



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MEXICO**

**PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS QUÍMICAS**

Diseño de potenciales externos para inducir el plegamiento  
de proteínas en dinámica molecular

**PROYECTO DE INVESTIGACIÓN  
PARA OPTAR POR EL GRADO DE**

**MAESTRA EN CIENCIAS**

PRESENTA

Q. Tania González García

Dr. Rubén Santamaría Ortiz  
Instituto de Física, UNAM  
**Tutor**

Ciudad de México, enero de 2018



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

## **Lugar donde se desarrolló el proyecto**

Departamento de Física Teórica, Instituto de Física, UNAM

## **Agradecimientos**

A la Universidad Nacional Autónoma de México por brindarme la oportunidad y los recursos para realizar mis estudios de maestría.

Al Consejo Nacional de Ciencia y Tecnología por la beca otorgada para poder realizar mis estudios de maestría. No. de becario: 583782.

## Abreviaturas

- **DM:** Dinámica Molecular
- **DFT:** *Density Functional Theory* (Teoría de los funcionales de la densidad)
- **PDB:** *Protein Data Bank* (Banco de datos de proteínas)
- **ADN:** Ácido desoxirribonucleico
- **QM:** *Quantum Mechanics* (Mecánica Cuántica)
- **MM:** *Molecular Mechanics* (Mecánica Molecular)
- **A $\beta$ :**  $\beta$ -amiloide
- **SMD:** *Steered Molecular Dynamics* (Dinámica Molecular Dirigida)
- **RMN:** Resonancia Magnética Nuclear
- **RMSD:** *Root Mean Square Deviation* (Desviación cuadrática media)

# Índice general

<b>1. Introducción</b>	<b>4</b>
<b>2. Antecedentes</b>	<b>5</b>
2.1. Plegamiento de proteínas . . . . .	5
2.2. Dinámica molecular . . . . .	7
<b>3. Objetivo e hipótesis</b>	<b>9</b>
3.1. Objetivos particulares . . . . .	9
3.2. Hipótesis . . . . .	9
<b>4. Modelo y metodología</b>	<b>10</b>
4.1. Modelo teórico . . . . .	10
4.2. Modelo molecular con disolvente implícito . . . . .	10
4.2.1. Lagrangiano . . . . .	10
4.2.2. Ecuaciones de movimiento . . . . .	11
4.3. Modelo molecular con disolvente explícito . . . . .	12
4.3.1. Potenciales Externos . . . . .	15
4.4. Metodología . . . . .	15
<b>5. Discusión de resultados</b>	<b>17</b>
5.0.1. Primera etapa: parametrización . . . . .	17
5.0.2. Segunda etapa: péptido $\beta$ -amiloide en disolvente implícito . . . . .	18
5.0.3. Segunda etapa: péptido trp-cage en disolvente implícito . . . . .	22
5.0.4. Tercera etapa: Plegamiento del péptido trp-cage con disolvente . . . . .	26
5.0.5. Cuarta etapa: plegamiento del péptido trp-cage en GROMACS . . . . .	30
<b>6. Conclusiones</b>	<b>32</b>
6.1. Proyectos a futuro . . . . .	32

# Introducción

Las simulaciones de dinámica molecular (DM) son una herramienta fundamental en el estudio de la función y la estructura de sistemas biológicos. Las simulaciones moleculares se consideran experimentos *in-silico* que aportan datos a nivel microscópico. Gracias al incremento en la capacidad de procesamiento del supercómputo científico y al desarrollo de paqueterías de simulación molecular[1, 2], ahora es posible simular sistemas biológicos de millones de átomos, como ribosomas[3] y cápsides virales[4]. No obstante, simular sistemas biológicos con técnicas de DM implica resolver las ecuaciones de movimiento para una gran cantidad de partículas, por lo que generalmente se emplean potenciales de interacción simples conocidos como campos de fuerza. La principal desventaja de los campos de fuerza es que éstos contienen una gran cantidad de parámetros ajustables, los cuales no siempre son transferibles a moléculas para las que no fueron parametrizados.

Actualmente, predecir de manera precisa la estructura nativa de las proteínas conociendo sólo su secuencia de aminoácidos es uno de los principales retos en el área de Biología Molecular. Las simulaciones moleculares son una de las principales herramientas para el estudio del plegamiento de proteínas. No obstante, explorar el espacio conformacional de una proteína es una tarea compleja debido al gran número de conformaciones que ésta puede adoptar, además es fácil que la proteína quede atrapada en mínimos locales durante una simulación molecular. Por esta razón, se han desarrollado diversos algoritmos de DM enfocados a mejorar la generación de ensamblajes conformacionales representativos[5, 6]. Por otro lado, el proceso de plegamiento se lleva a cabo en tiempos de microsegundos ( $\mu s$ ) a milisegundos ( $ms$ ) dependiendo del tamaño de la proteína, por lo tanto, observar el mecanismo de plegamiento requiere de simulaciones extensas con un alto costo computacional [7].

En este trabajo se pretende avanzar en el desarrollo de métodos más eficientes de dinámica molecular, a través del uso de potenciales externos para inducir el plegamiento de proteínas. Reduciendo de esta manera el tiempo computacional empleado en una simulación de DM. La implementación de un potencial externo en el método de dinámica molecular cumple con dos funciones: i) introducir efectos del medio en el proceso de plegamiento como el efecto hidrofóbico o la presencia de proteínas chaperonas; ii) acelerar la exploración del espacio conformacional para ahorrar tiempo computacional.

En trabajos previos se han obtenido resultados exitosos al implementar potenciales externos en el plegamiento de polímeros orgánicos modelo[8]. Como sistema de estudio en esta investigación, se eligió el péptido sintético *trp-cage*[9] de 20 residuos de aminoácidos y el fragmento (25-35) del péptido  $\beta$ -amiloide [10, 11]. Trabajar con miniproteínas nos permite emplear potenciales de interacción *ab-initio* con el método de funcionales de la densidad (DFT, por sus siglas en inglés). No obstante, el modelo de potenciales externos puede ser utilizado en conjunto con campos de fuerza para el estudio de proteínas de mayor tamaño.

# Antecedentes

## Plegamiento de proteínas

Las proteínas son la base de la maquinaria molecular que constituye a los seres vivos. Son polímeros de aminoácidos cuya secuencia es producto de la traducción del material genético celular. Conocer la estructura de las proteínas es fundamental ya que de ésta depende su funcionalidad. Las primeras estructuras de proteínas se obtuvieron en los años 50 con los estudios realizados por Max Perutz[12] y John Kendrew[13] con cristalografía de rayos-X. Actualmente existen grandes bases de datos como el Protein Data Bank (PDB)[14] que almacena más de 120,000 estructuras de proteínas. Sin embargo, existe un gran número de proteínas cuya estructura no ha sido determinada. Por esta razón, sería de gran utilidad encontrar una manera de predecir la estructura de las proteínas sin tener que analizarlas experimentalmente, sólo conociendo su secuencia de aminoácidos. Para alcanzar este objetivo se debe entender el mecanismo y los factores que dirigen el plegamiento de proteínas.

El plegamiento de proteínas es uno de los problemas más importantes sin resolver en el área de Biología y Biofísica Molecular[15]. El problema del plegamiento de proteínas consiste en resolver cómo la secuencia de aminoácidos dicta la estructura tridimensional de una proteína. Para su estudio se ha dividido en tres áreas, el problema termodinámico, el cinético y el computacional [16, 17].

El problema termodinámico consiste en comprender el balance energético que determina la estabilidad de una proteína en su estado nativo. La estructura nativa de una proteína es el resultado de la formación de interacciones atómicas en la proteína y su medio, las fuerzas interatómicas mantienen a la proteína en su conformación más estable. De alguna manera, la información de la estructura tridimensional de una proteína se encuentra codificada en su secuencia de aminoácidos, ya que ahí radica la diferencia entre una proteína y otra.

Diversas interacciones participan en el plegamiento de proteínas, como los puentes de hidrógeno, los puentes disulfuro y las interacciones electrostáticas. No obstante es difícil saber en qué medida contribuyen cada una de estas interacciones al plegamiento de las proteínas. La naturaleza anfífilica de las proteínas propicia la formación de estructuras organizadas donde los residuos apolares tienden a agruparse en el interior. Este fenómeno, conocido como efecto hidrofóbico, podría ser el principal factor del proceso de plegamiento de acuerdo con diversos estudios reportados en la literatura[18, 16].

El problema cinético consiste en entender el mecanismo de plegamiento de las proteínas. El mecanismo de plegamiento involucra un principio general el cual explica cómo la secuencia de aminoácidos y las condiciones del disolvente influyen en la trayectoria que sigue una proteína para alcanzar su estado nativo desde cualquier conformación. En 1969 Cyrus Levinthal[19] planteó la pregunta de cómo las proteínas pueden adoptar su conformación nativa tan rápido, algunas incluso en microsegundos, cuando explorar de manera puramente aleatoria todas las conformaciones posibles tomaría una enorme cantidad de años. Por lo que se concluyó que las proteínas no siguen una búsqueda aleatoria para llegar a su estado plegado sino que se presume un mecanismo de plegamiento[16].

Se han desarrollado modelos de mecánica estadística[20, 21] enfocados al estudio del espacio conformacional y al cálculo de la función de partición de proteínas. Se ha propuesto que la superficie de energía libre conformacional de una proteína solvatada, puede ser descrita con una función matemática  $F(\phi, \psi, X)$  que depende de sus grados de libertad. La representación gráfica de esta función es una superficie cuya forma es similar a la de un embudo, donde hay un número mucho menor de estructuras compactas y estables que de conformaciones desplegadas. El mínimo global de las superficies de energía corresponde al estado nativo de la proteína. No obstante, de manera local las superficies de energía pueden ser más complejas y presentar barreras cinéticas, intermediarios o mínimos locales como se muestra en la figura 2.1. Las superficies de energía conformacional, ilustran cómo el proceso de plegamiento puede iniciar desde distintos estados de alta energía, siguiendo distintas rutas conformacionales, para llegar a un único mínimo global.

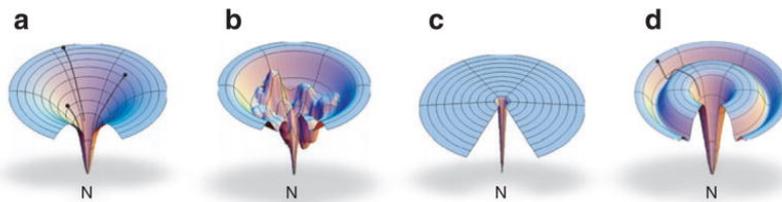


Figura 2.1: Superficies de energía libre conformacional  $F(\psi, \phi, \kappa)$ . (a) Superficie suave correspondiente a un plegamiento rápido, (b) superficie con trampas cinéticas, (c) superficie donde el plegamiento sigue un mecanismo difusivo, (d) superficie con un intermediario. Figura captada de referencia [16].

El problema computacional consiste en desarrollar algoritmos capaces de predecir la estructura nativa de una proteína, dada la secuencia de aminoácidos. La creación de bases de datos como el Protein Data Bank (PDB) ha sido fundamental para el desarrollo de diversos algoritmos bioinformáticos, los cuales se basan en la comparación de estructuras como los modelos de homología de secuencias, también llamados *template-based*[17]. Por otro lado, se han desarrollado algoritmos basados en modelos físicos como los de dinámica molecular y campos de fuerza. Sin embargo, el principal problema de los métodos físicos radica en la exploración del espacio conformacional. Se han propuesto estrategias donde se realizan de manera simultánea dinámicas con condiciones diferentes de temperatura[22], lo que permite explorar múltiples regiones del espacio conformacional. Así mismo, el método *accelerated molecular dynamics*[5] reduce artificialmente las barreras de potencial impidiendo permanecer en mínimos locales. Otra estrategia consiste en penalizar las conformaciones que han sido visitadas durante la simulación molecular, lo que se conoce como *metadynamics*[6].

Gracias al cómputo de alto rendimiento y a la optimización de algoritmos de DM, ha sido posible encontrar estructuras nativas de proteínas pequeñas. Sin embargo, los algoritmos de DM resultan ineficientes para predecir estructuras de proteínas de mayor tamaño. Así como los algoritmos bioinformáticos no han logrado predecir con buena precisión estructuras que no contengan secuencias homólogas en el PDB. Por lo que se han propuesto diversos métodos híbridos que combinan algoritmos bioinformáticos y físicos, entre los cuales destaca el método de ROSETTA[23], el cual ha obtenido muy buenos resultados en la competencia mundial de predicción de estructuras de proteínas CASP (Critical Assessment of protein Structure Prediction)[24].

## Dinámica molecular

La metodología de dinámica molecular consiste en simular computacionalmente la evolución temporal de sistemas moleculares dentro del marco de la mecánica clásica. En una dinámica molecular se calculan las trayectorias de los átomos, esto es, sus posiciones  $X$  y momentos  $P$  en el tiempo, generalmente en intervalos de 1 a 2 femtosegundos (fs). El paso de tiempo debe de ser un orden de magnitud menor que la frecuencia del movimiento con el periodo de oscilación más alto en el sistema, con un paso de tiempo mayor se pierde precisión en el proceso de integración y se pueden generar comportamientos caóticos [25, 26]. El cálculo de las trayectorias en una DM se realiza a través de la integración numérica de las ecuaciones de movimiento dadas por la segunda ley de Newton.

