



BENEMÉRITA UNIVERSIDAD  
AUTÓNOMA DE PUEBLA

---

FACULTAD DE CIENCIAS QUÍMICAS Y CENTRO DE  
QUÍMICA- INSTITUTO DE CIENCIAS

POSGRADO EN CIENCIAS QUÍMICAS

TESIS

CARACTERIZACIÓN DEL POTENCIAL  
ELECTROSTÁTICO DE LARGO ALCANCE EN  
EL INTERIOR DE LA CHAPERONINA  
HSP60-HSP10 Y SU POSIBLE EFECTO  
SOBRE EL PLEGAMIENTO DE PROTEÍNAS

PARA OBTENER EL TÍTULO DE:  
MAESTRÍA EN CIENCIAS QUÍMICAS EN EL ÁREA DE  
FISICOQUÍMICA

PRESENTA:  
LBT. LUCIA JUANA PEÑA ORTIZ

DIRECTOR DE TESIS:  
DR. JUAN MANUEL SOLANO ALTAMIRANO

PUEBLA, PUE. DICIEMBRE 2022





**BUAP**

**A QUIEN CORRESPONDA:**

Por este conducto la Comisión Revisora formada por el Dr. Julio Manuel Hernández Pérez, Dra. Nora Hilda Rosas Murrieta, Dr. Alexandra Deriabina, y el Dr. Joaquín Barroso Flores, informan haber leído y aprobado para su impresión, la tesis denominada **“Caracterización del potencial electrostático de largo alcance en el interior de la chaperonina HSP60-HSP10 y su posible efecto sobre el plegamiento de proteínas”**, que presenta la alumna **Lucía Juana Peña Ortiz** para obtener el grado de Maestra en Ciencias Químicas.

Se extiende la presente a los veinticuatro días del mes de noviembre del año dos mil veintidós.

**ATENTAMENTE**

Dr. Julio Manuel Hernández Pérez

Dra. Nora Hilda Rosas Murrieta

Dra. Alexandra Deriabina

Dr. Joaquín Barroso Flores



# Dedicatoria

*Con cariño a mis padres, hermanos y, especialmente, mi esposo cuyo apoyo incondicional ha sido sumamente importante para mí. Mis logros son sus logros.*

# Agradecimientos

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

Al consejo sinodal conformado por el Dr. Joaquín Barroso Flores, el Dr. Julio Manuel Hernández Pérez, la Dra. Nora Hilda Rosas Murrieta y la Dra. Alexandra Deriabina, por sus observaciones y comentarios.

A la Dra. Viridiana Vargas Castro, cuyos conocimientos y aportaciones sobre la herramienta de dinámica molecular contribuyeron en gran medida en la realización de esta tesis.

Finalmente a mi asesor, el Dr. Juan Manuel Solano-Altamirano por apoyo en la realización de esta tesis pero, sobre todo, por la confianza para trabajar en su grupo de investigación, pues hasta hace dos años fui completamente desconocida para él.

Participación en el *VIII Encuentro con la Calorimetría* (ocurrido el día 28 de noviembre del 2022): Lucia J. Peña Ortiz, Viridiana Varga-Castro, Juan Manuel Solano-Altamirano, *Estudio del campo electrostático de largo alcance en el interior de la proteína HSP60*.

Participación en la *XX Reunión Mexicana de Fisicoquímica Teórica* (ocurrida los días 17-19 de Noviembre del 2022): Lucia J. Peña Ortiz, Viridiana Varga-Castro, Juan Manuel Solano-Altamirano, *Estudio del campo electrostático de largo alcance en el interior de la proteína HSP60*.

Participación en el simposium *Técnicas computacionales aplicadas a las ciencias Químico-Biológicas* (ocurrido virtualmente los días 7 y 8 de abril del 2022): Lucia J. Peña Ortiz. *¿Por qué debe importarnos la dinámica molecular?: Principios básicos y aplicaciones*.

Participación en la *XIX Reunión Mexicana de Fisicoquímica Teórica* (ocurrida virtualmente los días 18 y 19 de Noviembre del 2021): Lucia J. Peña Ortiz, Viridiana Varga-Castro, Juan Manuel Solano-Altamirano, *Estudio del campo de fuerzas efectivas en el interior de la proteína de choque térmico 60*.

# Índice general

<b>Índice de figuras</b>	VII
<b>Índice de tablas</b>	IX
<b>1. Introducción</b>	1
<b>2. Antecedentes</b>	4
2.1. El problema del plegamiento de proteínas . . . . .	4
2.2. Plegamiento de proteínas en el entorno biológico . . . . .	8
2.3. Retos computacionales . . . . .	11
2.3.1. Dinámica molecular . . . . .	12
<b>3. Objetivos e hipótesis</b>	14
3.1. Objetivos particulares . . . . .	14
3.2. Hipótesis . . . . .	15
<b>4. Herramientas y métodos</b>	16
4.1. Formalismos básicos de dinámica molecular clásica . . . . .	16
4.1.1. Esquemas de integración . . . . .	16
4.1.2. Función de energía potencial . . . . .	18

## ÍNDICE GENERAL

---

4.1.3. Condiciones de frontera	19
4.1.4. Equilibración: ajuste de temperatura y presión	21
4.2. Metodología	23
4.2.1. Preparación de los sistemas de estudio	23
4.2.2. Establecimiento de la configuración inicial y definición de las condiciones de equilibrio	24
4.2.3. Producción de las simulaciones de dinámica molecular	25
4.2.4. Análisis de datos	26
<b>5. Resultados y discusión</b>	<b>30</b>
5.1. Estabilidad estructural de la HSP60 en los diferentes sistemas moleculares de estudio	31
5.2. Mapas de la distribución de potencial electrostático instantánea y promedio de la HSP60 en los diferentes sistemas moleculares de estudio	34
<b>6. Conclusiones</b>	<b>44</b>
<b>Referencias</b>	<b>50</b>

# Índice de figuras

2.1. Resumen de los conceptos generales que actualmente se aceptan como válidos a la hora de explicar el fenómeno del plegamiento de proteínas. . . . .	5
2.2. Modelo del embudo para el plegamiento de una proteína. . . . .	6
2.3. Papel de las chaperonas moleculares en el contexto del modelo de paisaje energético del plegado de una proteína. . . . .	7
2.4. Estructura tridimensional de los diferentes tipos de chaperoninas. . . . .	9
2.5. Organización estructural del complejo de chaperoninas HSP60-HSP10. . . . .	10
4.1. Ejemplo de condiciones de contorno periódicas bidimensional. . . . .	21
5.1. Evolución temporal de la desviación cuadrática media (RMSD) para cada uno de los sistemas moleculares de estudio de la HSP60. . . . .	32
5.2. Evolución temporal de la desviación cuadrática media (RMSD) para la proteína cytC. . . . .	33
5.3. Mapas de la distribución del potencial electrostático de largo alcance instantánea del sistema HSP60-HSP10 en el plano YZ. . . . .	35
5.4. Mapas de la distribución de largo alcance instantánea del sistema HSP60-HSP10-cytC en el plano YZ. . . . .	36

## ÍNDICE DE FIGURAS

---

5.5. Mapas de la distribución del potencial electrostático de largo alcance instantánea del sistema HSP60 en el plano YZ . . . . .	38
5.6. Mapas de la distribución de potencial electrostático de largo alcance promedio para el sistema HSP60 desde diferentes planos. . . . .	39
5.7. Mapas de la distribución de potencial electrostático de largo alcance promedio para el sistema HSP60-HSP10 desde diferentes planos. . . . .	40
5.8. Mapas de la distribución de potencial electrostático de largo alcance promedio para el sistema HSP60-HSP10 desde diferentes planos. . . . .	41
5.9. Mapas de la distribución del potencial electrostático de largo alcance instantánea de la proteína cytC en el plano YZ. . . . .	42
5.10. Mapas de la distribución de potencial electrostático de largo alcance promedio para CitC desde diferentes planos. . . . .	43
6.1. Ejemplo de cuadrícula de potencial para calcular el campo promedio de la distribución de largo alcance de una molécula de interés en <i>genAveragePME</i> . . . . .	48
6.2. Esquema general del proceso para el cálculo, generación de datos y representación de la distribución de potencial electrostático de largo alcance con ayuda de los programas <i>genAveragePME</i> y <i>manipdx</i> diseñados a propósito para realizar esta tarea. . . . .	49

# Índice de tablas

4.1. Resumen de parámetros para la construcción de las configuraciones iniciales	
para los sistema moleculares de estudio de la HSP60. . . . .	25

# Resumen

Todas las proteínas necesitan plegarse en una estructura tridimensional para realizar su función, a partir de su secuencia de aminoácidos. Este proceso está gobernado por una serie de factores importantes como las interacciones hidrofóbicas, la presencia de proteínas chaperonas y el balance entálpico-entrópico, entre otros. Las chaperonas comprenden un amplio grupo de proteínas involucrado en procesos de asistencia en el plegamiento de otras proteínas. La importancia de las chaperonas nace de la necesidad de las proteínas de interaccionar con al menos una chaperona durante su proceso de biosíntesis celular. De lo anterior, se intuye que estudiar el plegamiento de proteínas debería incluir, de alguna manera, el efecto de las chaperonas. Por lo tanto, en este trabajo analizamos la posibilidad de imitar el efecto de una chaperona, donde la distribución de potencial electrostático interna de esta representaría un papel importante sobre el proceso de plegamiento de una proteína. Para ello, utilizamos simulaciones de dinámica molecular clásica para mapear la distribución de potencial electrostático de largo alcance de la chaperona HSP60 sola y de la HSP60 con otras proteínas. Los resultados muestran un patrón no aleatorio de la distribución espacial del potencial electrostático de largo alcance aparentemente invariante a lo largo de la porción de trayectoria analizada. En consecuencia, parece factible simular el efecto de una chaperonina a través de un campo de fuerzas externo aproximado que reproduzca el patrón observado.

# Abstract

In order to properly function, every protein must fold from its sequence of amino acids into a three-dimensional structure. This process is governed by a series of important factors, such as hydrophobic interactions, the enthalpic-entropic balance, and the presence of chaperones. Chaperones are a wide group of proteins involved in assisting the folding of other proteins, which render the chaperones to be of great importance in cellular biosynthetic processes. Hence, it becomes intuitive that the effects of chaperones on the protein folding process should be included somehow in the study of protein folding. We believe that the electrostatic potential at the interior of a chaperone should play an important role. Therefore, in this work we performed Molecular Dynamics Simulations of the HSP60 chaperone alone, as well as of several complexes comprising the HSP60 and other proteins. Subsequently, we analyzed the long-range electrostatic potential over a small portion of the MD trajectory of the previous systems. We found that the spatial distribution of the long-range electrostatic potential shows a non-random pattern, which is apparently invariant over the analyzed trajectory portion. Consequently, it seems feasible to simulate the effects of chaperones on the folding of proteins, by adding an external force to Molecular Dynamics Simulations of proteins-to-be-folded.

# Introducción

Uno de los grandes problemas abiertos en la bioquímica estructural es el denominado *plegamiento de proteínas*, el cual plantea la duda sobre cómo a partir de la información contenida en el ADN de las células se podría predecir la estructura tridimensional de una proteína funcional. Diferentes disciplinas como la bioinformática, la modelización *in silico* y, más recientemente, la inteligencia artificial han conseguido logros destacables en este sentido. Sin embargo, el trabajo de resolver la estructura de estas macromoléculas sigue siendo complicado y costoso.

Particularmente, a principios de los años 60 se creía que toda la información necesaria para conseguir que una proteína se plegará en su conformación funcional se hallaba contenida en su secuencia aminoacídica [1]. Sin embargo, hoy se sabe que no siempre sucede así, ya que bajo ciertas circunstancias (como la alta concentración macromolecular, el pH, la temperatura y algunas consideraciones energéticas de tipo entálpicas y entrópicas) las proteínas no pueden alcanzar por sí solas su estructura tridimensional funcional [2]. En condiciones fisiológicas, las células disponen de un sofisticado sistema de proteínas, colectivamente conocidas como *chaperonas moleculares* que asisten el plegamiento y ensamble de otras proteínas [3,4]. Un grupo importante dentro de esta familia de proteínas son las *chaperoninas*, las cuales son agregados moleculares con una cavidad aislada cilíndrica, donde se lleva a cabo el plegado de las proteínas sin interferencias [5]. A la fecha, numerosas metodologías de carácter funcional y estructural han permitido comprender el mecanismo que emplean las

chaperoninas para ejercer su función. Sin embargo, poco se conoce sobre los aspectos físicos y químicos en el interior de dichos sistemas con el fin de asistir el correcto plegamiento de determinadas proteínas.

