 6-31g*//3-21g AcProNHCH3 RLD 4/23/93

DIHEDRALS

CC CT1 CT2 CA 0.0400 3 0.00 ! ALLOW ARO

! 2.7 kcal/mole CH3 rot in ethylbenzene, adm jr, 3/7/92

CC CT1 NH1 C 0.2000 1 180.00 ! ALLOW PEP POL

! ala dipeptide update for new C VDW Rmin, adm jr., 3/3/93c

CC CT2 NH1 C 0.2000 1 180.00 ! ALLOW PEP POL

! Alanine dipeptide; NMA; acetate; etc. adm jr., 3/3/93c

CD CP1 N C 0.0000 1 180.00 ! ALLOW PRO PEP

! 6-31g* AcProNH2, ProNH2, 6-31g*//3-21g AcProNHCH3 RLD 4/23/93

La mayoría de los parámetros obtenidos de la base de datos contiene información extra a cerca de cuándo y cómo fueron obtenidos. Muchos otros parámetros se puede obtener de la base de datos disponible en <http://www.rcsb.org/pdb/home/home.do>