Para un sistema de  $n$  partículas en el espacio físico tridimensional, con un sistema de coordenadas cartesianas (x,y,z), la fuerza que actúa sobre cada partícula  $i$  será el vector  $\vec{F}_i(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N, t)$  con componentes:  $F_{x_i}, F_{y_i}, F_{z_i}$ , la cual depende de las posiciones de las  $n$  partículas. En el sistema cartesiano de coordenadas, las ecuaciones de movimiento se pueden expresar en componentes como se muestra en la ecuación 2.1.

$$\begin{aligned} F_{x_i}(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N, t) &= m \frac{d^2 x_i}{dt^2}, \\ F_{y_i}(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N, t) &= m \frac{d^2 y_i}{dt^2}, \quad i = 1, 2, \dots, n \\ F_{z_i}(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N, t) &= m \frac{d^2 z_i}{dt^2} \end{aligned} \quad (2.1)$$

Dada una conformación inicial, por ejemplo las coordenadas cartesianas obtenidas con un experimento de cristalografía de rayos X de una proteína, las velocidades iniciales se asignan aleatoriamente a cada átomo, satisfaciendo una distribución de Maxwell-Boltzmann con un valor de temperatura asignado. Conociendo las posiciones y velocidades iniciales sólo resta conocer la fuerza que actúa en cada átomo para calcular la posición en el siguiente instante de tiempo  $t_0 + dt$ .

Cuando un sistema es conservativo la fuerza sobre cada partícula se puede obtener a partir del gradiente del potencial como se muestra en la ecuación 2.2. Un sistema es conservativo cuando su energía total se conserva, esto es cuando la suma de energía cinética  $T$  y potencial  $U$  es constante  $T + U = cte$  [27].

$$\vec{F}_i(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_n, t) = -\nabla_i U(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_n, t) \quad (2.2)$$

La energía potencial  $U$  en un sistema molecular se puede obtener de cálculos de mecánica cuántica o bien, a través de modelos clásicos de potenciales de interacción. Estos últimos se emplean frecuentemente en la simulación de sistemas biomoleculares como proteínas, membranas biológicas y ADN, ya que por la gran cantidad de átomos involucrados y la complejidad de los algoritmos que conforman la teoría cuántica, no es posible resolver la ecuación de Schrödinger para este tipo de moléculas. Los potenciales de interacción clásicos, mejor conocidos como campos de fuerza, representan una manera simple de calcular las fuerzas interatómicas. No obstante, en su forma funcional involucran varios parámetros que deben ser ajustados para poder obtener resultados precisos.

# Objetivo e hipótesis

El objetivo principal de este proyecto es avanzar en el desarrollo de métodos más eficientes de dinámica molecular, a través del uso de potenciales externos para inducir el plegamiento de proteínas. Reduciendo de esta manera el tiempo computacional empleado en la exploración del espacio conformacional durante una simulación computacional.

## Objetivos particulares

- Proponer y evaluar diferentes funciones de potencial para inducir el plegamiento de proteínas.
- Encontrar los parámetros de las funciones de potencial que optimicen la búsqueda de estructuras plegadas y reduzcan el tiempo de simulación.
- Evaluar el uso de potenciales externos en simulaciones de proteínas con y sin disolvente explícito.

## Hipótesis

Se sabe que el plegamiento de proteínas se lleva a cabo en tiempos del orden de los microsegundos a milisegundos y que depende de diversos factores como la secuencia de aminoácidos, las interacciones hidrofóbicas, la presencia de proteínas chaperonas, el balance entálpico-entrópico, entre otros. Diseñar un método preciso que tome en cuenta todos los factores que intervienen en el proceso de plegamiento, y sea capaz de alcanzar estructuras plegadas en tiempos de cómputo accesibles es sumamente complejo. Los potenciales externos representan una aproximación relativamente directa y simple para modelar los efectos que diversos factores tienen en el proceso del plegamiento de proteínas, por lo que su implementación ayudará a encontrar estructuras plegadas de forma eficiente.

# Modelo y metodología

## Modelo teórico

En este trabajo se aplicó un método de muestreo para la búsqueda de conformaciones de mínima energía de proteínas. El modelo propuesto integra métodos de dinámica molecular dentro del marco de la teoría de Langevin con la introducción de potenciales externos para dirigir el plegamiento de proteínas. La aproximación de Langevin permite equilibrar el sistema a una temperatura dada, evitando de esta manera el uso de termostatos que generalmente se emplean en simulaciones de DM, los cuales introducen partículas ficticias en las ecuaciones de movimiento con parámetros ajustados adicionales.

El modelo teórico puede aplicarse a simulaciones de proteínas donde el efecto del disolvente en la proteína se considera de manera implícita en los términos estocásticos de la ecuación de Langevin, los cuales se describirán más adelante. Por otro lado, el modelo puede aplicarse a una proteína en una celda de disolvente explícito, donde los átomos que se encuentran en la frontera de la celda siguen una dinámica de tipo Langevin, mientras que los que se encuentran en el interior de la celda siguen una dinámica Newtoniana. En ambos casos se pueden introducir potenciales externos para dirigir el plegamiento de las proteínas. En esta sección se describe de manera breve el método desarrollado por nuestro grupo, una discusión más detallada puede consultarse en las referencias [28] y [8].

## Modelo molecular con disolvente implícito

### Lagrangiano

El modelo teórico propuesto puede considerarse una generalización de la aproximación de Langevin-Zwanzig (ver referencia [29]) para  $n$  partículas, con la adición de un término extra en el Lagrangiano correspondiente al potencial externo, como se muestra en la ecuación 4.1. Suponemos que el sistema se encuentra inmerso en un baño térmico compuesto por una colección de osciladores armónicos.

$$\begin{aligned} \mathcal{L}(\{q_{ij}, x_i\}) = & \sum_i \frac{m_i}{2} \dot{x}_i^2 + \sum_i \sum_j \frac{m_{ij}}{2} \dot{q}_{ij}^2 - U(\{x_i\}) - V_p(\{x_i\}) \\ & - \sum_i \sum_j \left[ \frac{m_{ij} w_{ij}^2}{2} q_{ij}^2 - c_{ij} q_{ij} x_i + \frac{c_{ij}^2}{2m_{ij} w_{ij}^2} x_i^2 \right] \end{aligned} \quad (4.1)$$

El Lagrangiano (ecuación 4.1) depende de las coordenadas generalizadas de las partículas del sistema y de sus velocidades,  $\{x_i, \dot{x}_i\}$ , y de las partículas del baño térmico,  $\{q_{ij}, \dot{q}_{ij}\}$ . El doble índice de las partículas del baño térmico indica que cada partícula  $i$  del sistema interactúa con un baño térmico independiente. Los primeros dos términos del Lagrangiano representan la energía cinética de las partículas del sistema y del baño térmico. El potencial de interacción entre partículas del sistema  $U(\{x_i\})$ , se calcula resolviendo la ecuación de Schrödinger. El potencial externo está descrito por la cantidad  $V_p(\{x_i\})$ , cuya forma funcional es la principal aportación de este trabajo. La interacción del sistema con el baño térmico está descrita con una aproximación armónica correspondiente al último término del Lagrangiano, donde  $m_{ij}$  y  $w_{ij}$  son la masa y la vibración respectivamente de la partícula  $ij$ . La interacción entre las partículas del sistema y las del baño térmico es de forma bilineal, donde  $c_{ij}$  son coeficientes de acoplamiento y están relacionados con el coeficiente de fricción. Las partículas del baño térmico no interactúan entre sí.

## Ecuaciones de movimiento

Las ecuaciones de movimiento para una partícula con posición y velocidad generalizadas,  $(\zeta, \dot{\zeta})$ , se obtienen resolviendo la ecuación de Euler-Lagrange (ecuación 4.2).

$$\frac{\partial \mathcal{L}}{\partial \zeta} - \frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{\zeta}} = 0 \quad (4.2)$$

Al aplicar la ecuación de Euler-Lagrange al Lagrangiano (ecuación 4.1) se obtienen las siguientes ecuaciones para las partículas del sistema (ecuación 4.3) y del baño térmico (ecuación 4.4):

$$m_i \ddot{x}_i = \sum_j \left( c_{ij} q_{ij} - \frac{c_{ij}^2}{m_{ij} w_{ij}^2} x_i \right) - \frac{\partial(U + V_p)}{\partial x_i} \quad (4.3)$$

$$m_{ij} \ddot{q}_{ij} = c_{ij} x_{ij} - m_{ij} w_{ik}^2 q_{ij} \quad (4.4)$$

Se observa en las expresiones anteriores que las ecuaciones de movimiento de las partículas del sistema y del baño término se encuentran acopladas. Esto significa que la trayectoria  $x_i$  de la partícula  $i$  depende de las coordenadas de las  $n$  partículas del baño térmico  $\{q_{ij}\}_{j=1}^n$ . A su vez, la trayectoria de las partículas del baño térmico depende de la coordenada  $x_i$  de la partícula  $i$ .

Las ecuaciones de movimiento se pueden desacoplar resolviendo la ecuación diferencial de las partículas del baño térmico (ecuación 4.4) y sustituyendo en la ecuación de las partículas del sistema (ecuación 4.3). La solución es una ecuación de movimiento tipo Langevin como la que se muestra en la expresión 4.5. Los detalles para obtener la ecuación de movimiento se pueden consultar en las referencias [28, 29, 8].

$$m_i \ddot{x}_i = - \frac{\partial[U(\{x_i\}) + V_p(\{x_i\})]}{\partial x_i} + G_i(t) - m_i \xi_i \dot{x}_i \quad (4.5)$$

El primer término del lado derecho de la ecuación 4.5 representa la fuerza que actúa sobre la partícula  $i$  debida al potencial de interacción entre partículas  $U(\{x_i\})$  y al potencial externo  $V_p(\{x_i\})$ . El segundo término corresponde a una fuerza de origen estocástico ejercida por las partículas del baño térmico sobre la partícula  $i$ , esta fuerza introduce las fluctuaciones térmicas del medio. El último término es una fuerza de origen disipativa y representa la viscosidad del medio, donde  $\xi_i$  es el coeficiente de fricción local. Los últimos dos términos simplifican la ecuación de movimiento y favorecen el intercambio de energía del sistema con el medio, no obstante, convierten a la ecuación de movimiento irreversible en el tiempo y a la energía total del sistema no conservativa. En la figura 4.1 se ilustra una proteína en el baño térmico.

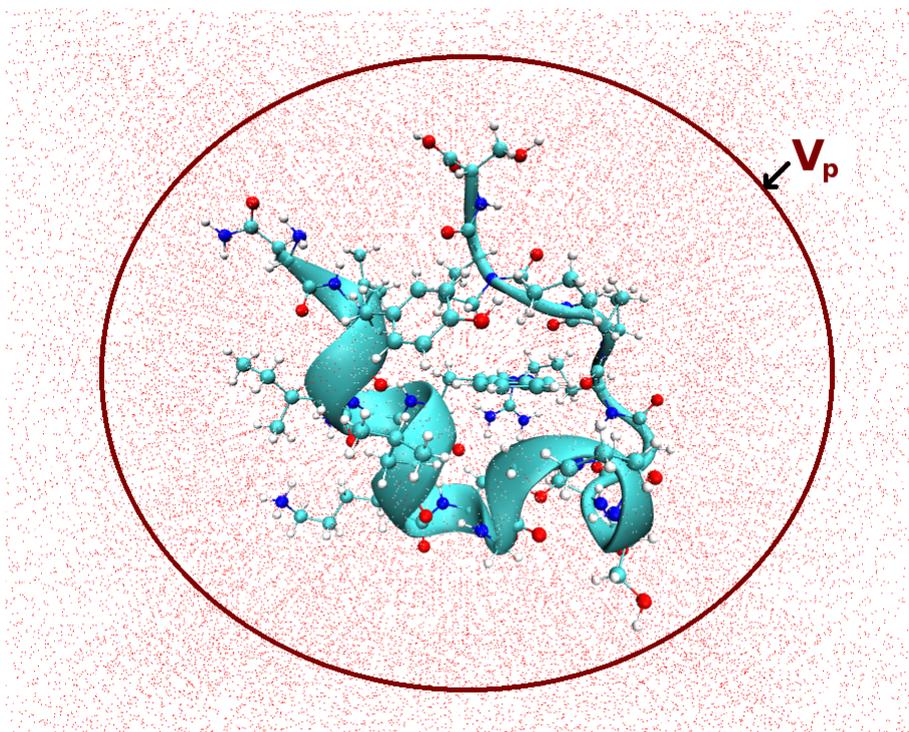


Figura 4.1: Péptido trp-cage en un baño de disolvente implícito. El baño térmico se representa con una malla roja punteada. El potencial externo se representa con el círculo rojo.