Junto a las metodologías experimentales, los métodos teóricos constituyen un complemento necesario de cara a obtener una descripción detallada de los aspectos estructurales y dinámicos de sistemas físicos, químicos y biológicos, particularmente cuando se trata de simular condiciones de entorno y situaciones químicas concretas que no son fáciles de medir experimentalmente. A este respecto, la dinámica molecular (DM) se ha convertido en una de las herramientas más útiles para estudiar el comportamiento dinámico de estructuras microscópicas. Sin embargo, a menudo presenta limitaciones como el del muestro conformacional insuficiente, campos de fuerza empíricos poco transferibles y la escala de tiempo real que, entre otras cosas, hace poco práctica su aplicación en el estudio del proceso de plegado de proteínas. Recientemente la implementación del modelo de potenciales externos en conjunto con campos de fuerzas ha obtenido resultados exitosos en el plegamiento de polímeros orgánicos y algunos péptidos pequeños. Estos cumplirían principalmente dos funciones: i) el de introducir efectos del medio en el proceso de plegamiento como el efecto hidrofóbico o la presencia de proteínas chaperonas y ii) el de acelerar el muestro conformacional para ahorrar tiempo computacional [6].

Con el interés de que en nuestro laboratorio de trabajo se desarrolle un método más eficiente de dinámica molecular, en donde se simule la presencia de una chaperonina (sin que ésta se incluya explícitamente) en el proceso de plegado de una proteína, en esta tesis analizamos algunas propiedades físicas del interior de una chaperonina específica. Lo anterior constituye el primero de varios pasos (en un proyecto de largo plazo), y en este primer paso (como proyecto de tesis de maestría) abordamos la prueba de concepto, es decir, buscamos obtener información que nos confirme que es posible (o no) reemplazar (en simulaciones) la chaperona por algún campo de fuerza externo y/o una pared confinante. Para alcanzar el objetivo de dicho primer paso, como sistema de estudio de esta investigación, se eligió a la chaperonina HSP60 [7,8]. A partir de esta estructura se construyeron tres sistemas. El sistema HSP60, cuyo único componente sería la estructura de la HSP60 (vacío en nuestro ejem-

plo artificial), el sistema HSP60-HSP10, cuyo componente sería el complejo entre la HSP60 y la co-chaperonina HSP10 (también vacío en nuestro ejemplo artificial y el sistema HSP60-HSP10-CitC, cuyos componentes además del complejo entre la HSP60 y la HSP10, incluiría en su interior a la proteína del citocromo C (de aquí en adelante referida como CitC). Una vez construidos los sistemas, se realizaron simulaciones de DM y posteriormente se realizó un análisis de la distribución del potencial electrostático de largo alcance durante cierto periodo de tiempo. Encontramos que la distribución espacial del potencial electrostático de largo alcance en el interior de la chaperonina no es aleatorio, sino que muestra cierta estructura como función de las coordenadas, y además es aparentemente invariante a lo largo de la porción de las trayectorias analizadas. Estos dos hechos nos confirman que es viable, desde una perspectiva física, proponer un campo de fuerza externo y/o una pared confinante en lugar de la chaperonina.

## Antecedentes

### 2.1. El problema del plegamiento de proteínas

Desde principios de la década de los 50 se hizo evidente que la función de una proteína depende de su estructura tridimensional única. Los investigadores habían correlacionado claramente las propiedades específicas de las proteínas funcionales con configuraciones moleculares específicas y la pérdida de las propiedades específicas que ocurre en la proteína desnaturalizada con los cambios que ocurrían en su configuración [9]. A esto, se sumaría la serie de experimentos realizados por C. Anfinsen y colaboradores en 1961. Anfinsen y su equipo de trabajo pusieron de manifiesto que toda la información necesaria para conseguir que una proteína se plegase en un ambiente determinado se hallaba contenida en su secuencia aminoacídica. En forma específica, C. Anfinsen demuestra la existencia de un estado termodinámicamente estable para las proteínas y supondría un mínimo en la energía libre de Gibbs (conocida como “la hipótesis termodinámica de Anfinsen”) [10]. Este trabajo ubicaría al problema de plegamiento de proteínas dentro del campo de estudio de los procesos físicos, dando así inicio a una serie de investigaciones que pretenderían establecer, bajo bases físicoquímicas, la naturaleza de los mecanismos involucrados [11]. Más adelante, en contraposición a ello, este modelo de plegamiento fue discutido por Cyrus Levinthal, en consideración al poco tiempo que requiere el proceso de plegamiento en la naturaleza. Según él, una búsqueda al azar (representada por la hipótesis del control termodinámico de

Anfinsen) de todas las conformaciones proteicas posibles no puede darse en realidad, dado el enorme tiempo que requeriría. Con esto, Levinthal concluyó que las proteínas no siguen una búsqueda aleatoria para llegar a su estado plegado, sino que se presume un mecanismo de plegamiento en el que 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 [6]. Este hecho se conoce como la paradoja de Levinthal y representa la hipótesis sobre el control cinético del proceso de plegamiento de proteínas.

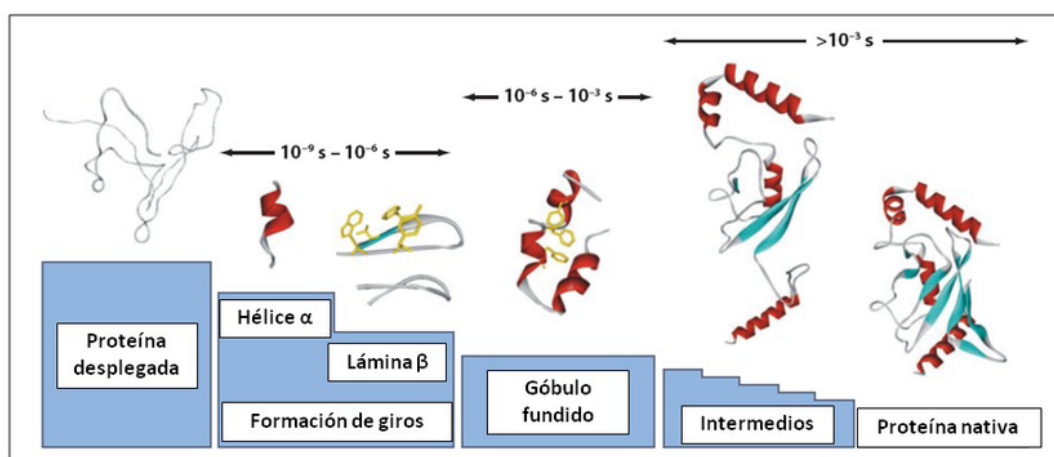


Figura 2.1: Resumen de los conceptos generales que actualmente se aceptan como válidos a la hora de explicar el fenómeno del plegamiento de proteínas [1]. Si consideramos genéricamente los modelos clásicos de plegamiento, el primer evento del plegamiento consiste en el colapso de interacciones hidrofóbicas de largo alcance (“colapso hidrofóbico”) acompañado de la aparición de pre-intermediarios como el glóbulo fundido o *molten globule* caracterizados por el nivel significativamente alto de estructura secundaria. Los sucesos intermediarios en escala de tiempo de plegamiento aparecen después de la formación del *molten globule* y antes de la etapa limitante del proceso de plegamiento que genera la estructura nativa funcional.

Por otro lado, a fin comprender cómo una cadena polipeptídica puede evitar verse atrapada en la paradoja de Levinthal, se propusieron diferentes modelos de plegamiento tanto a partir de observaciones experimentales como desde consideraciones teóricas. En 1953, los experimentos llevados a cabo por Thomas E. Creighton y col. con el inhibidor pancreático bovino vendrían a ser el detonante del comienzo de una intensa carrera, aun no acabada, hacia la comprensión de los mecanismos de plegamiento de proteínas, en los cuales juegan un papel fundamental los métodos de predicción de estructuras. Thomas fue capaz de demos-

trar que durante el plegamiento del inhibidor pancreático bovino se producían una serie de estados intermedios que se podían “atrapar” [1]. En otras palabras, Thomas pudo demostrar que el camino que puede seguir una proteína desde su estado desplegado hasta alcanzar la estructura tridimensional nativa no es necesariamente único y requiere de la aparición de intermedios (ver Figura 2.1). Estos intermedios, según la hipótesis del control cinético, corresponderían a mínimos locales de la energía libre de Gibbs, de ahí su relativa estabilidad. Como consecuencia, la identificación y caracterización estructural de los intermedios de plegamiento se vio transformada en un tema de importancia crucial (ver Figura 2.1).

A todo esto, los métodos mecano-estadísticos supusieron un modelo unificado del plegamiento proteico basado en la superficie energética efectiva. Dicho modelo llamado nuevo enfoque o nueva visión (del inglés *new view*) es representado por un embudo de plegamiento y describe el comportamiento termodinámico y cinético de la transformación de un conjunto de moléculas desplegadas a un estado predominantemente funcional (ver Figura 2.2).

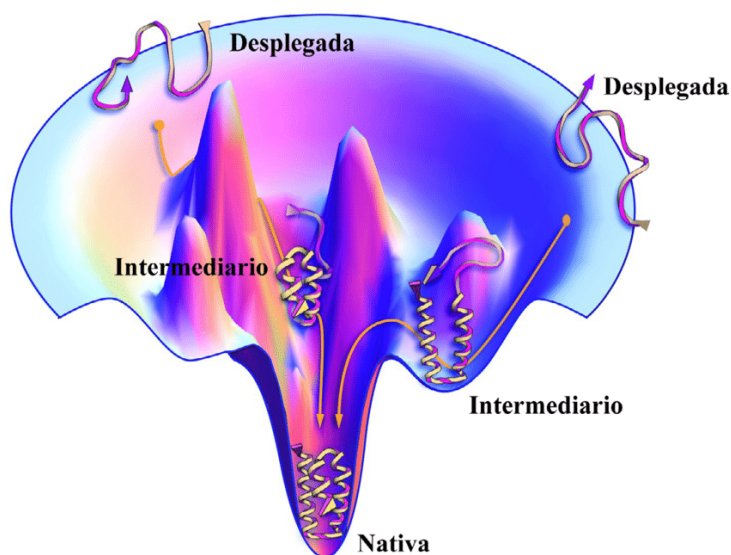


Figura 2.2: **Modelo del embudo para el plegamiento de una proteína** [2]. Cada conformación que la proteína puede adoptar tiene asociada una energía determinada; al conjunto de las energías correspondientes a todas las conformaciones posibles se le conoce como paisaje energético. Con base en diversos estudios teóricos y experimentales se propone que el paisaje energético de las proteínas tiene forma de embudo, donde el estado desnaturalizado se compone de un gran conjunto de conformaciones de alta energía y entropía. El espacio conformacional se reduce conforme la proteína se acerca al estado de menor energía (nativo) en el fondo del embudo.

De acuerdo con este modelo, hay pequeñas rutas paralelas y cada cadena polipeptídica seguiría su propia ruta individual. Cercano al fondo del embudo de plegamiento, el número de conformaciones de la proteína disminuye como lo hace a su vez la entropía. Cuanto más inclinada sea la pendiente más rápido será el plegamiento (ver Figura 2.2). A partir de modelo fue posible entender las cinéticas de plegamiento de proteínas, así como el mal plegamiento proteico y su posible agregación. Diversos resultados experimentales apoyan esta teoría y demuestran de manera creciente que la cadena polipeptídica inicialmente extendida se pliega a través de una población heterogénea de intermediarios parcialmente plegados en un equilibrio fluctuante [6].