## Modelo molecular con disolvente explícito

Este modelo puede aplicarse a sistemas en una celda de disolvente explícito, para lo cual se integra un potencial adicional de confinamiento  $V_c(\{x_i\})$ . La función del potencial  $V_c(\{x_i\})$  es mantener a las moléculas de disolvente y a la proteína dentro de la celda de manera análoga a un contenedor. Este potencial es de tipo exponencial y decae rápidamente cuando la distancia entre los átomos y la frontera de la celda aumenta. Por lo tanto, la fuerza que el potencial  $V_c(\{x_i\})$  ejerce en los átomos que se encuentran en el interior de la celda es despreciable.

En la figura 4.2 se ilustra la celda de disolvente. Los átomos que se encuentran en la frontera de la celda son los únicos que interactúan con el baño térmico y por lo tanto son los únicos que siguen una dinámica de tipo Langevin. Los átomos que se encuentran en el interior de la celda siguen una dinámica Newtoniana.

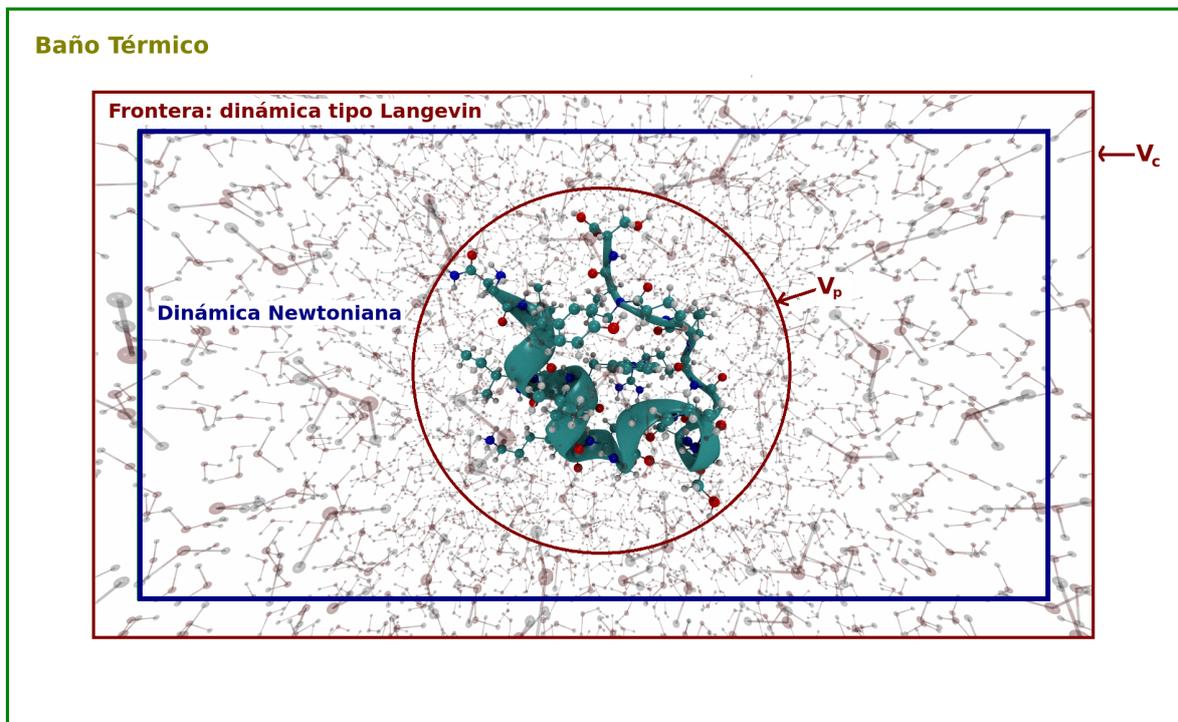


Figura 4.2: Péptido trp-cage en celda de disolvente. Los átomos en la frontera siguen una dinámica de Langevin con potenciales externos. Los átomos al interior de la celda siguen una dinámica Newtoniana con potenciales externos. Los potenciales de confinamiento  $V_c$  y de plegamiento  $V_p$  se ilustran en rojo.

La celda de disolvente es lo suficientemente grande para que la proteína no interactue con las fronteras, además el potencial externo la mantiene en el centro de la celda durante la simulación. El potencial externo o de plegamiento  $V_p(\{x_i\})$ , actúa únicamente sobre los átomos de la proteína. El potencial de confinamiento  $V_c(\{x_i\})$  actúa sobre todos los átomos en la celda. Por lo tanto, las ecuaciones de movimiento para la proteína y las moléculas de disolvente en cada región de la celda están dadas por las ecuaciones que se muestran en la figura 4.3.

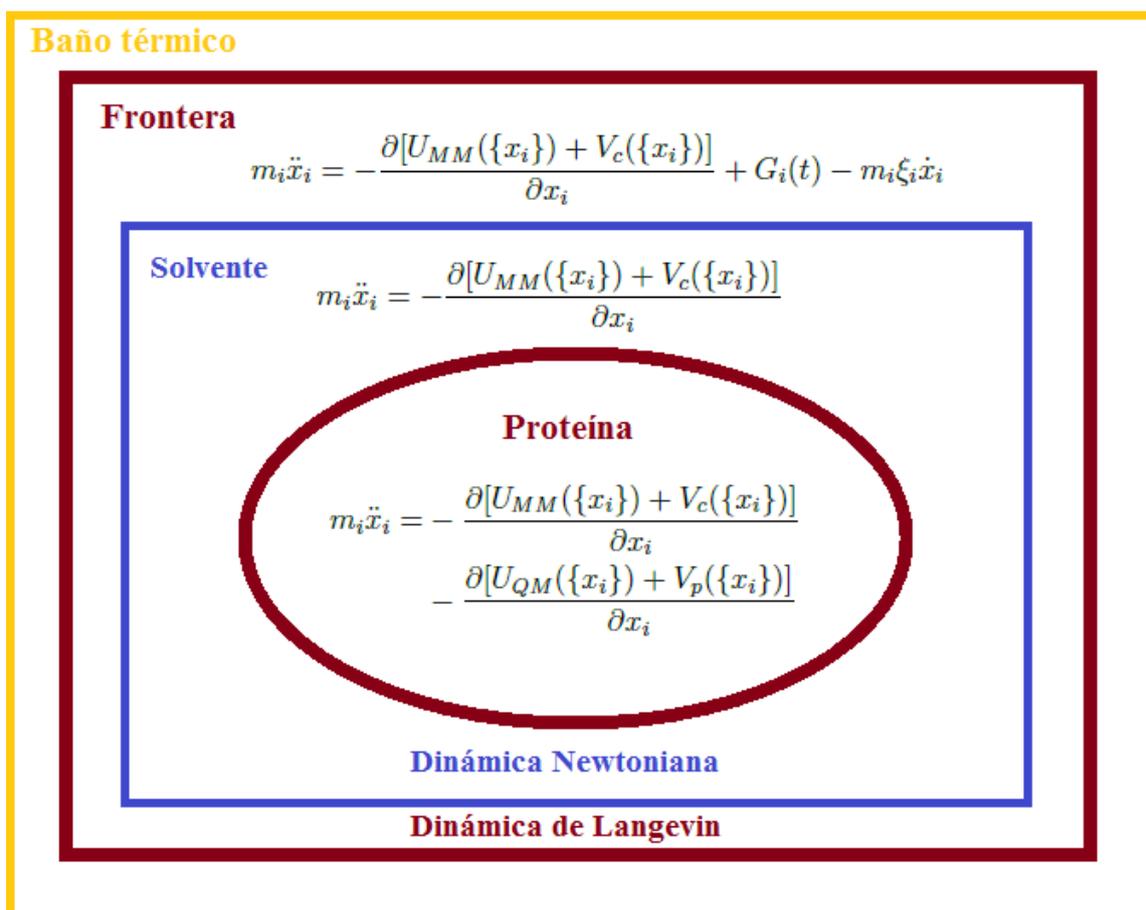


Figura 4.3: Ecuaciones de movimiento para cada región de la celda. La ecuación en el elipse corresponde a la ecuación de movimiento de los átomos de la proteína. Las ecuaciones fuera del elipse corresponden a las ecuaciones de movimiento de los átomos del disolvente en la frontera (ecuación de tipo Langevin) y en el interior de la celda (ecuación Newtoniana).

En las ecuaciones de la figura 4.3 se distinguen dos potenciales de interacción entre partículas, el potencial  $U_{MM}$  y el potencial  $U_{QM}$ , esto se debe a que el número de moléculas de agua necesarias para solvatar a una proteína es muy grande y no es posible evaluar el potencial del disolvente con cálculos *ab-initio*, por esta razón se recurre al uso de potenciales clásicos para el disolvente. No obstante, el potencial de interacción entre los átomos de la proteína  $U_{QM}(\{x_i\})$  es de naturaleza cuántica, se obtiene de resolver la ecuación de Schrödinger independiente del tiempo bajo la aproximación Born-Oppenheimer (ver ecuación 4.6), teniendo de esta manera simulaciones del tipo mecánica cuántica-mecánica molecular (QM/MM por sus siglas en inglés).

$$\hat{H}\psi(\{x_i\}) = E\psi(\{x_i\}); \quad U_{QM}(\{x_i(t)\}) = E|_{\{x_i(t)\}} \quad (4.6)$$

El potencial clásico  $U(\{x_i\})$  representa la interacción entre las moléculas del disolvente así como la interacción del disolvente con la proteína. Existen diversos modelos de disolvente explícito, no obstante la mayoría contienen los términos de interacción Coulómbica y de Lennard-Jones. El potencial de interacción entre la proteína y el disolvente se calcula también con los potenciales clásicos de Coulomb y de Lennard-Jones cuyos parámetros se pueden obtener de cualquier campo de fuerza para proteínas. Los campos de fuerza empleados se describen en la sección de metodología.

## Potenciales Externos

El potencial externo  $V_p(x_i)$  puede adoptar distintas formas funcionales. En este trabajo se proponen tres formas funcionales simples para inducir el plegamiento. El potencial externo puede actuar sobre todos los átomos de la proteína o bien, sobre un conjunto de ellos, como el conjunto de átomos que conforman el esqueleto peptídico, los carbonos alfa o los átomos de las cadenas laterales únicamente. Las formas funcionales propuestas se muestran a continuación.

$$V_h(\{x_i\}) = \sum_i \frac{1}{2}kr_i^2 \quad (4.7)$$

$$V_g(\{x_i\}) = -V_0 \sum_i e^{-r_i^2/(2\sigma^2)} \quad (4.8)$$

$$V_s(\{x_i\}) = \sum_i e^{\alpha(r_i-R)} \quad (4.9)$$

Donde,

$$r_i = [(x_i - x_o)^2 + (y_i - y_o)^2 + (z_i - z_o)^2]^{1/2} \quad (4.10)$$

La primera expresión (ecuación 4.7) es un potencial armónico, donde  $k$  es la constante de fuerza del oscilador. La siguiente expresión (ecuación 4.8) es un potencial de tipo Gaussiano atractivo, donde los parámetros  $V_0$  y  $\sigma$  son el peso de la función y la desviación estándar. En la última expresión (ecuación 4.9) se emplea una función exponencial con un factor de normalización que le da un carácter esférico al potencial, donde  $R$  representa el radio de la esfera. Los parámetros  $(x_o, y_o, z_o)$  en la ecuación 4.10 son el origen del potencial. Los valores de los parámetros empleados en las diferentes funciones de potencial se calcularon de manera que la fuerza inducida por el potencial externo no superara las fuerzas de enlaces y que se redujera el tiempo de simulación.

## Metodología

Durante este proyecto se realizaron múltiples simulaciones para evaluar el uso de potenciales externos en dinámicas moleculares tipo QM/MM de proteínas. Este trabajo se dividió en cuatro etapas. Las cuales se enlistan a continuación.

**Primera etapa.** Durante la primera etapa, se realizaron simulaciones con dos péptidos distintos: i) una proteína sintética con estructura secundaria y terciaria de 20 residuos de aminoácidos, diseñada especialmente para optimizar su plegamiento[9]. Este tipo de estructura plegada se ha denominado como *motivo trp-cage*, por lo que se usará este nombre para referirnos a esta proteína. ii) El fragmento A $\beta$ (25-35) de 11 residuos del péptido  $\beta$ -amiloide, con estructura secundaria de hélice  $\alpha$ [10, 11]. Las primeras simulaciones con ambos péptidos se realizaron con el objetivo de encontrar los parámetros óptimos de las funciones del potencial externo. Los parámetros no han sido evaluados con otros péptidos.

**Segunda etapa.** En una segunda etapa, se realizaron simulaciones para evaluar la aplicación de los potenciales externos en la búsqueda de estados plegados de ambos péptidos: motivo trp-cage y el fragmento (25-25) $\beta$ -amiloide. Posteriormente, se realizaron simulaciones de relajación en ausencia del potencial externo. El objetivo de estas simulaciones fue permitir que la proteína adoptara una conformación menos estresada, dada únicamente por las fuerzas de interacción entre los átomos de la proteína y por el efecto del baño térmico. Las simulaciones de relajación se realizaron con el modelo molecular descrito en la sección anterior pero sin la contribución del potencial externo  $V(\{x_i\})$ .

**Tercera etapa.** Dado que el modelo de Langevin no toma en cuenta las interacciones de manera individual entre las moléculas de agua con la proteína, y por ende no se consideran los efectos electrostáticos ni de Van der Waals, los cuales pueden ser fundamentales en el proceso de plegamiento de proteínas, en la tercera etapa de este trabajo se realizaron simulaciones con la proteína en una celda de disolvente explícito. Para cumplir con este objetivo, se implementó en el método el modelo de tres sitios TIP3P desarrollado por Jorgensen et. al.[30] con moléculas de agua flexibles[31]. Las interacciones entre las moléculas de la proteína con las moléculas de agua explícitas se calcularon empleando el campo de fuerza OPLS-AA[32]. No obstante, las cargas de los átomos de las proteínas se promediaron con las cargas de Mulliken obtenidas con el cálculo de estructura electrónica en cada paso de tiempo (ver ecuación 4.11).

$$\bar{q}_{atom}(t) = \frac{1}{2} (q_{OPLS-AA} + q_{DFT}(t)) \quad (4.11)$$

**Cuarta etapa.** Con fines comparativos, en la última etapa de este trabajo se realizó una simulación de mecánica molecular por 20 ns de la proteína trp-cage con disolvente explícito a temperatura y volumen constante (ensamble canónico NVT). En esta simulación se empleó el campo de fuerza OPLS-AA [32], el modelo de disolvente explícito TIP3P[30] y el algoritmo de acoplamiento de temperatura de Nosé-Hoover. La integración de las ecuaciones de movimiento se realizó con el algoritmo *Velocity-Verlet*[33]. Esta simulación se realizó con la paquetería de simulación molecular GROMACS[1].

La integración de las ecuaciones de movimiento de tipo Langevin se realizó con una variante del algoritmo de Verlet, el cual se puede consultar en las referencias [34, 35, 28]. Así mismo, el paso de tiempo en todas las simulaciones fue de 1 *fs*. El software para resolver las ecuaciones de movimiento se programó en FORTRAN. El código es libre y se puede obtener en [36].

Se eligió el método de teoría funcionales de la densidad, DFT (por sus siglas en inglés), para la resolución de la ecuación de Schrödinger, de la que se obtuvo el potencial de interacción de las proteínas, en la aproximación de gradiente generalizado. El método de DFT tiene la ventaja de tener una buena precisión y un bajo costo computacional comparado con métodos de función de onda, como los métodos post-Hartree-Fock. Se emplearon los funcionales de Becke[37] y Lee[38] de intercambio y correlación respectivamente con un funcional de dispersión[39]. Se usó como base el conjunto de Gaussianas 6-31G. Las ecuaciones monoeléctricas de Kohn-Sham se resolvieron de manera autoconsistente con un criterio de convergencia de  $10^{-6}$  ua en la matriz de densidad y energía. Los cálculos de estructura electrónica se realizaron empleando el programa TeraChem[40] ya que se encuentra implementado para su funcionamiento en plataformas de tipo GPU.

# Discusión de resultados

Un fragmento de 11 residuos del péptido intrínsecamente desordenado  $\beta$ -amiloide, así como un péptido de 20 aminoácidos con estructura secundaria y terciaria, fueron empleados para probar el efecto de los potenciales externos en la dinámica del plegamiento de proteínas. Se produjeron estructuras desplegadas de ambos péptidos, aplicando fuerzas de osciladores armónicos (ecuación 4.7) en los extremos de las proteínas bajo un esquema *steered molecular dynamics* (SMD). Todas las simulaciones realizadas posteriormente parten de las estructuras desplegadas que se muestran en la figura 5.1.

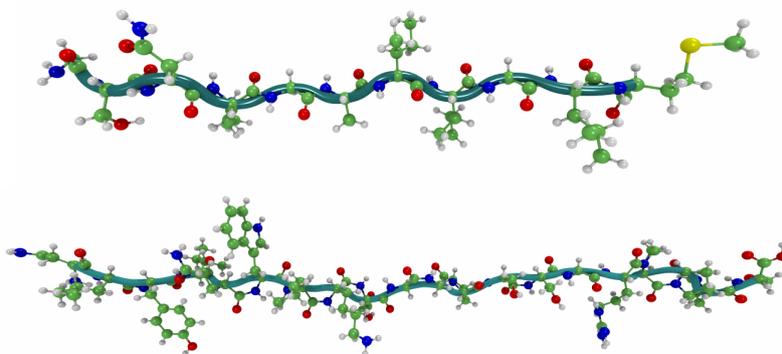


Figura 5.1: Estructuras iniciales. Arriba: Estructura desplegada del fragmento 25-35 del péptido  $\beta$ -amiloide, con secuencia de aminoácidos GSNKGAIIGLM. Abajo: Estructura desplegada del péptido trp-cage con secuencia NLYIQWLKDGGPSSGRPPPS.

## Primera etapa: parametrización

Se realizaron simulaciones de 10 (ps) con las dos proteínas de estudio para determinar el valor de los parámetros de las funciones del potencial externo. La función del potencial armónico (ver ecuación 4.7) depende de la constante de fuerza  $k$ , esta constante puede cambiar en el tiempo con una rapidez  $v_k = \Delta k / \Delta t$  como se muestra en la expresión 5.1, lo cual resulta útil cuando se desea mantener la fuerza constante.

$$k_i = v\Delta t + k_{i-1} \quad (5.1)$$

Donde  $i = 1, 2, \dots, n$  representa el número de iteración en el tiempo, siendo el tiempo total de la simulación  $\tau = n\Delta t$ . Por otro lado, la función Gaussiana (ecuación 4.8) y la función exponencial (ecuación 4.9) dependen de las amplitudes  $V_0$  y  $A = e^{-\alpha R}$  respectivamente. La fuerza externa que se ejerce sobre las partículas del sistema es directamente proporcional a estas constantes. En la función de potencial esférico el parámetro  $R$  aumenta  $\beta$  veces cada  $n$  pasos de tiempo, de acuerdo con la siguiente expresión.

$$R_i = \beta R_{i-1}; \quad i = n\Delta t \quad (5.2)$$

El origen del potencial  $\vec{o} = (x_o, y_o, z_o)$  en las tres funciones es arbitrario, no obstante, en el potencial esférico se fijó al centro de masa de la proteína  $\vec{r}_{CM}$ . Los parámetros obtenidos para cada función se muestran en la tabla 5.1.

Tabla 5.1: Parámetros para cada función de potencial externo

Potencial	Parámetros
<b>Armónico</b> $V_h(\{x_i\}) = \sum_i \frac{1}{2} k r_i^2$	$k = [0.01, 0.5]$ (N/cm) $v = 0.02$ Ncm <sup>-1</sup> ps <sup>-1</sup> ; $k_0 = 0.01$ N/cm
<b>Gaussiano</b> $V_g(\{x_i\}) = -V_0 \sum_i e^{-r_i^2/(2\sigma^2)}$	<sup>a</sup> $\sigma = 9.0$ Å $V_0 = 0.1$ Ha
<b>Esférico</b> $V_s(\{x_i\}) = \sum_i e^{\alpha(r_i-R)}$	<sup>b</sup> $\alpha = 8.0$ 1/Å $\beta = 0.98$ $n = 100$ fs

<sup>a</sup> El parámetro  $\sigma$  depende del tamaño de la proteína. Una buena aproximación para este parámetro es el radio de giro de la proteína  $\sigma \approx r_g$ .

<sup>b</sup> El valor del parámetro  $\alpha$  se determinó en trabajos previos [28].

Los valores reportados en la tabla 5.1 se eligieron bajo dos criterios: i) que la fuerza ejercida por el potencial fuera menor a las fuerzas de interacción atómicas y ii) que las proteínas adoptaran estructuras plegadas en un tiempo menor a 10 ps. Los potenciales externos pueden actuar sobre todos los átomos del sistema o sólo sobre un grupo de ellos. En este trabajo se exploraron tres casos: en el primero se aplicó un potencial armónico a cada carbono  $\alpha$  del péptido A $\beta$ ; en el segundo se aplicaron los potenciales Gaussiano y esférico a todos los átomos pesados del péptido trp-cage; finalmente se aplicó el potencial esférico a los átomos pesados de las cadenas laterales de los residuos hidrofóbicos del péptido trp-cage. Al aplicar el potencial sobre un grupo de átomos se puede dirigir el plegamiento del péptido sin restringir todos los grados de libertad.

## Segunda etapa: péptido $\beta$ -amiloide en disolvente implícito

La formación de elementos de estructura secundaria en las simulaciones de dinámica molecular puede tomar varios nanosegundos. Se ha observado a través de la comparación entre estructuras de proteínas reportadas, que las secuencias locales similares adoptan arreglos tridimensionales similares, esto ocurre principalmente con las hélices ya que son arreglos de una sola región de la proteína, a diferencia de las láminas  $\beta$  que pueden abarcar diferentes regiones de una cadena polipeptídica. Podría ser de utilidad emplear potenciales externos para inducir la formación de elementos de estructura secundaria, que actúen sobre secuencias locales que puedan formar estructuras como hélices  $\alpha$ .

Para probar esta hipótesis, se realizó una simulación de 25 ps del fragmento 25-35 del péptido  $A\beta$  con estructura de hélice  $\alpha$ . Se dirigió al péptido desplegado hasta su estructura experimental aplicando un potencial armónico sobre cada carbono  $\alpha$ . El origen de cada potencial se fijó en las posiciones de los carbonos  $\alpha$  correspondientes de la estructura experimental. La estructura del fragmento 25-35 del péptido  $A\beta$  ha sido caracterizado con RMN por D'Ursi y colaboradores[11], ésta se encuentra reportada en el *Protein Data Bank* con clave 1QWP (ver figura 5.2).

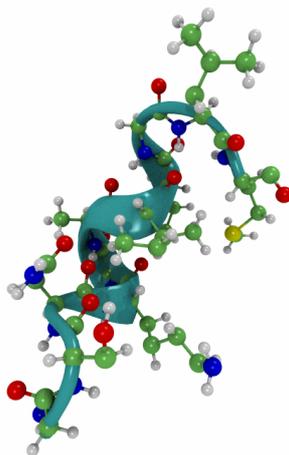


Figura 5.2: Estructuras experimental del fragmento 25-35 del péptido  $\beta$ -amiloide en agua.

Posteriormente se realizó una simulación por 10 ps en ausencia de potenciales externos para permitir al péptido adoptar una conformación natural y obtener datos estructurales. Las trayectorias obtenidas de las simulaciones de plegamiento y de relajación se analizaron a través del cálculo de la desviación cuadrática media (RMSD por sus siglas en inglés) con la estructura experimental como referencia, la desviación cuadrática media de las dos simulaciones se muestra en las figuras 5.3 y 5.4.

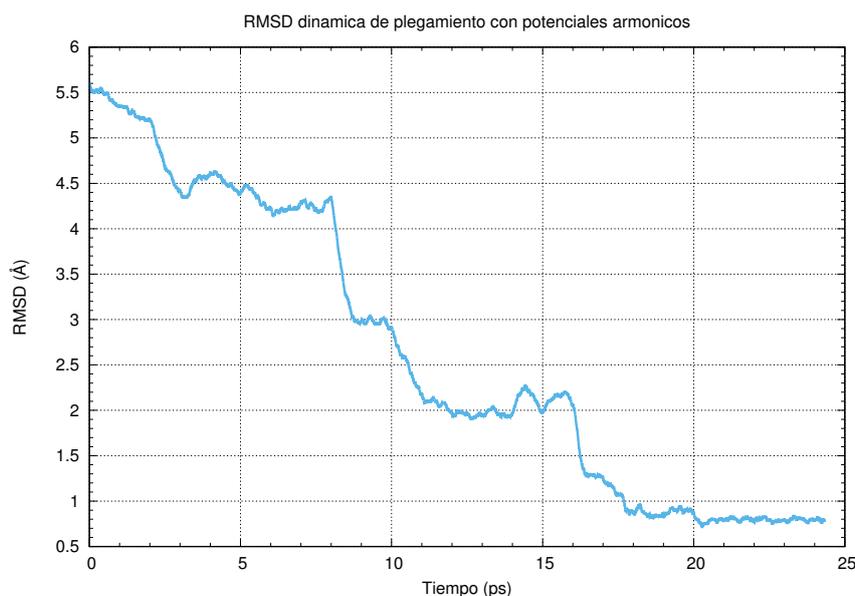


Figura 5.3: RMSD de la simulación de plegamiento del péptido  $\beta$ -amiloide en disolvente implícito con potenciales armónicos.