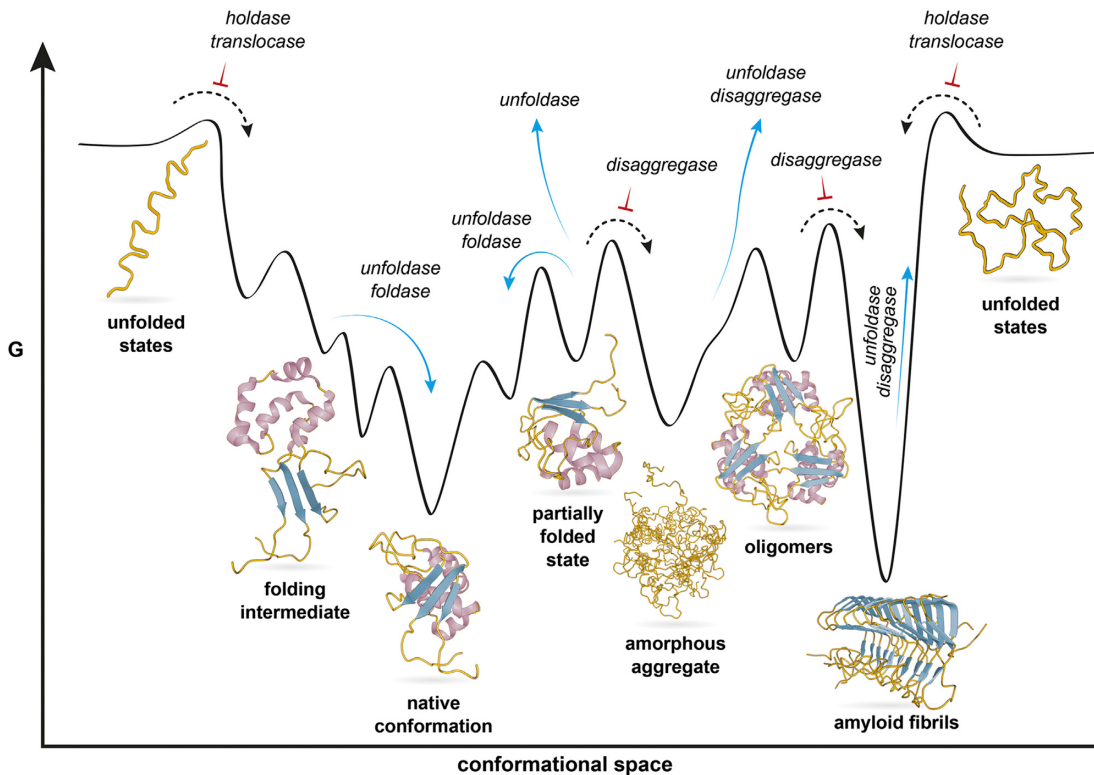


Figura 2.3: **Papel de las chaperonas moleculares en el contexto del modelo de paisaje energético del plegado de una proteína** [12]. En amarillo se muestra la cadena polipeptídica de una proteína con diferentes elementos estructurales secundarios resaltados en púrpura y azul. La proteína navega por la superficie de energía libre. Los varios mínimos locales representan diferentes estados conformacionales. Las chaperonas actúan en un amplio rango de procesos celulares que definen el rango de funciones listadas en letras itálicas. Las flechas azules indican transiciones que facilitan las chaperonas, mientras que las flechas negras con T invertida roja indican transiciones que las chaperonas evitan.

## 2.2. Plegamiento de proteínas en el entorno biológico

Desde su descubrimiento a finales de la década de los 70, las chaperonas moleculares han desafiado las reglas básicas del plegamiento proteico hasta ese momento establecidas. A raíz de diversos estudios experimentales, se ha comprobado que las chaperonas moleculares no solo previenen la agregación de conformaciones no nativas y mantienen a las proteínas en las pautas necesarias para su plegamiento correcto, sino que se encargan de prevenir interacciones no deseadas con otros componentes celulares, de dirigir el ensamblaje de proteínas de gran tamaño y complejos multiproteicos y, durante situaciones de estrés, de potenciar el plegamiento de numerosas proteínas (ver Figura 2.3) [3,4]. De esta manera, dentro del contexto del proceso de plegamiento de proteínas *in vivo*, la importancia de las chaperonas nace de la necesidad de las proteínas de interactuar con al menos una chaperona durante su proceso de biosíntesis celular. De ahí que parece razonable la aproximación al estudio del plegamiento utilizando este tipo de sistemas biológicos.

Una propiedad única de las chaperonas moleculares es la cavidad que proporcionan en la que las proteínas se pliegan. Se ha identificado la presencia de esta estructura en un grupo importante de proteínas de estructura homóloga y peso molecular en torno a 60 kDa (kilo Daltons) conocidas como *chaperoninas* (ver Figura 2.4) [13]. Esta estructura de tipo cilíndrica es significativamente diferente a otras superficies proteicas y es que se ha probado su capacidad para adherirse a una resina de cromatografía permitiendo la renaturalización de proteínas mediante el paso de éstas por una columna empaquetada con el complejo resina-*minichaperonina* [14]. Asimismo, ya se ha demostrado el interés en el estudio teórico del efecto de las chaperonas sobre el plegamiento de una proteína imitando los efectos de GroEL-GroES, el complejo más representativo de la familia de las chaperoninas, mediante un confinamiento espacial. Los resultados sugieren que los efectos primarios de este tipo de sistemas moleculares pueden ser imitados de una manera simple utilizando modelos simplificados (como *Coarse Grain*) y ser utilizados para facilitar el plegamiento de proteínas en simulaciones de dinámica molecular. Además, los resultados apoyan la idea en la que el confinamiento espacial juega un rol importante en el plegamiento de las proteínas confina-

das [15-17].

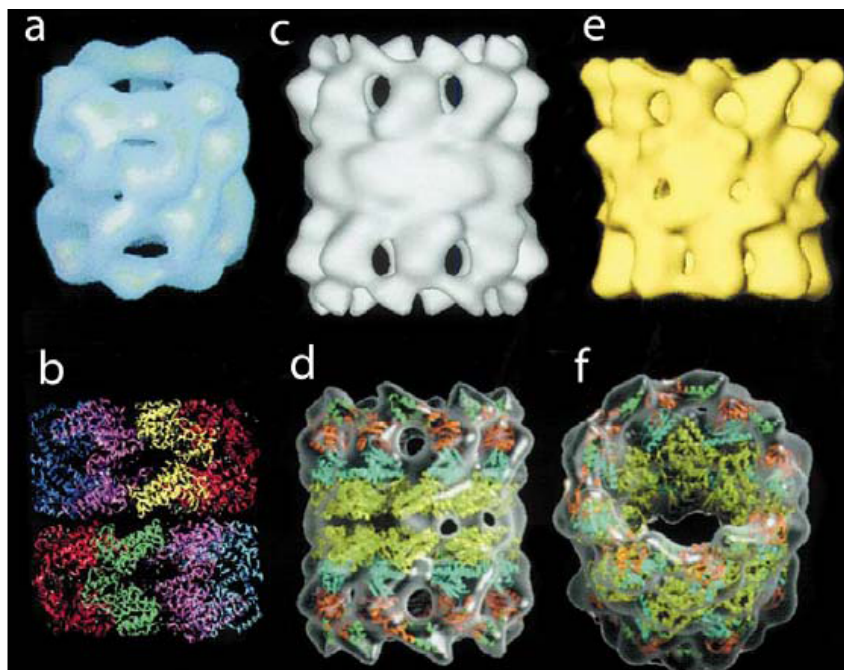


Figura 2.4: **Estructura tridimensional de los diferentes tipos de chaperoninas** [18] En la parte superior se muestra las reconstrucciones tridimensionales obtenidas por criomicroscopía de (a) GroEL [19], (c) Termosoma [20] y (e) CCT (del inglés *Chaperonin Containing TCP-1*) [21]. En (b) se muestra la estructura atómica de GroEL, mientras que en (d) y (f) se muestran dos vistas del ajuste de la estructura atómica modificada del Termosoma en la reconstrucción tridimensional mostrado en (c). Se ha caracterizado distintos tipos de conformaciones para este tipo de proteínas en presencia y/o ausencia de ATP, que dan como resultado el confinamiento espacial de la proteína sustrato. En la Termosoma y la CCT, la unión de ATP genera la conformación cerrada debido a la presencia de protuberancias a la entrada de la cavidad que la cierran al producirse el contacto con el sustrato, mientras que GroEL, además de la unión de ATP, el cierre de la cavidad se produce por la unión de oligómeros de bajo peso molecular denominados cochaperoninas (por ejemplo, GroES). El sustrato que se ha unido a la chaperonina queda así cerrado en su cavidad y, es entonces, que se inicia el proceso de plegamiento que dará como resultado una proteína funcional.

Derivado de las observaciones anteriores, en este trabajo investigamos la posibilidad de imitar el efecto de las chaperoninas sobre el proceso de plegamiento de una proteína desde un punto de vista físico, donde la distribución de carga de los residuos al interior de la cavidad representan un efecto importante sobre el plegamiento de aquella. Asimismo, se ha encontrado que las chaperonas presentan una distribución única de residuos en sus cavidades interiores que tienen grandes fracciones de residuos cargados, particularmente, el comple-

jo de chaperoninas GroEL-GroES de *T. thermophilus*, presenta al interior de su cavidad un 70% de residuos cargados, cuyo papel es estabilizar el plegamiento de proteínas dentro de la cámara aumentando el efecto hidrofóbico [22,23]. En nuestro caso, se eligió como sistema de estudio en esta investigación al complejo de chaperoninas HSP60-HSP10 (ver Figura 2.5), cuya similaridad estructural se asemeja con algunos de los miembros más representativos de la familia de las chaperoninas presentados en la Figura 2.4.

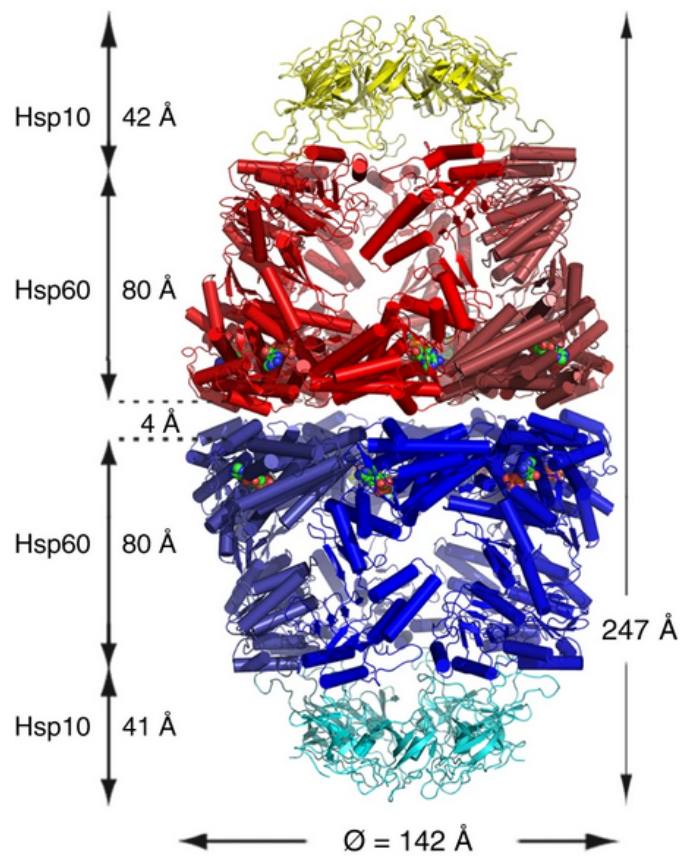


Figura 2.5: **Organización estructural del complejo de chaperoninas HSP60-HSP10** [8]. En azul y rojo se muestran los anillos heptaméricos cis y trans de la HSP60, mientras que en amarillo y cian se muestran los anillos heptaméricos en forma de capucha de la co-chaperonina HSP10. De acuerdo a Gomez-Llorente, Y. y col., la HSP60 se ensambla como oligómeros tetradecaméricos en forma de barril que están organizados en dos anillos heptaméricos. Aquí, las paredes internas del cilindro proporcionan puntos de unión para proteínas desplegadas a través de parches hidrofóbicos expuestos en los extremos de la cavidad [24]. La reacción de plegamiento de proteínas para este complejo requiere la unión de ATP y la hidrólisis por cada subunidad de la cochaperonina HSP10. La HSP10 forma anillos heptaméricos que unen a ambos extremos del cilindro de una manera dependiente del ATP. En virtud de ello, la HSP10 proporciona un entorno aislado para el plegamiento de proteínas, evitando así la agregación prematura de las proteínas sustrato.