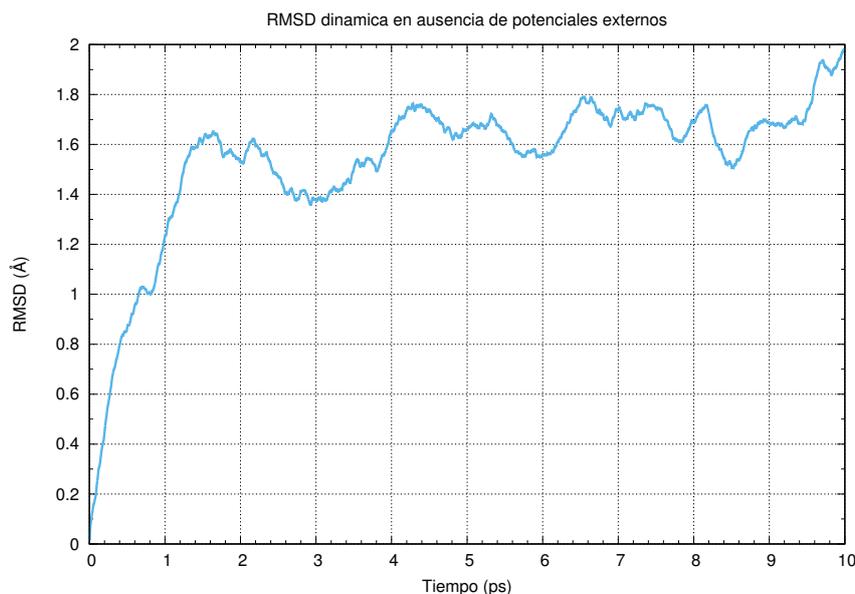


Figura 5.4: RMSD de la simulación del péptido  $\beta$ -amiloide en disolvente implícito sin potenciales externos.

El análisis de RMSD se aplicó únicamente a los esqueletos de los péptidos. Se aprecia en las figuras 5.3 y 5.4 que se alcanza un RMSD de 0.7 Å (RMSD ideal  $\leq 1$  Å) en la simulación con los potenciales armónicos. Esto indica que se llegó a una estructura muy cercana a la estructura experimental. No obstante, en la simulación de relajación se puede observar que el RMSD aumenta hasta 2 Å. Las estructuras finales de cada simulación se muestran en la figura 5.5.



Figura 5.5: Izquierda: Estructura obtenida con la aplicación de potenciales de oscilador armónico alineada con la estructura experimental. Derecha: Estructura final de la simulación de relajación comparada con la estructura experimental. La estructura experimental se representa en ambas figuras en transparencia.

También se llevó a cabo un análisis de la estructura secundaria de las conformaciones obtenidas, donde se determinó que no se reprodujo la estructura experimental de hélice  $\alpha$ , ya que los ángulos de torsión no son los adecuados. Esto se puede deber a que la fuerza ejercida por los potenciales armónicos impide que la cadena rote para formar los giros de una hélice  $\alpha$ . Mientras que la estructura experimental contiene un 36.4% de hélice  $\alpha$  y un 18.2% de hélice  $3_{10}$ . Las estructuras obtenidas en las simulaciones no muestran elementos de hélice sino giros  $\beta$ . El contenido de estructura secundaria se calculó con el programa PDBsum.

Durante la simulación de relajación la proteína alcanza una estabilidad estructural lo cual se ve reflejado en la convergencia del RMSD. La energía potencial DFT de las dinámicas de plegamiento y relajación se presenta en las figuras 5.6 y 5.7.

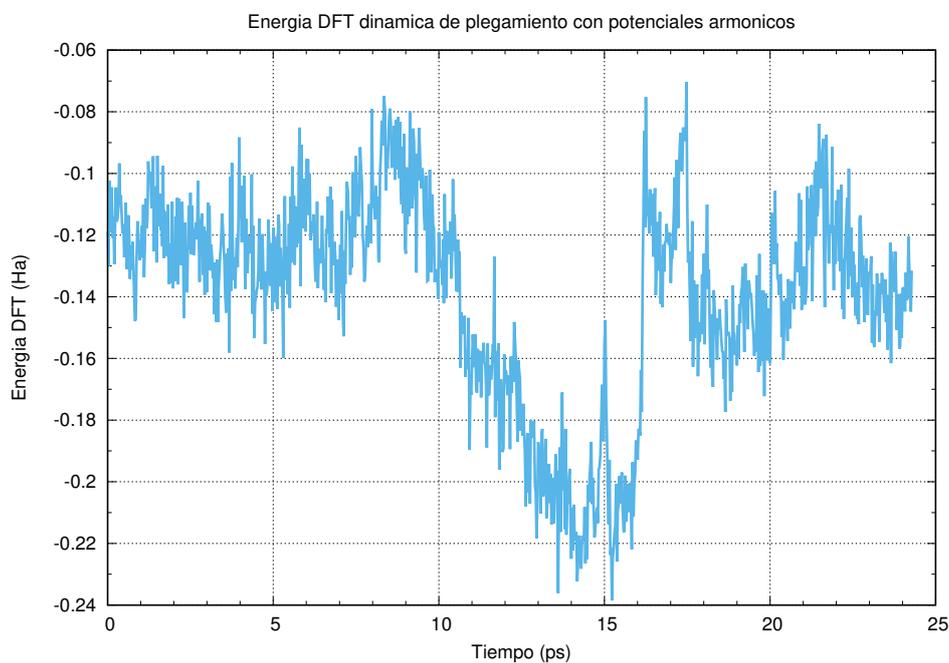


Figura 5.6: Energía potencial del péptido  $A\beta$ . Dinámica de plegamiento. La energía fue calculada en cada paso de tiempo con TeraChem empleando el método DFT (ver metodología).

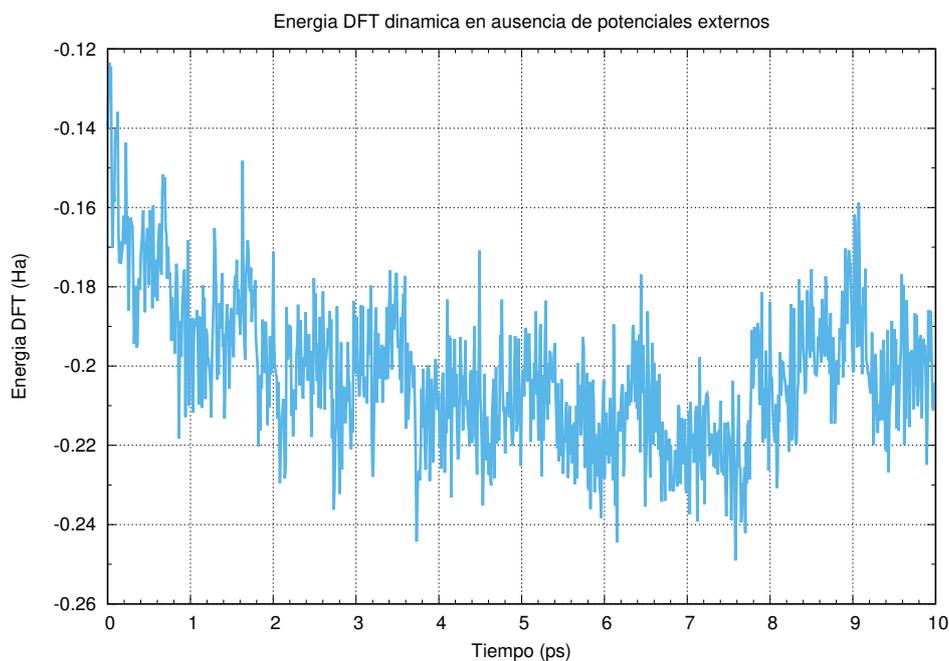


Figura 5.7: Energía potencial del péptido  $A\beta$ . Dinámica de relajación. La energía fue calculada en cada paso de tiempo con TeraChem empleando el método DFT (ver metodología).

Se observa en las figuras 5.6 y 5.7 que durante la simulación de plegamiento la proteína experimenta cambios en la energía potencial mayores que en ausencia de potenciales externos. Esto sugiere que el potencial externo ayuda al péptido a alcanzar conformaciones de menor energía en periodos de tiempo cortos. En este caso el péptido alcanza una conformación de baja energía en 15 ps, no obstante a partir de los 15 ps se observa un incremento en la energía, posiblemente debido a que la fuerza del potencial externo mantiene a la proteína en un estado compacto donde la repulsión electrónica aumenta. En la dinámica de relajación la energía DFT converge a un valor de -0.213 Ha. La energía en ambas gráficas es la diferencia de energía DFT en cada pasó de tiempo menos la energía de referencia, que corresponde a la energía DFT del estado desplegado inicial,  $\Delta E_{DFT}(t) = E_{DFT}(t) - E_{ref}$ .

## Segunda etapa: péptido trp-cage en disolvente implícito

El péptido trp-cage presenta elementos conservados de estructura secundaria y terciaria. Este péptido está constituido por 20 residuos de aminoácidos con secuencia 1NL-YIQWLKDGGPSSGRPPPS20. Su estructura consiste en una hélice  $\alpha$  del residuo Leu 2 a Lys 8, una pequeña hélice  $3_{10}$  de los residuos 10-13 y un núcleo hidrofóbico compuesto por tres residuos de prolinas (pro 12, pro 18 y pro 19) y los residuos aromáticos Tyr 3 y Trp 6. Este péptido se sintetizó y caracterizó a través de técnicas experimentales por Neidigh y colaboradores[9]. Las simulaciones realizadas se compararon con la estructura experimental, para lo cual se realizó el análisis de RMSD. La estructura experimental se obtuvo del PDB con clave 1L2Y y se muestra en la figura 5.8.

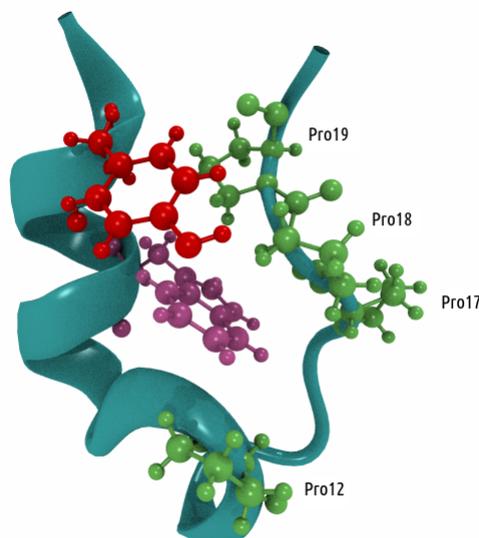


Figura 5.8: Estructura experimental obtenida con RMN[9] reportada en el PDB con clave 1L2Y. Los residuos que conforman el motivo trp-cage se resaltan con la representación CPK. Tirosina 3 (Rojo), triptofano 5 (lila), prolinas 12, 17, 18 y 19 (verde).

Se realizó una dinámica de tipo Langevin (ver ecuación 4.5) por 5 ps, donde se aplicó un potencial externo de tipo Gaussiano (ver ecuación 4.8) a los átomos pesados del péptido para inducir el plegamiento del péptido trp-cage. Posteriormente se realizó el análisis de RMSD de la trayectoria respecto a la estructura experimental, los resultados se muestran en las figuras 5.9 y 5.10. Así mismo se realizó una simulación de relajación por 5 ps en ausencia de potenciales externos, el análisis de RMSD también se muestra en las figuras 5.9 y 5.10. El análisis de RMSD en ambas simulaciones se aplicó únicamente al esqueleto peptídico de la proteína.

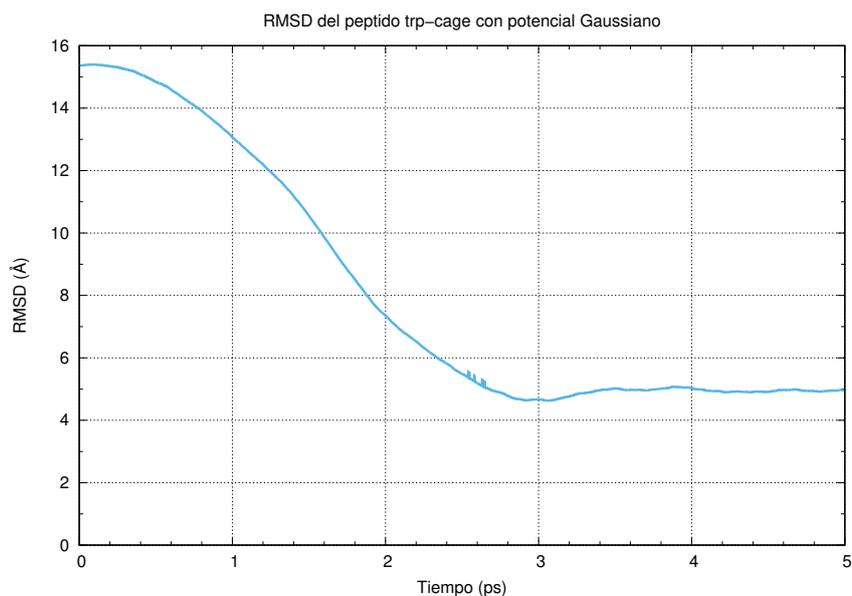


Figura 5.9: RMSD de la simulación de plegamiento con un potencial Gaussiano.

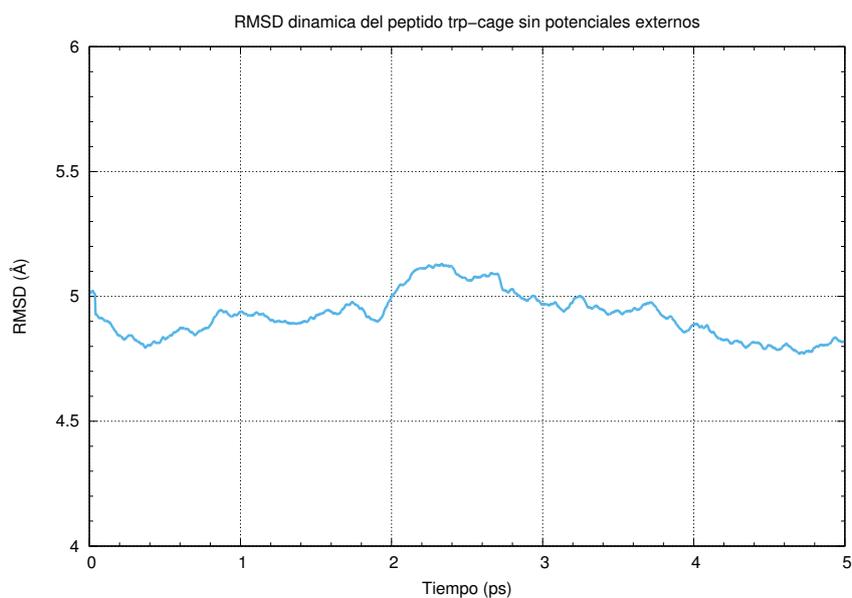


Figura 5.10: RMSD de la simulación sin potenciales externos.

En las figuras 5.9 y 5.10 se observa que el RMSD disminuye de 16 Å hasta 5 Å en menos de 5 ps. Así mismo, se aprecia un mínimo alrededor de los 3 ps, lo cual sugiere que las estructuras en esta zona de la trayectoria son más cercanas a la estructura nativa del péptido. No obstante durante la simulación no se observan elementos de estructura secundaria o terciaria. En la etapa de relajación el RMSD se mantiene alrededor de 5 Å por lo que el péptido se puede encontrar en un mínimo local. Las estructuras finales se muestran en la figura 5.11.

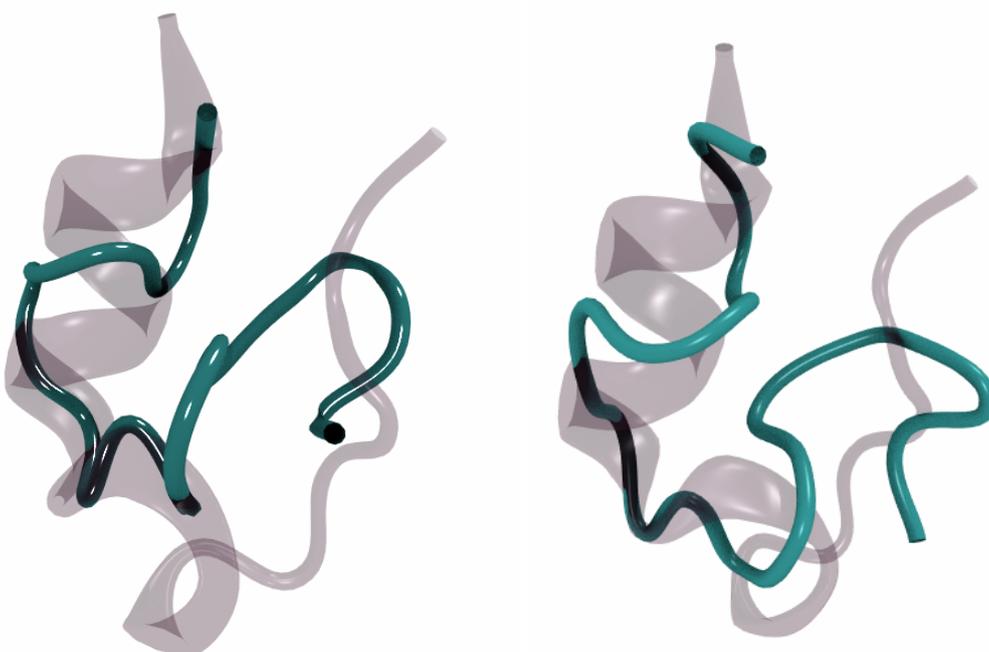


Figura 5.11: Izquierda: Estructura obtenida con la aplicación de potenciales de un potencial Gaussiano alineada con la estructura experimental. Derecha: Estructura final de la simulación de relajación comparada con la estructura experimental. La estructura experimental se representa en ambas figuras en transparencia sin cadenas laterales.

La energía potencial DFT se muestra en las figuras 5.12 y 5.13. En éstas se observa que el potencial ayuda a la proteína a atravesar un máximo de energía potencial (2-3 ps), posteriormente la energía disminuye hasta -0.4 Ha, por lo que el potencial induce a la proteína a adoptar una conformación de menor energía. En la simulación en ausencia de potenciales externos la energía DFT del péptido converge a un valor de 0.57 Ha, lo cual indica que el péptido alcanzó una estabilidad estructural. La energía reportada en las figuras 5.12 y 5.13 está dada por la diferencia entre la energía en cada paso de tiempo menos la energía del estado desplegado. El potencial Gaussiano aunque lleva a la proteína a una conformación de menor energía, ésta no alcanza la estructura nativa. Por lo que se requiere llevar a cabo simulaciones en periodos mayores de tiempo.

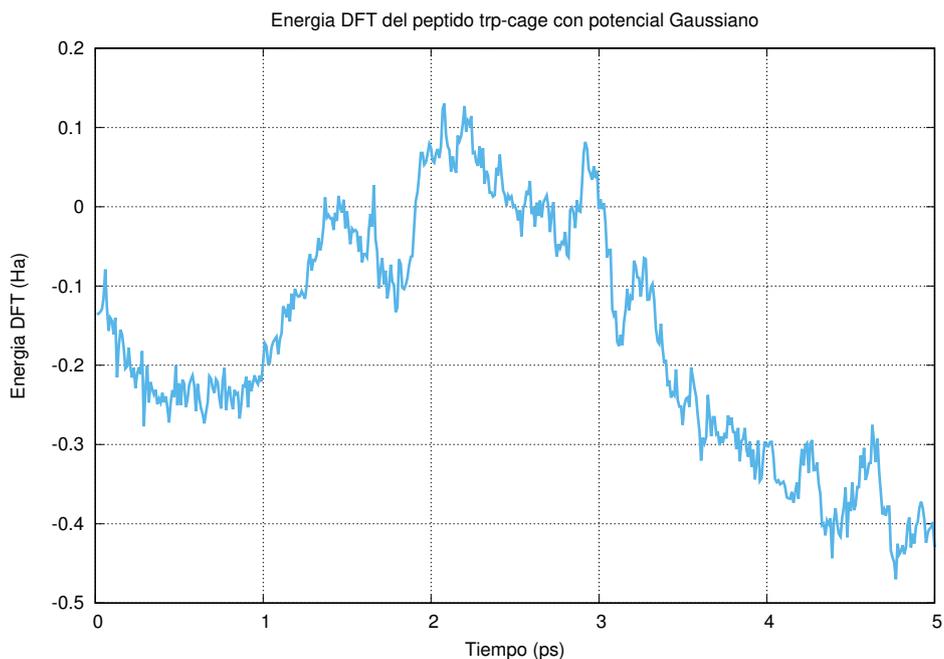


Figura 5.12: Energía potencial del péptido trp-cage. Dinámica de plegamiento.

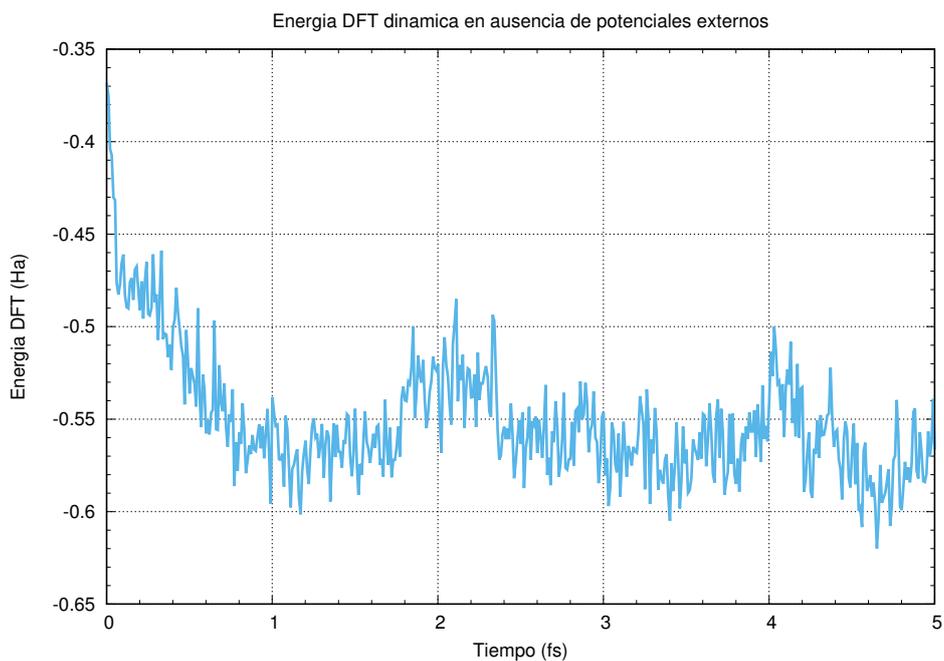


Figura 5.13: Energía potencial del péptido trp-cage. Dinámica de relajación.

Un factor importante que no se toma en cuenta en una dinámica tipo Langevin es la interacción electrostática entre las moléculas del disolvente y de la proteína, lo cual podría ser fundamental en la formación de elementos de estructura secundaria pero principalmente en la formación de núcleos hidrofóbicos. Por esta razón se implementó en el código el modelo de disolvente explícito TIP3P, para estudiar el efecto de potenciales externos en la dinámica de plegamiento del péptido trp-cage con disolvente explícito.

### Tercera etapa: Plegamiento del péptido trp-cage con disolvente

Se construyó una celda de disolvente triclinica con dimensiones  $45 \times 22 \times 22$  ( $\text{\AA}^3$ ) donde el péptido trp-cage se posicionó en el centro como se muestra en la figura 5.14. Las simulaciones con la proteína solvatada tomaron una mayor cantidad de tiempo para alcanzar el equilibrio termodinámico debido al gran número de átomos. Se realizó una simulación de equilibrio por 100 ps restringiendo las posiciones de los átomos de la proteína para permitir al disolvente alcanzar el equilibrio térmico. La temperatura durante esta simulación se ilustra en la figura 5.15, donde se observa que se alcanza el equilibrio térmico del sistema con una temperatura de 297 K.

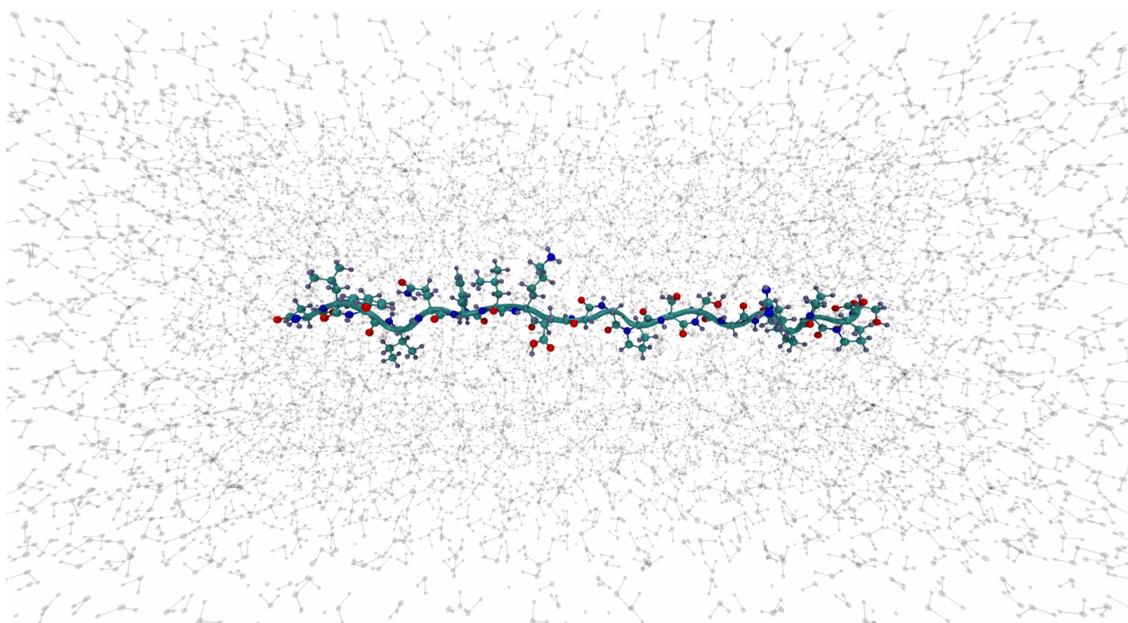


Figura 5.14: Estructura del péptido desplegado solvatado, caja triclinica con vista frontal.