## 2.3. Retos computacionales

El estudio del problema del plegamiento de proteínas también ha significado un problema computacional. Aquí el problema consiste en la capacidad de desarrollar algoritmos capaces de predecir la estructura nativa de una proteína, dada la secuencia de aminoácidos. La creación de bases de datos ha sido un paso importante para el desarrollo de diversos algoritmos bioinformáticos, los cuales se basan en la comparación de estructuras como los modelos de homología de secuencias o *template-based* [6]. Recientemente, la inteligencia artificial ha llegado por la puerta grande a este campo. Apenas el año pasado, el concurso conocido por sus siglas: CASP (*Critical Assessment of Protein Structure Prediction*) fue el escenario de todo un hito en su decimotercera edición, cuando uno de las organizaciones que participaban, DeepMind, quedó en primer lugar y con una amplia diferencia respecto al resto. DeepMind logró desarrollar un nuevo algoritmo basado en el aprendizaje profundo capaz de analizar, extraer y explotar los datos sobre la estructura tridimensional de las proteínas contenido en el Banco de Datos de Proteínas (PDB por sus siglas en inglés). El programa, bautizado como *AlphaFold 2*, se destacó particularmente en las predicciones de estructuras para las que no existían modelos previos, consideradas por los organizadores de la competición como las más difíciles. En general, aunque el logro técnico de *AlphaFold 2* ha recibido un reconocimiento generalizado, la exactitud de las predicciones aún no es lo suficientemente buena en un tercio de los casos y el algoritmo no revela ninguna información sobre el mecanismo del plegamiento de las proteínas [25,26].

Por otro lado, se han desarrollado algoritmos basados en modelos físicos como los de dinámica molecular y Monte Carlo. Sin embargo, el principal problema en estos métodos es que son muy exigentes en términos de potencia de cálculo y tiempo. Se han propuesto métodos alternativos que reducen la necesidad de cálculos enormes; estos incluyen “simplificar” moléculas al reducir el número de componentes (como en *Coarse Grain MD*). Asimismo, se ha sugerido la posibilidad de implementar el modelo de potenciales externos en conjunto con campos de fuerza que permitan reducir el tiempo computacional requerido. Estos cumplirían principalmente dos funciones: (i) el de introducir efectos del medio en el proceso

de plegamiento como el efecto hidrofóbico o la presencia de proteínas chaperonas y (ii) el de acelerar la exploración del espacio conformacional para ahorrar tiempo computacional. Se han obtenido resultados exitosos al implementar potenciales externos en el plegamiento de polímeros orgánicos modelo y algunos péptidos pequeños (20-35 aminoácidos) [6]; sin embargo, aun resulta necesario desarrollar, mejorar y aplicar prácticas alternativas que permitan optimizar eficazmente los algoritmos de dinámica molecular.

### 2.3.1. Dinámica molecular

La metodología de dinámica molecular permite simular computacionalmente la evolución temporal de sistemas moleculares dentro del marco de la mecánica clásica. En cada paso de dinámica molecular se calculan las fuerzas que actúan sobre todas las partículas y se integran (resuelven) las ecuaciones de movimiento dadas por la segunda ley de Newton, generalmente en intervalos de 1 a 2 femtosegundos (fs):

$$m_i \ddot{\vec{r}}_i = \vec{f}_i(\vec{r}_1, \dots, \vec{r}_N), \quad (2.1)$$

donde  $\vec{f}_i$  es la fuerza total ejercida sobre la partícula  $i$ ,  $m$  su masa y  $\vec{r}_i$  su posición. Las coordenadas de las partículas calculadas en cada paso junto con sus velocidades y fuerzas específicas un punto en el espacio-fase. Al final de la simulación, tenemos un conjunto de puntos en el espacio-fase llamado trayectoria, desde la cual pueden conocerse las propiedades macroscópicas y otros datos relevantes del sistema en todo instante.

Si el conjunto de ecuaciones resultantes en [2.1] se aplican a un sistema en el que todas las fuerzas son conservativas, la fuerza sobre cada partícula puede obtenerse a partir del gradiente del potencial,

$$\vec{f}_i(\vec{r}_1, \dots, \vec{r}_N) = -\nabla V(\vec{r}_1, \dots, \vec{r}_N). \quad (2.2)$$

La energía potencial  $V$  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 proteí-

nas, 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, su forma funcional involucra varios parámetros que deben ser ajustados para poder obtener resultados precisos [6]. Los detalles específicos relacionados con esta metodología se discutirán en el Capítulo 4.

## Objetivos e hipótesis

El objetivo principal de este trabajo consiste en caracterizar el ambiente fisicoquímico, y en particular el potencial electrostático de largo alcance, en el interior de la chaperonina HSP60-HSP10 sola y de la chaperonina HSP60-HSP10 en complejo con una proteína pequeña y, con ello, obtener el punto de partida sobre cómo representar el efecto que una chaperonina ejercería sobre el plegamiento de una proteína, sin necesidad de incluirla explícitamente en simulaciones de dinámica molecular.

### 3.1. Objetivos particulares

- Construir sistemas de estudio adecuados para representar los efectos electrostáticos de largo alcance en el interior de la chaperonina HSP60-HSP10.
- Efectuar simulaciones de dinámica molecular de la HSP60-HSP10 sola y de la HSP60 en complejo con CitC.
- Analizar la evolución temporal de la distribución de potencial electrostático de largo alcance de cada sistema de estudio con la ayuda de herramientas de análisis existentes y otras implementadas *ex profeso*.

## **3.2. Hipótesis**

La distribución del potencial electrostático de largo alcance en el interior de una chaperonina es invariante en el tiempo y, espacialmente, no es aleatoria.

## Herramientas y métodos

En este trabajo se aplicó un método de muestro de la distribución de potencial electrostático de largo alcance de la proteína HSP60 en diferentes sistemas de estudio. El método propuesto integra el algoritmo de malla de partícula de Ewald (PME) dentro del marco de la dinámica molecular clásica. A continuación, revisaremos el formalismo básico de algunos temas más importantes relacionados con la dinámica molecular clásica y abordaremos brevemente el método desarrollado por nuestro grupo para llevar a cabo esta tesis.

### 4.1. Formalismos básicos de dinámica molecular clásica

#### 4.1.1. Esquemas de integración

En la metodología de dinámica molecular lo que se analiza es la evolución temporal del sistema molecular considerado. De modo que, la parte central de este método lo constituye el algoritmo de integración de las ecuaciones de movimiento de Newton, que son ecuaciones diferenciales ordinarias, no-lineales y de segundo orden, que deben ser resueltas numéricamente. Dadas las posiciones y velocidades a un tiempo inicial  $t_0$ , la tarea del algoritmo es, por tanto, obtener las posiciones iniciales y velocidades en el tiempo  $t_0 + \Delta t$ , donde  $\Delta t$  es el paso de tiempo (del inglés *time step*) en la simulación de dinámica molecular [27].

Existen numerosos algoritmos para integrar las ecuaciones de movimiento de Newton. Todos ellos convierten las ecuaciones diferenciales en ecuaciones en diferencias finitas. Uno de los algoritmos más comúnmente utilizado para la integración de las ecuaciones de movimiento es el algoritmo de Verlet también referido como Störmer Verlet. Para deducirlo, partimos del desarrollo en serie de Taylor de la posición,  $\vec{r}_i(t)$ , de una partícula  $i$  de masa  $m_i$  y velocidad  $\vec{v}_i(t)$ :

$$\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \vec{v}_i(t)\Delta t + \frac{\vec{a}_i(t)}{2!}(\Delta t)^2 + \frac{1}{3!}\frac{d^3\vec{r}_i}{dt^3}(\Delta t)^3 + O((\Delta t)^4), \quad (4.1)$$

donde la aceleración es  $\vec{a}_i(t) = \frac{\vec{f}_i(t)}{m_i}$ . Del mismo modo,

$$\vec{r}_i(t - \Delta t) = \vec{r}_i(t) - \vec{v}_i(t)\Delta t + \frac{\vec{a}_i(t)}{2!}(\Delta t)^2 - \frac{1}{3!}\frac{d^3\vec{r}_i}{dt^3}(\Delta t)^3 + O((\Delta t)^4). \quad (4.2)$$

Sumando ambos desarrollos, obtenemos

$$\vec{r}_i(t + \Delta t) + \vec{r}_i(t - \Delta t) = 2\vec{r}_i(t) + \vec{a}_i(t)(\Delta t)^2 + O((\Delta t)^4). \quad (4.3)$$

Así, la nueva posición  $\vec{r}_i$  buscada, en el tiempo  $t + \Delta t$ , viene dada por

$$\vec{r}_i(t + \Delta t) = 2\vec{r}_i(t) - \vec{r}_i(t - \Delta t) + \vec{a}_i(t)(\Delta t)^2 + O((\Delta t)^4). \quad (4.4)$$

Para este caso, el error estimado para la nueva posición  $\vec{r}_i$  es del orden de  $(\Delta t)^4$ . Asimismo, de (4.4) puede verse que para evaluar la nueva posición  $\vec{r}_i(t + \Delta t)$  sólo necesitamos conocer la posición anterior en  $\vec{r}_i(t - \Delta t)$  y la aceleración en el tiempo  $t$ ; no se necesita la velocidad. Sin embargo, ésta la podemos calcular a partir de representar la diferencia entre la ecuación (4.1) y (4.2) como sigue,

$$\vec{r}_i(t + \Delta t) - \vec{r}_i(t - \Delta t) = 2\vec{v}_i(t)\Delta t + O((\Delta t)^3). \quad (4.5)$$

donde  $\vec{v}_i(t)$  es igual a

$$\vec{v}_i(t) = \frac{\vec{r}_i(t + \Delta t) - \vec{r}_i(t - \Delta t)}{2\Delta t} + O((\Delta t)^3). \quad (4.6)$$

Aquí, el error en la velocidad es menor, del orden  $(\Delta t)^3$ .

Se han desarrollado algoritmos de integración más precisos como *velocity Verlet* y *leap-frog*, que resuelven detalles relacionados con la aproximación de las velocidades [28]. Aún así, todos ellos son métodos aproximados, y debe tenerse en cuenta que el error de estimación siempre aumenta rápidamente con el paso de tiempo. La elección del algoritmo es, por lo tanto, fruto de un compromiso entre el grado de precisión requerido y el esfuerzo computacional.

#### 4.1.2. Función de energía potencial

Las interacciones entre los constituyentes de un sistema molecular pueden describirse de manera precisa mediante una función de energía potencial  $V$  que dependa de la configuración del sistema. En general, la energía potencial  $V$  de un sistema de  $N$  átomos puede expresarse en función de las coordenadas de los átomos individuales, de la distancia entre ellos, de la posición relativa entre átomos, etc. Cada uno de estos términos pueden verse como un conjunto de contribuciones de tipo dispersivas, electrostáticas, etc. denominado *campo de fuerza*. En la actualidad existen diferentes campos de fuerza diseñados para diversos propósitos. Un resumen de los diferentes modelos de campos de fuerzas pueden consultarse en [29]. Por poner un ejemplo, en (4.7) puede observarse la forma funcional del campo de fuerza OPLS (del inglés, *Optimized Potentials for Liquid Simulations*), el cual es ampliamente utilizado en simulaciones de sistemas biomoleculares y orgánicos [30-33].

$$\begin{aligned}
 V(\vec{r}_1, \dots, \vec{r}_N) = & \sum_{\text{enlaces}} \frac{1}{2} k_b (\vec{r} - \vec{r}_0)^2 + \sum_{\text{ángulos}} k_\theta (\theta - \theta_0)^2 + \sum_{\text{torsiones}} \sum_n \frac{1}{2} V_i^n [1 + \cos(n\phi_i - \phi_i)] + \\
 & + \sum_{i=1}^{N-1} \sum_{j=i+1}^N \left\{ 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} + \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j e^2}{r_{ij}} \right\} f_{ij}
 \end{aligned}
 \tag{4.7}$$

Para tal caso, el primer y segundo término representan la energía entre los átomos unidos covalentemente y la energía debida a la geometría de los orbitales electrónicos involucrados en el enlace covalente, respectivamente. Ambos términos son descritos por potenciales

armónicos, ya que son muy fuertes y fluctúan solo ligeramente alrededor de sus valores de equilibrio a temperatura ambiente. El tercer término representa la energía para torcer un enlace debido al orden de enlace (por ejemplo, enlaces dobles) y enlaces vecinos o pares solitarios de electrones. Este último es descrito mediante una expansión de la función coseno y puede tomar cualquier valor dentro de  $360^\circ$  dependiendo de la altura de la barrera entre las conformaciones de baja energía, lo que hace que la precisión de la barrera del potencial diedro sea crucial para algunas propiedades de polímeros. Finalmente, el cuarto término representa la energía no enlazada entre todos los pares de átomos, que puede descomponerse en energías de van der Waals (primer término de la suma) y electrostáticas (segundo término de la suma), descritos mediante el potencial de Lennard Jones y de Coulomb [34].