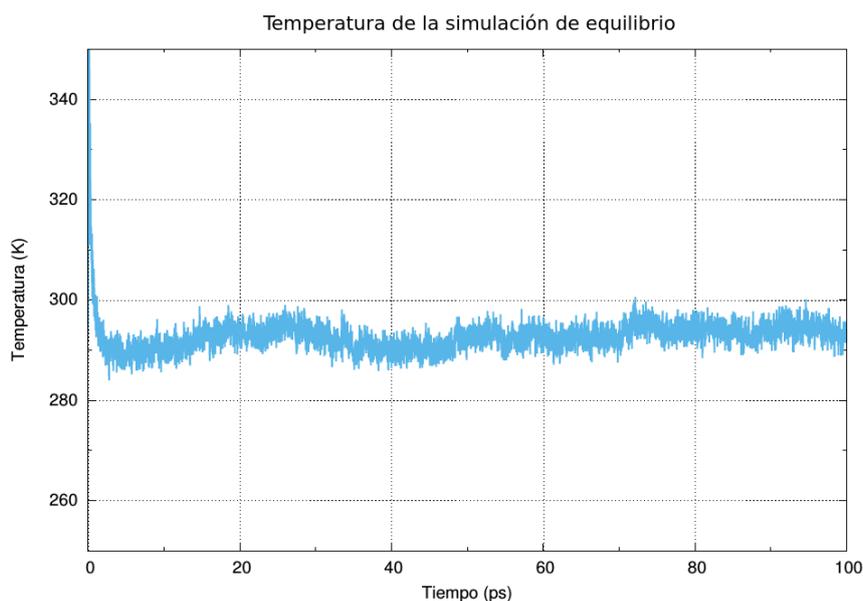


Figura 5.15: Temperatura durante simulación de equilibrio.

Teniendo al sistema en equilibrio termodinámico, se realizaron dos simulaciones con disolvente explícito empleando un potencial esférico de acuerdo con la ecuación 4.9. En la primera simulación se aplicó el potencial externo a todos los átomos pesados de la proteína. En la segunda simulación se aplicó el potencial externo únicamente a los átomos pesados de las cadenas laterales de los residuos hidrofóbicos, esto con el objetivo de introducir el efecto hidrofóbico. Por cada simulación de plegamiento se realizó una dinámica en ausencia de potenciales externo para permitir a la proteína adoptar una conformación menos estresada. Los resultados del análisis de RMSD de éstas simulaciones se muestran en las figuras 5.16 y 5.17.

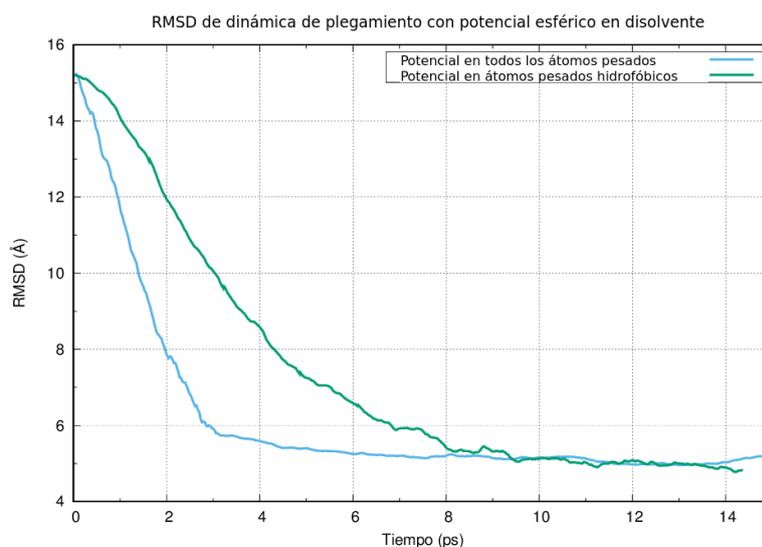


Figura 5.16: RMSD de la simulación de plegamiento del péptido trp-cage en disolvente con potencial esférico. La dinámica donde se aplicó el potencial sobre todos los átomos pesados se muestra en azul. En verde se ilustra la dinámica donde se aplicó el potencial únicamente sobre los átomos pesados de los residuos hidrofóbicos.

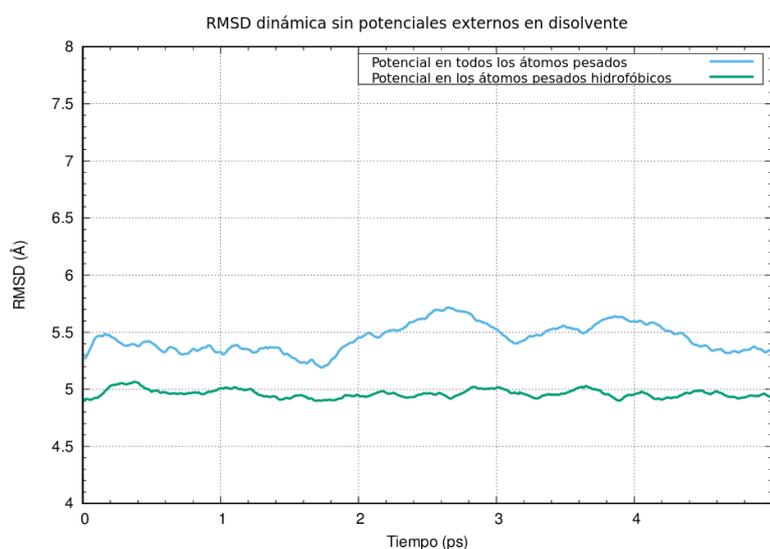


Figura 5.17: RMSD de la simulación del péptido trp-cage en disolvente sin potenciales externos. Relajación de la dinámica donde se aplicó el potencial sobre todos los átomos pesados se muestra en azul. En verde se ilustra la relajación de la dinámica donde se aplicó el potencial únicamente sobre los átomos hidrofóbicos.

En la figura 5.16 se observa que al aplicar el potencial externo únicamente a los átomos hidrofóbicos no se reduce el RMSD de la estructura obtenida. No obstante en las simulaciones de relajación (figura 5.17) se observa que la estructura donde se aplicó el potencial sobre los residuos hidrofóbicos alcanza un RMSD menor por 0.5 Å. También, se aprecia en la figura 5.16 que cuando se aplica el potencial a todos los átomos pesados del péptido, éste llega a una estructura más compacta en menos tiempo que cuando se aplica a los átomos de residuos hidrofóbicos.

La energía potencial DFT para ambas simulaciones se muestra en las figuras 5.18 y 5.19. Donde se observan comportamientos diferentes de la energía en ambas simulaciones. La simulación de plegamiento donde se aplicó el potencial externo a todos los átomos pesados del péptido (gráfico color azul figura 5.18) muestra un incremento asintótico de la energía a partir de los 12 ps, mientras que en la simulación donde se aplicó el potencial externo a los átomos hidrofóbicos (gráfico color verde figura 5.18) se aprecia una disminución de la energía entre los 6 ps y 8 ps.

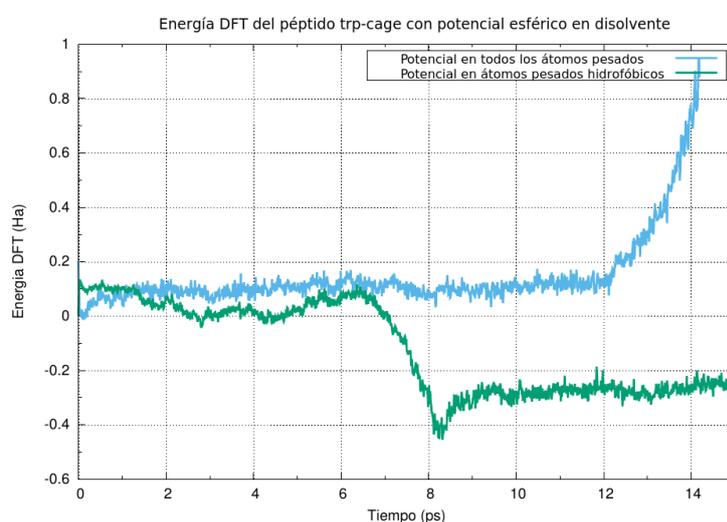


Figura 5.18: Energía potencial del péptido trp-cage. Dinámicas de plegamiento.

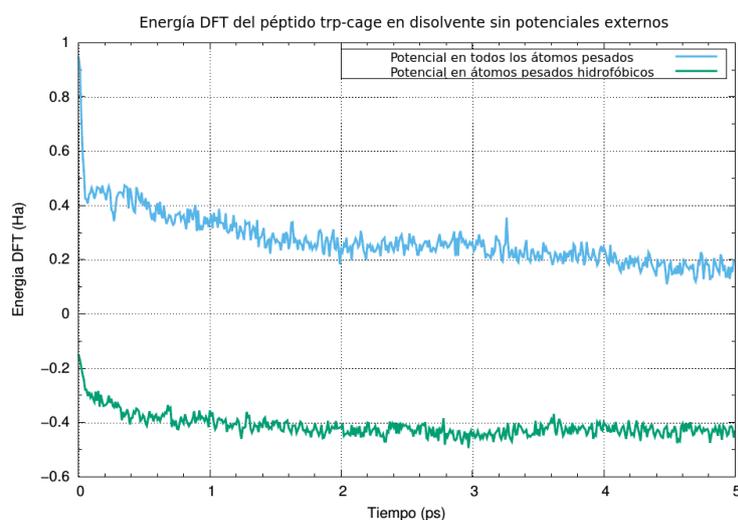


Figura 5.19: Energía potencial del péptido trp-cage. Dinámicas de relajación.

En el primer caso, el incremento de la energía se debe a que el potencial lleva al péptido a una estructura muy compacta donde la repulsión electrónica se vuelve dominante. Así mismo, se observa que la energía no disminuye en ningún momento a lo largo de la simulación. En el segundo caso, donde se aplicó el potencial externo a los átomos hidrofóbicos de los 6 ps a los 8 ps la energía potencial de la proteína disminuye considerablemente, se puede observar un mínimo a los 8.25 ps. Este comportamiento puede atribuirse a la formación de interacciones hidrofóbicas del péptido. Las estructuras antes y después de esta transición se muestran en la figura 5.20. En las simulaciones de relajación se observa que la energía converge. En la simulaciones de relajación (figura 5.19) se observa que donde se aplicó el potencial externo sobre los átomos pesados hidrofóbicos, el valor de la energía que alcanza el sistema al retirar el potencial es menor que en la dinámica donde se aplicó el potencial externo a todos los átomos pesados.

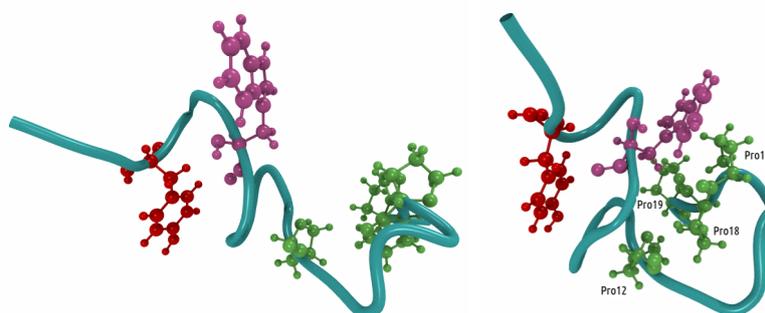


Figura 5.20: Estructuras del péptido trp-cage en la simulación de plegamiento con el potencial externo aplicado a los átomos hidrofóbicos. Izquierda: estructura al tiempo 6.4 ps. Derecha: estructura al tiempo 8.25 ps. El código de colores es el mismo que en la figura 5.8

En la figura 5.20 se aprecia que la estructura al tiempo de 6.4 ps se encuentra en un estado desplegado mientras que al tiempo 8.25 ps el péptido adopta una conformación más compacta donde existen contactos hidrofóbicos entre el triptofano (en color lila) con las prolinas 17 y 19, así como entre las prolinas 12 y 18. No obstante este arreglo se desvía del observado experimentalmente, ya que en la estructura experimental el triptofano se posiciona por debajo de los anillos de prolinas 17, 18 y 19. Las estructuras finales de las simulaciones de relajación, se muestran en la figura 5.21.