Una característica distintiva de los parámetros de OPLS es que se optimizaron para adaptarse a las propiedades experimentales de los líquidos, como la densidad y el calor de vaporización, además de instalar perfiles de torsión en fase gaseosa. La descripción correcta de estas propiedades hace que OPLS y sus variantes, por ejemplo, OPLS-*all atoms* y OPLS-*united atoms*, representen bien las cantidades de los términos relacionados a los parámetros no enlazantes, incluido el enlace por puente de hidrógeno y el tamaño de la molécula [35].

### 4.1.3. Condiciones de frontera

La forma clásica de eliminar los efectos de borde de un sistema de volumen finito durante la ejecución de una trayectoria de dinámica molecular es aplicar condiciones de contorno periódicas. Para ello, los átomos del sistema a simular se colocan en una caja de simulación de tamaño  $L$ , que está rodeada de una red infinita y periódica de celdas idénticas de sí misma (ver Figura 4.1.)

De esta forma, una partícula no sólo interacciona con el resto de las  $N - 1$  partículas en la caja de simulación, sino también con todas sus imágenes periódicas en las réplicas de la caja de simulación. Así, en ausencia de fuerzas externas y si se asume que el potencial es aditivo por pares ( $v_2$ ), se obtiene que la energía potencial del sistema de  $N$  partículas en cualquier

celda puede expresarse como,

$$V(\vec{r}_1, \dots, \vec{r}_N) = \frac{1}{2} \sum_{i,j,\vec{n}} 'v_2(|\vec{r}_{ij} + \vec{n}L|), \quad (4.8)$$

donde  $\vec{n} = (n_x, n_y, n_z)$  siendo  $n_x, n_y$  y  $n_z$  números enteros arbitrarios, y el símbolo prima indica que para el caso  $i = j$  no se considera  $\vec{n} = 0$  [27].

La suma en (4.8) se realiza empleando la convención de imagen mínima, que consiste en considerar que cada partícula interactúa sólo con las imágenes de las otras  $N - 1$  partículas cercanas a ellas, tal y como se puede observar en la Figura 4.1 donde una partícula dada interacciona con las ubicadas en una caja dibujada en línea discontinua centrada en dicha partícula. En la práctica, la aplicabilidad de tal enfoque depende en gran medida de las propiedades del sistema bajo investigación y, en particular, del alcance del potencial de interacción. Por ejemplo, si se trata de potenciales de corto a moderado alcance se suelen trincar las interacciones, haciéndolas nulas más allá de un radio de corte  $r_c$ . La situación es enteramente diferente cuando los átomos en el sistema presentan cargas o dipolos eléctricos permanentes. Aquí las interacciones electrostáticas resultantes no pueden ser truncadas. En este caso, se utilizan métodos especializados como la suma de Ewald o la malla de partículas de Ewald (*particle mesh Ewald*, PME), ver sección 4.2.4. Si bien el concepto se aplicaría estrictamente a un sistema cristalino, este ha mostrado ser de gran utilidad para el estudio de propiedades dinámicas de sistema como las proteínas solvatadas, y serán las condiciones de fronteras usadas en este trabajo.

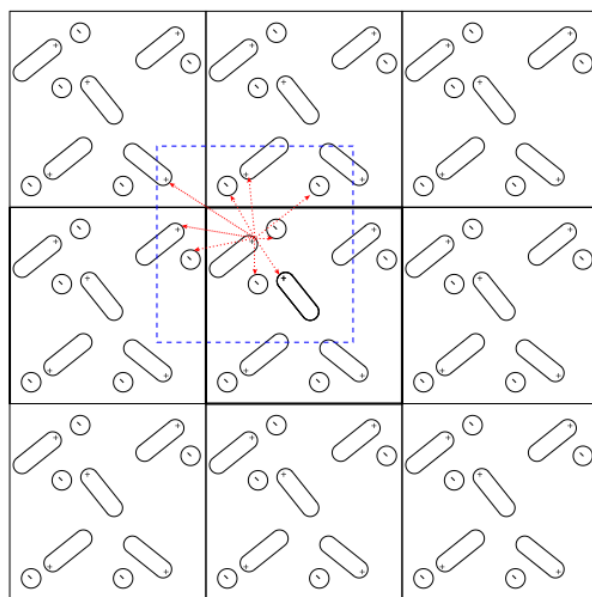


Figura 4.1: **Ejemplo de condiciones de contorno periódicas bidimensional** [36]. La caja central se ha representado en trazo grueso, mientras que las partículas que se encuentran dentro de la caja de líneas discontinuas son las que interactúan con la partícula que se encuentra en su centro según la convención de mínima imagen. Las interacciones se representan por las flechas de trazo rojo.

#### 4.1.4. Equilibración: ajuste de temperatura y presión

En termodinámica estadística, un *ensamble* se refiere al conjunto de estados microscópicos descritos por un conjunto de variables de estado macroscópicamente observables [29]. En una simulación de dinámica molecular, si no se impone alguna restricción mediante algoritmos adicionales sobre el sistema molecular considerado, por defecto, el número de partículas ( $N$ ), el volumen ( $V$ ) y la energía total ( $E$ ) se mantienen constantes; haciendo, al ensamble microcanónico ( $NVE$ ), el conjunto natural de la técnica de dinámica. Sin embargo, si se desea simular condiciones parecidas a las condiciones experimentales de procesos biológicos, deberán emplearse los ensambles canónico ( $NVT$ ) e isotérmico-isobárico ( $NPT$ ), donde además de mantener  $N$  y  $V$  constantes, también la temperatura ( $T$ ) y la presión ( $P$ ).

Durante la trayectoria de dinámica molecular, la temperatura de un sistema de  $N$  partículas, que se mueven en un espacio  $d$ -dimensional, puede determinarse mediante la expresión:

$$T = \frac{1}{d(N-1)k_B} \sum_{i=1}^N m_i v_i^2, \quad (4.9)$$

donde  $k_B$  es la constante de Boltzmann,  $T$  la temperatura,  $d(N - 1)$  los grados de libertad del sistema y  $m_i$  y  $v_i$  la masa y velocidad del átomo  $i$ . Por tanto en un sistema con  $N$  constante, es necesario incidir sobre las velocidades de los átomos. El procedimiento más sencillo es reescalar las velocidades a modo de mantener el valor de la temperatura constante. Para ello, la estrategia mas apropiada es “acoplar” al sistema a un “baño térmico”. Mediante este acople se lleva la temperatura del sistema a la del baño mediante una relajación de tipo exponencial que sigue la cinética de primer orden de la forma:

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau}, \quad (4.10)$$

donde  $T_0$  es la temperatura de referencia (del baño externo),  $T$  la temperatura instantánea calculada en la etapa de dinámica molecular y  $\tau$  una constante. Las velocidades se reescalan multiplicando aquellas obtenidas en la etapa de dinámica por un factor  $\lambda$ :

$$v'_i = \lambda v_i. \quad (4.11)$$

Eligiendo el valor de  $\lambda$  se puede elegir la “fuerza” del acoplamiento al baño: a mayores valores de esta constante, se obtienen acoplamientos más débiles y, viceversa.

Para el caso en el que se desea mantener la presión constante en el sistema, se procede de manera análoga al caso de la temperatura, pero ahora reescalando las posiciones de los átomos. De esta manera, el cálculo de la presión durante la trayectoria de dinámica molecular se puede obtenerse mediante la expresión,

$$P = \frac{2}{3} V (E_k + \frac{1}{2} \sum_{i < j} \vec{r}_{ij} \vec{f}_{ij}), \quad (4.12)$$

donde  $V$  es el volumen,  $E_k$  la energía cinética,  $\vec{r}_{ij}$  el vector que une al átomo  $i$  con el  $j$  y  $\vec{f}_{ij}$  es la fuerza ejercida por el átomo  $j$  sobre el átomo  $i$  [37].

En todo lo referido hasta ahora, se ha supuesto que se tienen las coordenadas cartesianas de los diversos componentes de los sistema bajo consideración en el tiempo inicial. Estos datos puede obtenerse a partir de experimentos de cristalografía de rayos X o, en situaciones

menos comunes, mediante métodos de modelado molecular.

## 4.2. Metodología

El enfoque de trabajo utilizado para la ejecución de las simulaciones de dinámica molecular se basó en la paquetería GROMACS 18.3 [38,39].

### 4.2.1. Preparación de los sistemas de estudio

El enfoque típico utilizado por GROMACS para la preparación del sistema considerado consiste en construir la topología para el soluto y/o proteína(s) de interés, que luego se modifica o actualiza para incluir otros componentes que se agregan al sistema (agua, iones, ligandos, etc.). Para nuestros propósitos, se tomó la estructura cristalina del complejo HSP60-HSP10 mitocondrial humano [7] y procesó de manera que se pudiera representar el análisis de la propiedad física deseada en diferentes sistemas moleculares de estudio. A este respecto, cada sistema molecular se etiquetó de acuerdo a lo que representaba: sistema HSP60, cuyo único componente sería la estructura de la HSP60 (vacío en nuestro ejemplo artificial), sistema HSP60-HSP10, cuyo componente sería el complejo entre la HSP60 y la HSP10 (también vacío en nuestro ejemplo artificial y cuyo ejemplo puede verse representado en la Figura 2.5) y el sistema HSP60-HSP10-CitC, cuyos componentes además del complejo entre la HSP60 y la HSP10, incluiría en su interior a la proteína CitC. Esta última se utilizó, bajo ciertas consideraciones [4], para un fin determinado, el de representar una perturbación en la distribución de potencial electrostático a lo largo de la trayectoria de la HSP60.

La generación de la estructura de partida del sistema HSP60-HSP10-CitC representó un reto importante ya que requirió de un paso previo de preparación respecto al resto de los sistemas de estudio. Primero, se procesaron las estructuras del complejo HSP60-HSP10 mitocondrial humano (PDB ID: 6MRC) [8] y de la proteína CitC (PDB ID: 3ZCF) [47] para eliminar

<sup>1</sup>La proteína CitC se seleccionó de acuerdo a dos criterios importantes: (1) que su modelo de plegado se ajustará a los discutidos en la sección 2.1 [40-46] y (2) que su estructura tridimensional no fuera mayor a 100 aminoácidos, de manera que representará el menor coste computacional en las simulaciones de dinámica molecular

iones metálicos y moléculas de solvente. Después cada estructura debidamente procesada se colocó en una celda de simulación triclinica configurada de manera que el centro geométrico de la(s) proteína(s) fuese el mismo con respecto a un mismo sistema de referencia [48]. En este punto, se generó la topología de cada una empleando el campo de fuerzas OPLS-*aa* [30] y, con ambas topologías, se construyó una topología única que incluiría los parámetros necesarios y otras definiciones moleculares para que GROMACS considerara a este conjunto de proteínas como una estructura de partida adecuada. Los detalles relacionados a la configuración del archivo de topología para el sistema HSP60-HSP10-cytC pueden revisarse en el tutorial de GROMACS *Protein-Ligand Complex* [49].

#### **4.2.2. Establecimiento de la configuración inicial y definición de las condiciones de equilibrio**

Con las estructuras en el formato correcto para el campo de fuerza elegido, así como la topología correspondiente, el siguiente paso en la construcción de un sistema de proteínas solvatadas fue definir un volumen alrededor de la(s) proteína(s) que será llenado con agua. La justificación por la que estos métodos deban emplearse pueden consultarse en la sección 4.1.3. Bajo esta conjetura, cada sistema de estudio se colocó en una celda de solvente explícito lo suficientemente grande empleando el modelo de moléculas de agua SPCE desarrollado por Berendsen y col. [50]. La práctica más convencional para un sistema como este es centrar a la(s) proteína(s) en una caja que satisfagan la convención de imagen mínima periódica. Existen muchas formas de caja que pueden usarse, en nuestro caso se utilizó una caja triclinica de tipo dodecaédrica que aunque no representa menor complejidad al tener un volumen más pequeño comparada por ejemplo, con una caja cúbica, satisface muy bien la convención de imagen mínima periódica (ver sección 4.1.3). Por otro lado, como es bien sabido, las moléculas por sí solas (proteínas, iones, etc.) pueden tener una carga formal asociada y, aunque esta convención refleja las condiciones experimentales típicas, las simulaciones de dinámica molecular suelen ser más relevantes cuando se consideran condiciones de electroneutralidad. Mas allá de este punto, las interacciones electrostáticas como resultado de la

interacción entre dos o más partículas cargadas en un sistema molecular suelen evaluarse con mayor frecuencia utilizando el método de malla de partículas Ewald (PME) (ver sección 4.2.4) que requiere un sistema electroneutral. Para conseguir las condiciones de electroneutralidad requeridas, se agregó una cantidad relevante de iones monoatómicos de sodio a la solución acuosa para contrarrestar cualquier carga neta de la(s) proteína(s) presente(s) (ver Tabla 4.1).

Con los sistemas solvatados y eléctricamente neutros, el último paso en la construcción de coordenadas incluyó un proceso de relajación/equilibración que tendría dos propósitos: (1) obtener un estado de baja energía empleando un proceso de minimización energética y (2) un estado termodinámico estable, en el que cada estructura energéticamente equilibrada se ajustó a una temperatura y presión constantes (ver sección 4.1.4).

Sistema de estudio	Volumen de la celda de simulación	No. de moléculas de agua SPCE	No. de iones monoatómicos de sodio
Sistema HSP60	7,759.23 nm <sup>3</sup>	219,095	126
Sistema HSP60-HSP10	14,602.40 nm <sup>3</sup>	436,893	98
Sistema HSP60-HSP10-CitC	14,602.36 nm <sup>3</sup>	436,498	89

Tabla 4.1: Resumen de parámetros para la construcción de las configuraciones iniciales para los sistemas moleculares de estudio de la HSP60.

### 4.2.3. Producción de las simulaciones de dinámica molecular

El esfuerzo computacional general involucró simulaciones de sistemas con alrededor de 500,000 átomos para 100 ns con disolvente explícito a condiciones estándar de 310 K y 1 bar. Las variables de temperatura y presión se controlaron empleando los algoritmos de Berendsen modificado [51] y de Parrinello-Rahman [52]. Las energías de Van der Waals se calcularon utilizando un corte suave de 10 Å. Las interacciones electrostáticas en los átomos de las proteínas se evaluaron utilizando el método de Ewald de malla de partículas (PME). La integración de las ecuaciones de movimiento mecánicas se realizó empleando el algoritmo *leap-frog* [53] y un paso de tiempo de 2 fs, cuya característica más importante es que este aproxima mejor el cálculo de las velocidades sin perder precisión (ver sección 4.1.1). En todos los casos se utilizaron condiciones de contorno periódicas.

#### 4.2.4. Análisis de datos

El análisis de datos contempló dos puntos importantes: (1) probar la estabilidad estructural de las proteínas en los diferentes sistemas de estudio y (2) representar adecuadamente las propiedades electrostáticas de estos sistemas. En virtud de ello, se consideraron las siguientes herramientas.

##### **Análisis estructural: Desviaciones medias cuadráticas**

En el estudio de las conformaciones de proteínas, se mide habitualmente la similitud en la estructura tridimensional mediante la desviación cuadrática media (o RMSD por sus siglas en inglés) de las coordenadas atómicas de proteínas superpuestas. Es muy útil para mostrar movimientos estructurales en las diferentes regiones del sistema molecular, incluso si no conllevan a energías diferentes, o para evaluar el movimiento relativo del ligando con respecto a la proteína, por mencionar dos de las aplicaciones más comunes.

Para calcular el RMSD de un par de estructuras (por ejemplo,  $x$  y  $y$ ), cada estructura debe representarse como un vector de coordenadas de longitud  $3N$  (suponiendo  $N$  átomos). El RMSD se obtiene a partir de calcular la raíz cuadrada del promedio de las distancias al cuadrado entre los átomos correspondientes de  $x$  y  $y$ ,

$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^N |x_i - y_i|^2} \quad (4.13)$$

Sin embargo, cuando se toman muestras de conformaciones moleculares a partir de dinámicas moleculares u otras formas de muestreo, suele darse el caso de que la molécula se aleje del origen y rote de forma arbitraria.

Para compensar estos problemas, suele utilizarse la desviación cuadrática media de la raíz mínima (o lRMSD por sus siglas en inglés) que tiene como objetivo obtener el mínimo RMSD sobre todas las posiciones y orientaciones relativas posibles de las dos conformaciones bajo consideración. Para representarla, primero debe realizarse una “alineación óptima” de las dos estructuras eliminando la traslación del centro de masa y, luego, calcular su RMSD me-

diante (4.13) [37, 54].

### Análisis y representación de las propiedades electrostáticas: Potencial de malla de partículas de Ewald (PME)

El método PME, desarrollado por Darden y col. [55, 56], es uno de los métodos más útiles para manejar el carácter de largo alcance de las interacciones electrostáticas (representadas por la ley de Coulomb) en sistemas moleculares periódicos de gran tamaño. En Gromacs, el último término de la ecuación (4.7) relacionado con los átomos que están fuera del radio de corte, se reemplaza por el PME, el cual describimos a continuación.

El algoritmo PME adopta el esquema de dividir la energía electrostática total ( $V$ ) de un sistema de  $N$  partículas cargadas en una caja, la cual se replica en todas direcciones, en una serie de distribuciones (usualmente gaussianas) de carga iguales a las puntuales del sistema relacionadas con una suma directa en el espacio cartesiano representada por  $V_{dir}$ , una suma recíproca en el espacio de Fourier representada por  $V_{rec}$  y un término constante  $V_0$ , comúnmente referido como potencial de corrección.

$$V = V_{dir} + V_{rec} + V_0, \quad (4.14)$$

donde,

$$V_{dir} = \frac{f}{2} \sum_{\vec{n}} \sum_i \sum_j q_i q_j \frac{\text{erfc}(\beta |\vec{r}_j - \vec{r}_i + \vec{n}|)}{|\vec{r}_j - \vec{r}_i + \vec{n}|}, \quad (4.15)$$

$$V_{rec} = \frac{f}{2\pi V} \sum_i \sum_j q_i q_j \sum_{\vec{m}} \frac{\exp(-(\pi \vec{m} \beta)^2 + 2\pi \vec{m} \cdot (|\vec{r}_i - \vec{r}_j|))}{\vec{m}^2}, \quad (4.16)$$

$$V_0 = -\frac{f\beta}{\sqrt{\pi}} \sum_i q_i^2, \quad (4.17)$$

erfc la función error complementaria

$$\text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty \exp(-t^2) dt, \quad (4.18)$$

y  $\beta$  el parámetro de la distribución gaussiana usado de modo

$$\rho_i(\vec{r}) = q_i \left( \frac{\beta}{\sqrt{\pi}} \right)^3 e^{-\beta^2 |\vec{r} - \vec{r}_i|^2}. \quad (4.19)$$

La ventaja de la expresión (4.14) es que los términos  $V_{\text{dir}}$  y  $V_{\text{rec}}$  pueden truncarse. Esto es, hay un valor óptimo de  $\beta$  para el que el efecto de truncamiento de ambos términos se minimiza, demostrándose que el costo computacional de evaluar las interacciones electrostáticas con esta metodología escala como  $N \log(N)$  (calcular directamente la contribución electrostática con suma de cargas escala como  $N^2$ ). Una revisión detallada de las técnicas existentes para representar los términos mostrados en (4.15) está dada por Toukmaji y Board (57).

Para nuestros fines, nos limitaremos a revisar el método de cálculo subyacente de una versión modificada del *plugin* PMEpot de VMD utilizado como herramienta para muestrear la distribución de potencial electrostático de largo alcance entre los átomos de la HSP60.

De acuerdo con Aksimentiev, A. y col. (58), PMEpot evalúa el término recíproco del método PME mostrado en (4.16), en cada punto  $\vec{r}_{j,k}$  de una malla tridimensional produciendo una cuadrícula de potencial electrostático suavizado. El suavizado es un resultado natural del método PME de aproximar la distribución de carga que cada átomo genéricamente cargado ( $q_i$ ) ejerce sobre cada punto de la cuadrícula ( $\vec{r}_{j,k}$ ) mediante una función gaussiana esférica de amplitud inversa como la mostrada en (4.19). Eligiendo adecuadamente la distancia entre los puntos de la cuadrícula ( $\Delta = \vec{r}_{j+1,k} - \vec{r}_{j,k}$ ) se puede entonces obtener una imagen refinada de la electrostática de largo del sistema molecular en cuestión. El plugin PME es, además, compatible con los programas estándar de visualización molecular. Cabe señalar, además, que el cálculo de  $V_{\text{rec}}$  resulta ser un proceso dinámico, ya que durante las simulaciones de dinámica molecular todos los átomos dentro de la caja de simulación se mueven libremente; se deduce, por lo tanto, que su contribución a  $V_{\text{rec}}$  será diferente, *en principio*, en cada instante de tiempo de la simulación de dinámica molecular. Tomando en cuenta este problema, para determinar la distribución de potencial electrostático de largo alcance promedio con respecto al tiempo ( $\bar{V}_{\text{rec}}$ ), se tomó una muestra espacial representativa de potenciales electrostáticos instantáneos  $V_{\text{rec}}$  calculados con PMEpot y se promedió en un intervalo de tiempo de la

trayectoria de dinámica molecular. Para este fin, desarrollamos y empleamos herramientas (*genAveragePME*, *manipdx*) exclusivamente diseñadas para este propósito en el Laboratorio de Software y Modelado Matemático, dirigido por el Dr. Juan Manuel Solano Altamirano. Para mayores detalles ver Apéndice A.

## Resultados y discusión

Como se mencionó en la introducción, en trabajo futuro, nos interesará extraer la esencia de los efectos que ejercen las chaperoninas en el plegado de proteínas. Esto significa, a nivel microscópico, que buscamos encontrar la esencia fisicoquímica que pudiera ejercer la chaperonina. De ser cierto que se pueda extraer información al respecto, y con el fin de añadirla en una simulación de DM (sin incluir la chaperonina en sí), **idealmente** la contribución debería tener características globales (es decir en un sentido matemático no local) e invariantes en el tiempo. Por supuesto, esta es una conjetura, y en este trabajo buscamos evidencia de su veracidad o falsedad. Para lograr lo anterior, comenzamos analizando la estabilidad estructural de la chaperonina. Posteriormente, analizamos el campo electrostático de largo alcance a lo largo y ancho del espacio ocupado por la chaperonina, de manera que nos de una idea de cómo se distribuye espacialmente  $V_{rec}$  y de cómo cambia, si lo hace, en el tiempo.

## 5.1. Estabilidad estructural de la HSP60 en los diferentes sistemas moleculares de estudio

Después de preparar las posiciones iniciales de los sistemas, según lo descrito en la sección 4.2, y realizar las simulaciones de DM, se analizaron las trayectorias mediante el cálculo de la desviación cuadrática media o RMSD (ver sección 4.2.4) tomando como referencia la estructura experimental, libre de iones y moléculas de solvente cristalizadas. En cada caso, se cambió la representación de la periodicidad y se eliminó la traslación del centro de masa de todos los átomos en la(s) proteína(s).