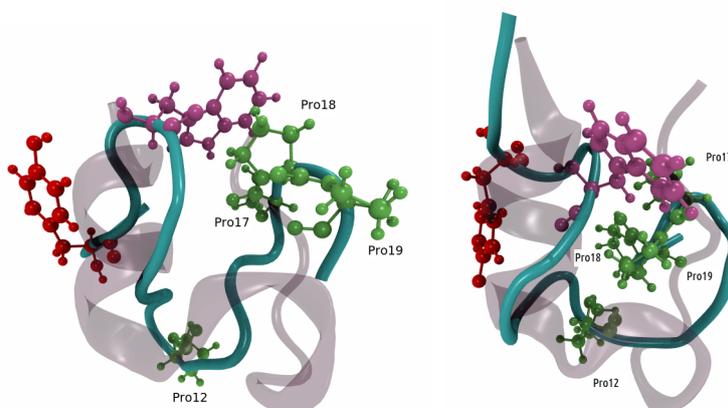


Figura 5.21: Péptido trp-cage en disolvente. Estructuras finales de la simulación de relajación. Izquierda: potencial esférico aplicado a todos los átomos del péptido. Derecha: Potencial aplicado a los átomos hidrofóbicos. El código de colores es el mismo que en la figura 5.8

En la figura 5.21 se observa, en ambos casos, que las estructuras obtenidas se desvían de la experimental, no obstante el potencial externo llevó al péptido a una conformación con un RMSD de 5 Å respecto a la estructura experimental. Así mismo, a pesar de que se pueden observar contactos hidrofóbicos entre los residuos de TRP y las prolinas 17, 18 y 19. El arreglo espacial de estos residuos no coincide con el arreglo del núcleo hidrofóbico observado en la estructura experimental (ver figura 5.8).

### Cuarta etapa: plegamiento del péptido trp-cage en GROMACS

Con el objeto de evaluar el desempeño del modelo de potenciales externos en comparación con modelos de dinámica molecular con campos de fuerza, se realizó una simulación de DM con el programa GROMACS en disolvente explícito del péptido trp-cage. Como configuración inicial se eligió la estructura desplegada del péptido en una celda de disolvente de las mismas dimensiones que en las simulaciones anteriores (ver figura 5.14). Se realizó una simulación NVT durante 20 ns a una temperatura de 297 K. Posteriormente se realizó el análisis de RMSD con la estructura experimental como referencia, los resultados se muestran en las figuras 5.22 y 5.23.

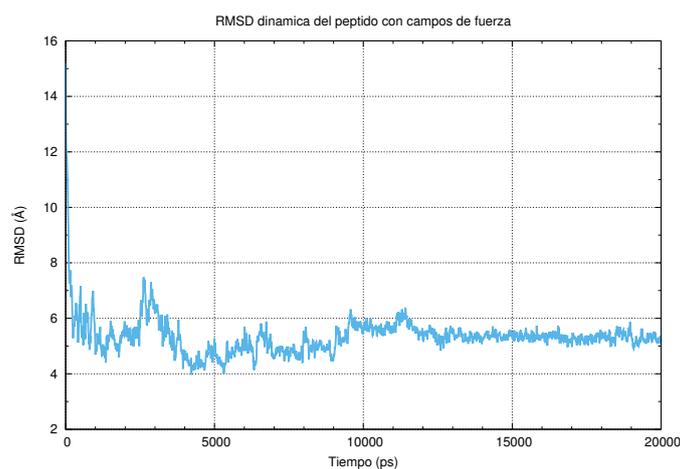


Figura 5.22: RMSD de la simulación del péptido trp-cage en disolvente con campos de fuerza.

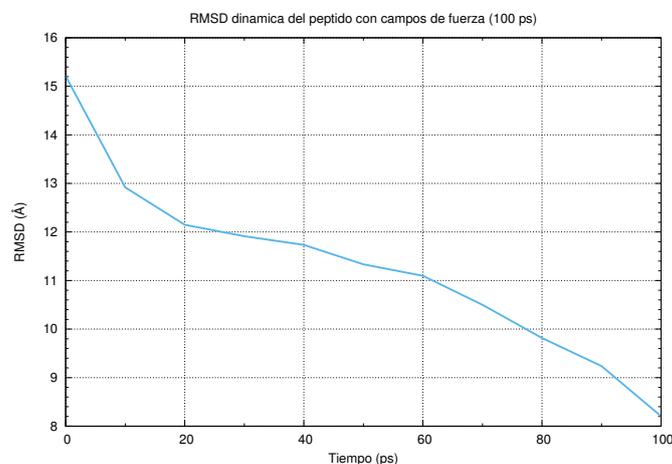


Figura 5.23: Ampliación primeros 100 ps del RMSD.

En la figura 5.22 se observa que el RMSD alcanza la convergencia a partir de los 12000 ps. La convergencia del RMSD indica que el péptido ha alcanzado una conformación estable, posiblemente se encuentre en un mínimo local. Las conformaciones exploradas con el modelo de DM con campos de fuerza presentan una desviación cuadrática media de alrededor 5 Å. En la gráfica 5.23 se muestran sólo los primeros 100 ps de la simulación, donde el RMSD alcanza apenas los 8 Å. La medida del RMSD de las conformaciones obtenidas con modelos convencionales de DM y campos de fuerza es similar al obtenido con el modelo propuesto en este trabajo en la mayoría de las simulaciones realizadas. No obstante, con el modelo propuesto en este trabajo se alcanza la convergencia en un tiempo de simulación al menos de un orden de magnitud menor que el tiempo de simulación necesario en una DM con campos de fuerza. La estructura obtenida se muestra en la figura 5.24.

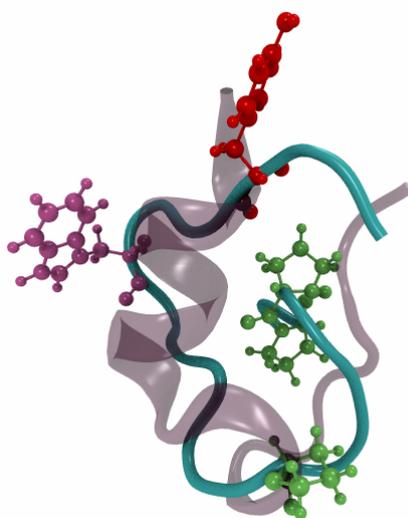


Figura 5.24: Estructuras del péptido trp-cage en la simulación de plegamiento con campos de fuerza en GROMACS. El código de colores es el mismo que en la figura 5.8

En la figura 5.24 se observa que la estructura obtenida con una simulación de 20 ns en GROMACS no presenta elementos de estructura secundaria o terciaria, así mismo tampoco se formaron contactos hidrofóbicos como los que se obtuvieron con el uso de potenciales externos en este trabajo. Así mismo, se aprecia que el residuo de triptofano queda expuesto al disolvente y no en el centro del núcleo hidrofóbico como en la estructura experimental.

## Conclusiones

- El uso de potenciales externos en las simulaciones de dinámica molecular ayudó a alcanzar una convergencia energética y estructural más rápido que con técnicas de DM y campos de fuerza convencionales.
- La aplicación de un potencial esférico en las simulaciones con disolvente favoreció la formación de contactos entre los residuos hidrofóbicos del péptido trp-cage, principalmente cuando el potencial se aplicó sólo a los átomos de los residuos hidrofóbicos.
- La aproximación teórica de Langevin para introducir efectos de temperatura y viscosidad en las fronteras del sistema, resultó efectiva ya que se alcanzó el equilibrio termodinámico en menos de 100 ps. Con lo que se evitó el uso de complejos algoritmos de termostatos usados normalmente en DM.
- En las estructuras obtenidas no se observaron los elementos de estructura secundaria y terciaria esperados. Esto puede deberse a que se necesita un tiempo mayor de simulación para que las proteínas exploren una mayor cantidad de conformaciones y encontrar su estado nativo.
- En la simulación realizada con campos de fuerza bajo las mismas condiciones de temperatura tampoco se alcanza la estructura nativa del péptido. Las conformaciones exploradas con DM y campos de fuerza presentan un RMSD similar al obtenido con el método propuesto.

## Proyectos a futuro

- Implementar el modelo de potenciales externos en software de dinámica molecular con campos de fuerza, como GROMACS y NAMD, para probar el modelo en simulaciones de nanosegundos (ns).
- Aplicar el modelo de energía libre en el método con potenciales externos, a través de la aproximación de Jarzynski [41].
- Realizar simulaciones con múltiples temperaturas empleando potenciales externos para vencer barreras de potencial y permitir al péptido explorar diferentes regiones del espacio configuracional.

---

# Referencias

- [1] Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindah, E. *SoftwareX* **2015**, *1-2*, 19–25.
- [2] Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. *Journal of chemical theory and computation* **2012**, *8*, 1542–1555.
- [3] Bock, L. V.; Blau, C.; Schröder, G. F.; Davydov, I. I.; Fischer, N.; Stark, H.; Rodnina, M. V.; Vaiana, A. C.; Grubmüller, H. *Nature Structural & Molecular Biology* **2013**, *20*, 1390–1396.
- [4] Freddolino, P. L.; Arkhipov, A. S.; Larson, S. B.; McPherson, A.; Schulten, K. *Structure* **2006**, *14*, 437–449.
- [5] Miao, Y.; Feixas, F.; Eun, C.; McCammon, J. A. *Journal of Computational Chemistry* **2015**, *36*, 1536–1549.
- [6] Micheletti, C.; Laio, A.; Parrinello, M. *Physical Review Letters* **2004**, *92*, 170601(4).
- [7] Karplus, M.; McCammon, J. A. *Nature Structural Biology* **2002**, *9*, 646–652.
- [8] Santamaria, R.; Jones, K.; Arroyo, M.; González-García, T. *Computational and Theoretical Chemistry* **2015**, *1068*, 72–80.
- [9] Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. *Nature Structural Biology* **2002**, *9*, 425–430.
- [10] Clementi, M. E.; Marini, S.; Coletta, M.; Orsini, F.; Giardina, B.; Misiti, F. *FEBS Letters* **2005**, *579*, 2913–2918.
- [11] D’Ursi, A. M.; Armenante, M. R.; Guerrini, R.; Salvadori, S.; Sorrentino, G.; Picone, D. *Journal of medicinal chemistry* **2004**, *47*, 4231–4238.
- [12] Perutz, M. F.; Rossmann, M. G.; Cullis, A. F.; Muirhead, H.; Will, G.; North, A. C. T. *Nature* **1960**, *185*, 416–422.
- [13] Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C. *Nature* **1958**, *181*, 662–666.
- [14] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic acids research* **2000**, *28*, 235–242.
- [15] Hubble, E. *Science* **2005**, *309*, 78b–102b.
- [16] Dill, K. A.; Ozkan, S. B.; Shell, M. S.; Weikl, T. R. *Annu.Rev.Biophys.* **2008**, *37*, 289–316.
- [17] Dill, K. A.; MacCallum, J. L. *Science (New York, N.Y.)* **2012**, *338*, 1042–1046.

- 
- [18] Costas, M. *Boletín de Educación Bioquímica* **1987**, *6*, 91–98.
- [19] Levinthal, C. *Mössbauer Spectroscopy in Biological Systems Proceedings* **1969**, *67*, 22–24.
- [20] Dill, K. A. *Biochemistry* **1985**, *24*, 1501–1509.
- [21] Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins: Structure, Function, and Bioinformatics* **1995**, *21*, 167–195.
- [22] Sugita, Y.; Okamoto, Y. *Chemical Physics Letters* **1999**, *314*, 141–151.
- [23] Bradley, P. *Science* **2005**, *309*, 1868–1871.
- [24] Mariani, V.; Kiefer, F.; Schmidt, T.; Haas, J.; Schwede, T. *Proteins: Structure, Function and Bioinformatics* **2011**, *79*, 37–58.
- [25] Grubmüller, H.; Heller, H.; Windemuth, A.; Schulten, K. *Molecular Simulation* **1991**, *6*, 121–142.
- [26] Choe, J. I.; Kim, B. *Bulletin of the Korean Chemical Society* **2000**, *21*, 419–424.
- [27] Feynman, R. P.; Leighton, R. B.; Sands, M. *The Feynman Lectures on Physics: The New Millennium Edition: Mainly Mechanics, Radiation, and Heat*; Basic Books: New York, 2015.
- [28] Santamaria, R.; de la Paz, A. A.; Roskop, L.; Adamowicz, L. *Journal of Statistical Physics* **2016**, *164*, 1000–1025.
- [29] Zwanzig, R. *Journal of Statistical Physics* **1973**, *9*, 215–220.
- [30] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *The Journal of Chemical Physics* **1983**, *79*, 926–935.
- [31] Neria, E.; Fischer, S.; Karplus, M. *The Journal of Chemical Physics* **1996**, *105*, 1902–1921.
- [32] Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- [33] Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. *The Journal of Chemical Physics* **1982**, *76*, 637–649.
- [34] Paterlini, M.; Ferguson, D. M. *Chemical Physics* **1998**, *236*, 243–252.
- [35] van Gunsteren, W.; Berendsen, H. *Molecular Physics* **1982**, *45*, 637–647.
- [36] Santamaria, R. <http://www.fisica.unam.mx/rso/publications2.html>, 2009.
- [37] Becke, A. D. *Physical Review A* **1988**, *38*, 3098–3100.
- [38] Lee, C.; Yang, W.; Parr, R. G. *Physical Review B* **1988**, *37*, 785–789.
- [39] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *The Journal of Chemical Physics* **2010**, *132*, 154104(19).
-

- [40] Ufimtsev, I. S.; Martinez, T. J. *Journal of Chemical Theory and Computation* **2009**, *5*, 2619–2628.
- [41] Jarzynski, C. *Physical Review Letters* **1997**, *78*, 2690–2693.