A partir de la Figura 5.1, queda claro que las estructuras de las proteínas en cada sistema de estudio permanecen estables durante la mayor parte de la simulación. Para cada caso, se pudo observar que el RMSD converge rápidamente hacia un valor de distancia objetivo (alrededor de los 40 ns) y permanece estable durante el resto de la trayectoria de dinámica molecular. Como consecuencia de ello, la estructura molecular del sistema HSP60 mostró la menor variación en su distancia promedio respecto a la estructura de referencia (~ 0.5 nm), seguido de los sistemas HSP60-HSP10 (~ 0.55 nm) y HSP60-HSP10-CitC (~0.57 nm), lo cual resulta interesante considerando que la estabilidad estructural de la HSP60, cuando se encuentran unida a otras proteínas (tal como ocurre en la célula) no cambia substancialmente. Estos resultados, sugieren, por lo tanto, que la estructura molecular de la HSP60, a tiempos mayores 100 ns, podría permanecer inalterada, independientemente del número y tamaño de proteínas a las que se encuentre asociada (como en el caso de los sistemas HSP60-HSP10 y HSP60-HSP10-CitC). En este sentido, valdría la pena simular sistemas HSP60-HSP10-X, donde X es una proteína de mayor tamaño que CitC, para determinar la posibilidad de que la HSP60 pueda verse alterada al albergar proteínas de mayor tamaño.

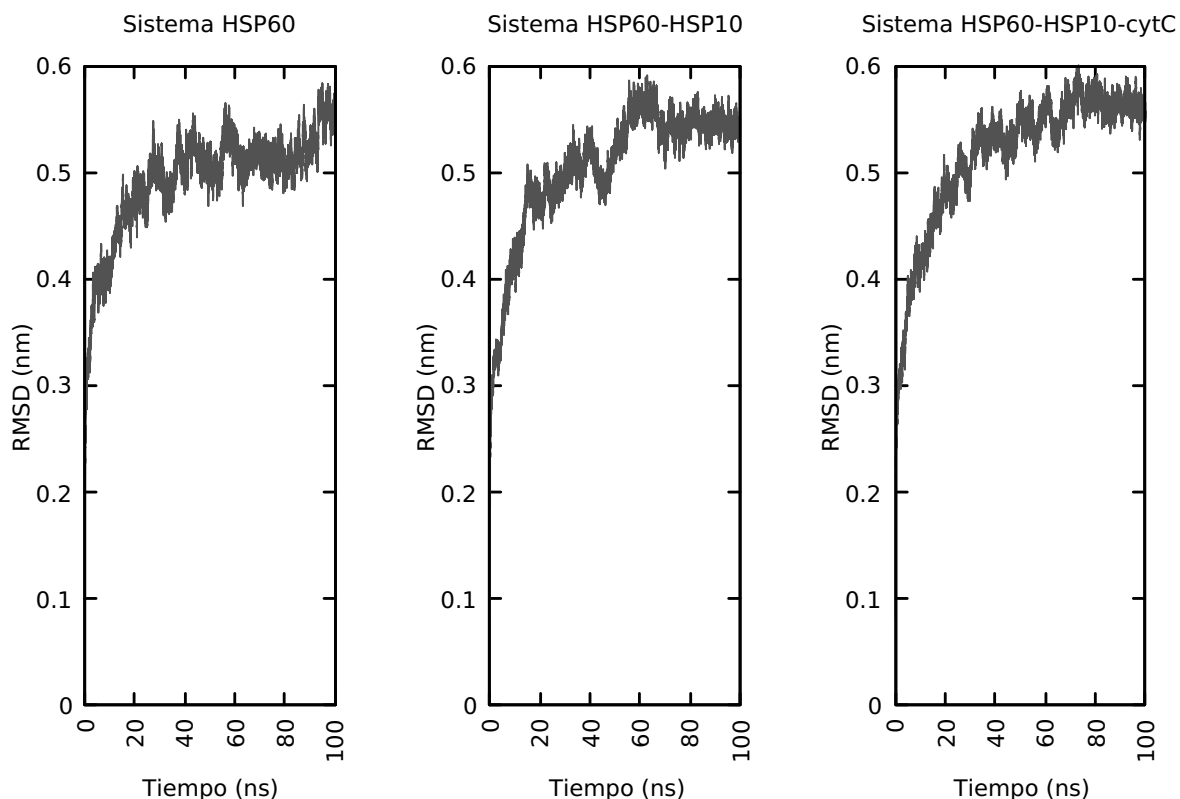


Figura 5.1: Evolución temporal de la desviación cuadrática media (RMSD) para cada uno de los sistemas moleculares de estudio de la HSP60.

Finalmente, para determinar si la HSP60 ejerce un efecto importante sobre la estructura de CitC, ésta última se sometió a una simulación de DM individual y se analizó el RMSD (ver Figura 5.2). Para este sistema, el análisis del RMSD mostró que la estructura de CitC dista de alcanzar rápidamente un valor de distancia objetivo ( $\sim 80$  ns de simulación), en comparación de lo observado en el sistema HSP60-HSP10-CitC cuando la HSP60 y CitC se encuentran juntas. Esto parece indicar que el efecto de la HSP60 sobre CitC es ayudarle a mantener un nivel de compactación adecuado, de manera que cuando CitC se encuentra sola ésta empieza a deformarse, haciendo que su equilibrio estructural sea más difícil de alcanzarse.

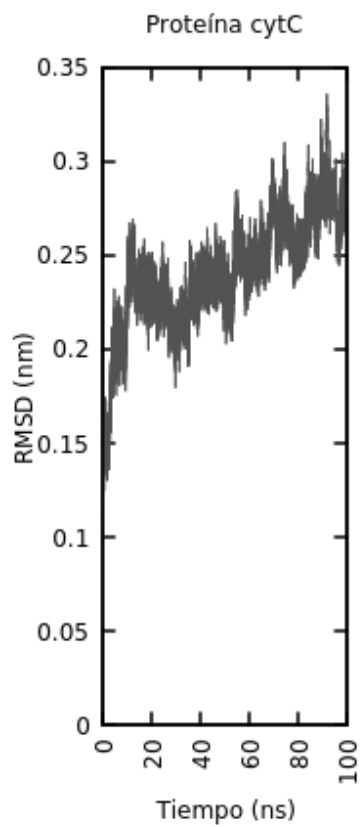


Figura 5.2: Evolución temporal de la desviación cuadrática media (RMSD) para la proteína cytC.

## 5.2. Mapas de la distribución de potencial electrostático instantánea y promedio de la HSP60 en los diferentes sistemas moleculares de estudio

Una vez que encontramos que los tres sistemas de estudio son estructuralmente estables, se procedió a calcular y representar los mapas de la distribución de potencial electrostático de largo alcance en la HSP60 (ver sección 4.2.4). Las figuras 5.3, 5.4 y 5.5 muestran los mapas de la distribución de potencial electrostático instantánea de 12 fotogramas seleccionados al azar<sup>1</sup> de los últimos 10 ns de las trayectorias de los diferentes sistemas moleculares de estudio para la HSP60. Los resultados apuntan a un patrón de comportamiento característico común en cada sistema. Por ejemplo, en los sistemas HSP60-HSP10 y HSP60-HSP10-cytC (ver Figura 5.3 y 5.4), la región conformada exclusivamente por los átomos de la HSP60 presenta una distribución de potencial electrostático que puede considerarse espacialmente homogénea y negativa (regiones de color rojo), mientras que lo contrario puede observarse para la región conformada por los átomos de la HSP10, donde se puede apreciar una distribución de potencial homogénea positiva (regiones de color azul). Adicionalmente, puede apreciarse una región de potencial nulo (regiones de color blanco) correspondiente a la interfase de las regiones de potencial electrostático de la HSP60 (negativas) y de las regiones potencial electrostático de la HSP10 (positivas). Lo anterior resulta interesante particularmente en el sistema HSP60-HSP10-cytC, donde se encuentra una estructura ajena (la de la proteína cytC) a la del complejo HSP60-HSP10. Para este caso, la distribución de potencial electrostático observada en la Figura 5.4 no se altera al encontrarse en el interior de la HSP60 y, al mismo tiempo, tampoco altera la distribución de potencial electrostático de la HSP60, es decir, se mantienen aproximadamente las mismas características que las observadas en el sistema HSP60-HSP10 (ver Figura 5.3).

<sup>1</sup>Realizamos la selección con un generador de números aleatorios en línea 59.

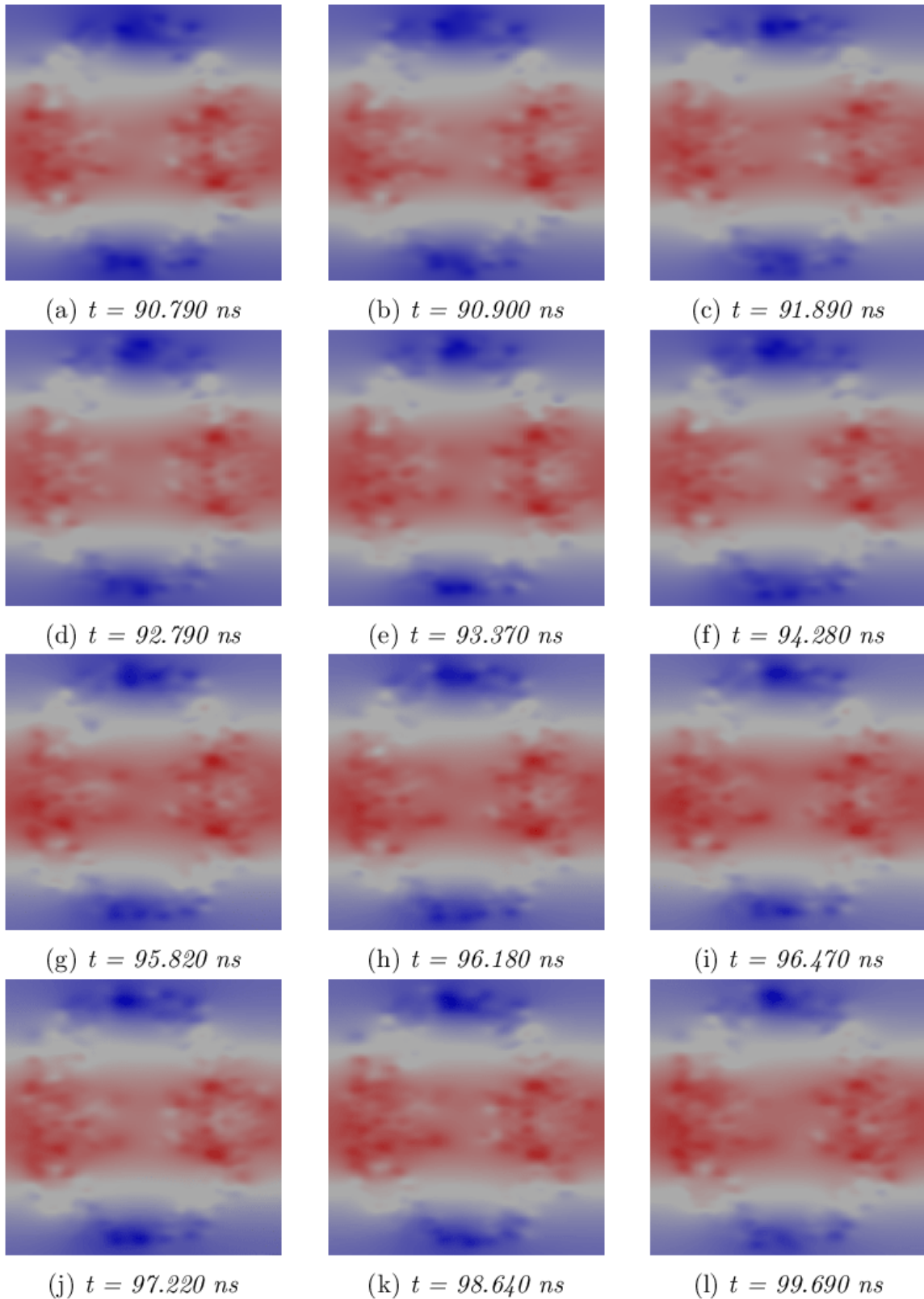


Figura 5.3: Mapas de la distribución del potencial electrostático de largo alcance instantánea del sistema HSP60-HSP10 en el plano YZ.

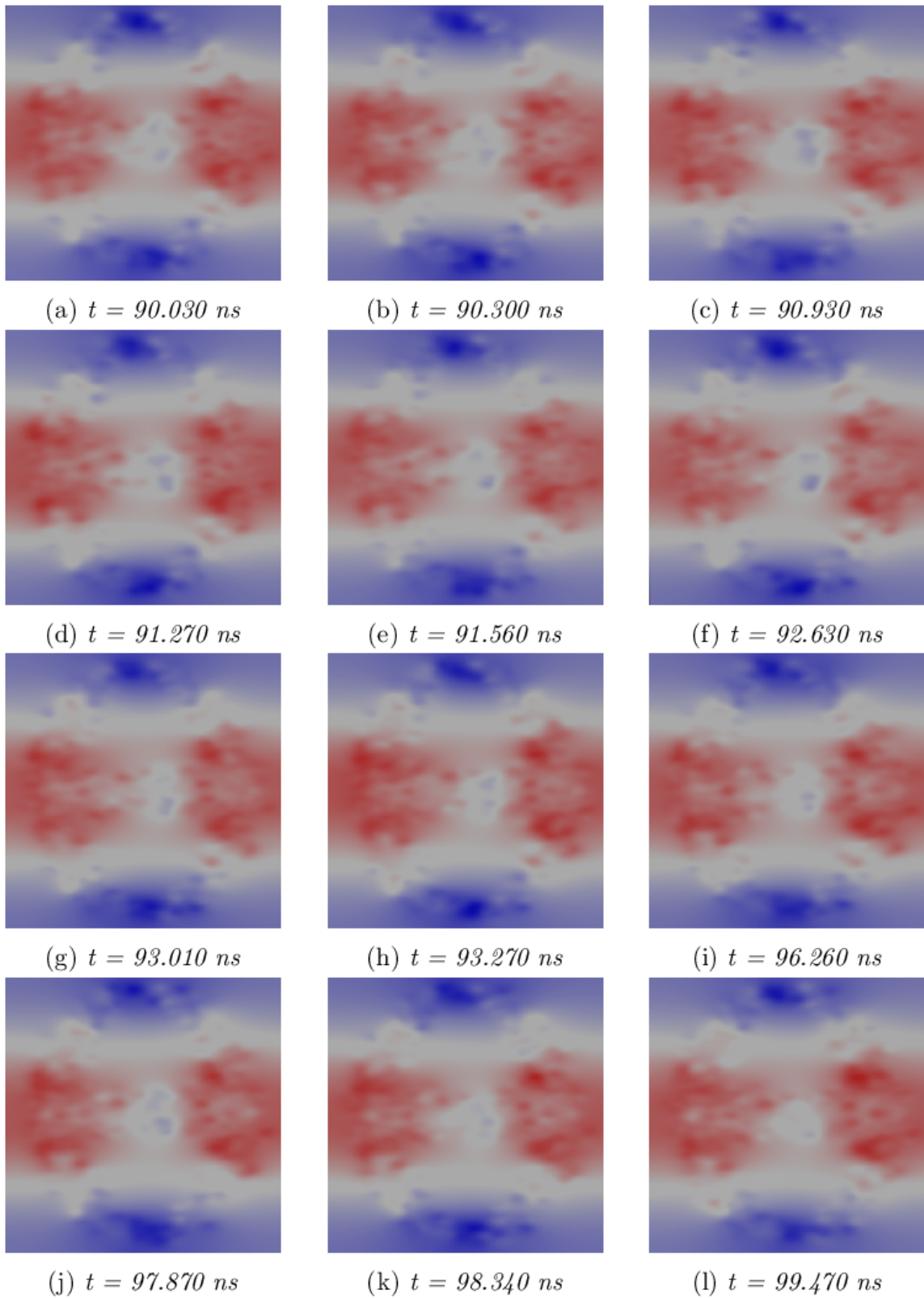


Figura 5.4: Mapas de la distribución de largo alcance instantánea del sistema HSP60-HSP10-cytC en el plano YZ.

El caso contrario a tales observaciones puede apreciarse en los mapas de la distribución de potencial electrostático instantánea del sistema HSP60 (ver Figura 5.5). Para este sistema, los átomos de la estructura de la HSP60 por sí solos muestran una distribución de potencial electrostático heterogénea con regiones de potencial electrostático positivas y negativas circundantes. Además, a diferencia del patrón observado en la distribución de potencial electrostático en los dos sistemas anteriores, en el interior de la HSP60 la distribución de potencial electrostático parece ser uniformemente positiva, desde el centro del sistema hasta los extremos de la HSP60 (ver Figura 5.8). Lo anterior puede suponer que la presencia de la HSP10 influye notablemente en el arreglo dimensional de la distribución de potencial de la HSP60 por sí sola alterando, como consecuencia, su distribución de potencial electrostático cuando esta presente.

Los mapas de potencial electrostático promedio de las Figuras 5.6, 5.7 y 5.8 obtenidos a partir de considerar 1,000 fotogramas extraídos de los últimos 10 ns de la trayectoria para cada sistema, presentan con mayor detalle cada una de las observaciones anteriores. Cada uno se calculó usando todos los átomos del sistema, un mallado tridimensional de  $120 \times 120 \times 120$  y un factor de Ewald de 0.25 (equivalente al parámetro  $\beta$  presentado en las ecuaciones 4.16 y 4.19). Las alturas de los planos representan la posición del centro de masa de todos los átomos del sistema (donde  $h/2$  es el centro de la caja de simulación) y la altura de la caja, respectivamente. En general, se puede apreciar que la distribución de potencial electrostático en cada sistema molecular representado para la HSP60, mantiene el mismo patrón de comportamiento que observado para los diferentes tiempos de la trayectoria en las Figuras 5.3, 5.4 y 5.5, en donde hay una distribución de potencial negativa en el centro del sistema y conforme se desplazada hacia la interface entre la distribución de potencial electrostático entre la HSP60 y la HSP10 ésta varía hasta hacerse neutra.

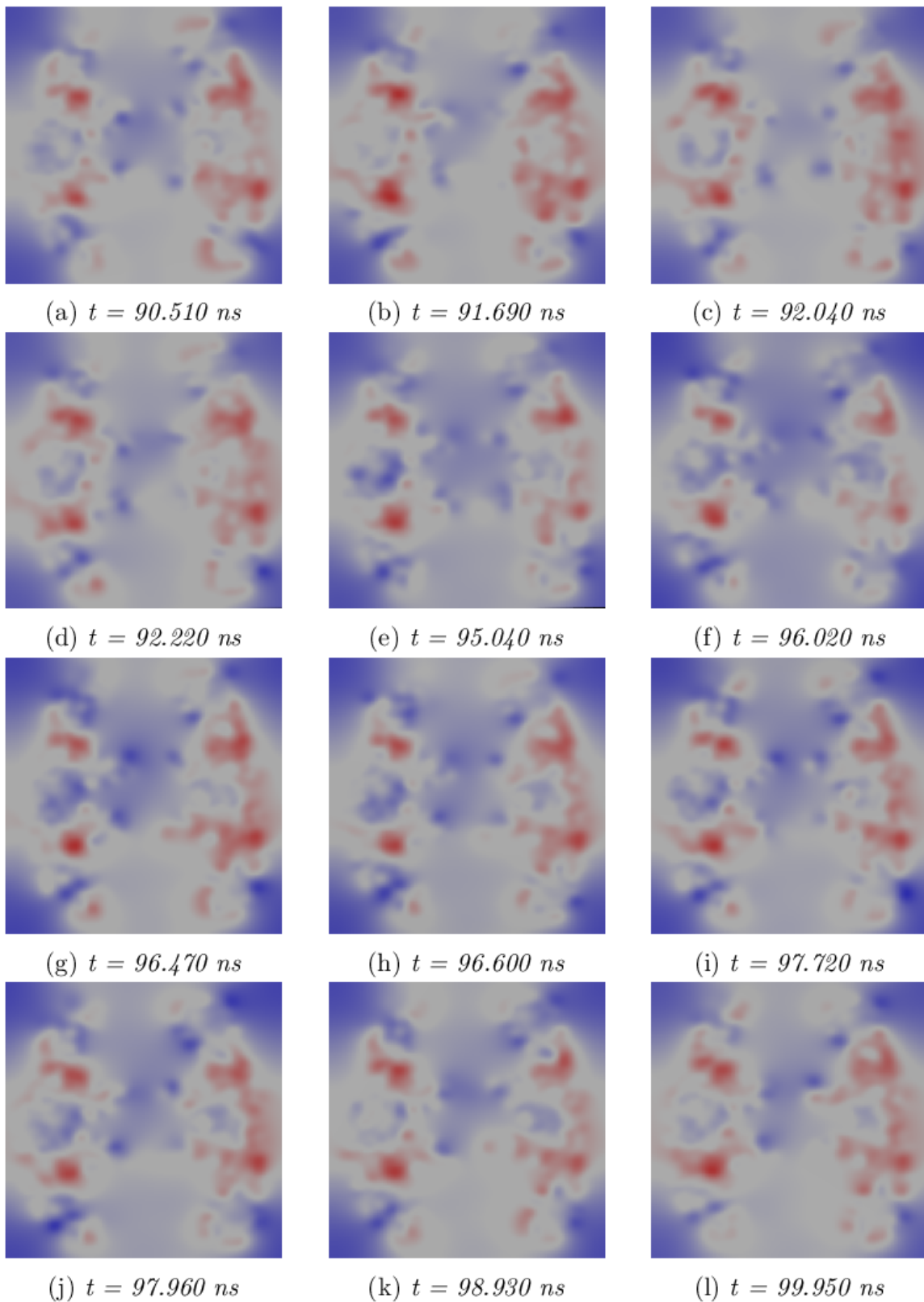
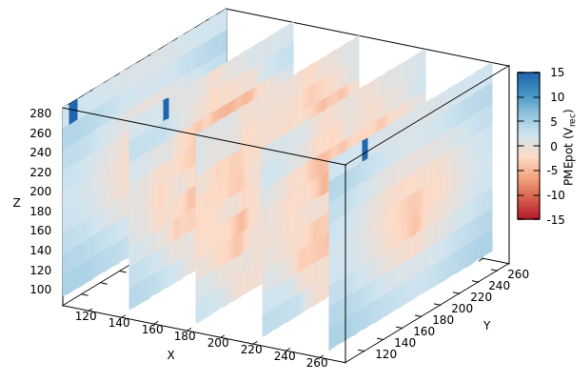
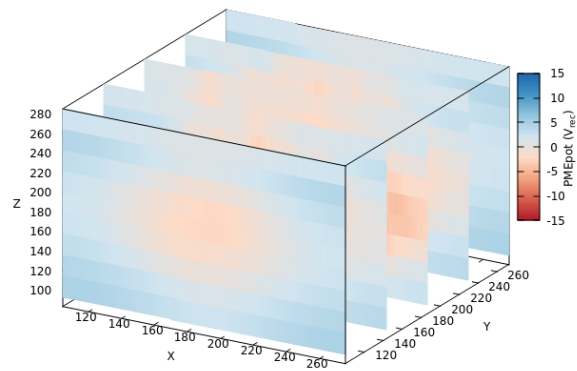


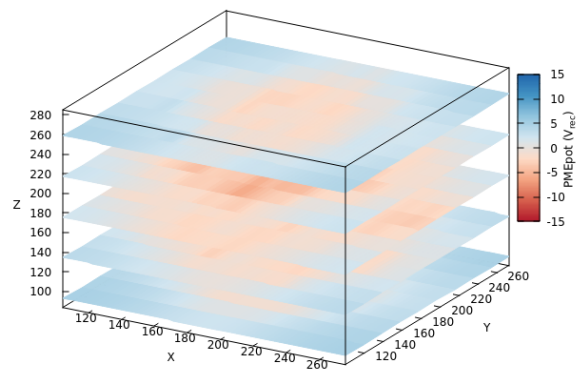
Figura 5.5: Mapas de la distribución del potencial electrostático de largo alcance instantánea del sistema HSP60 en el plano YZ.



(a) Plano YZ

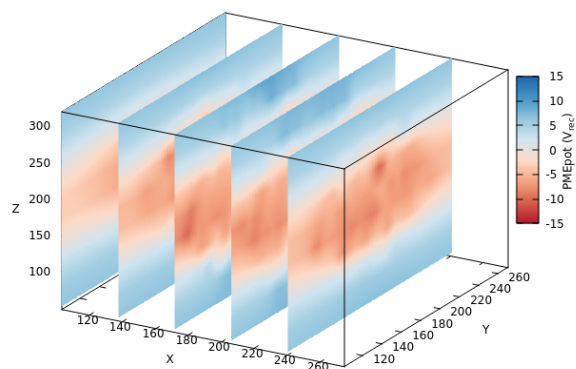


(b) Plano XZ

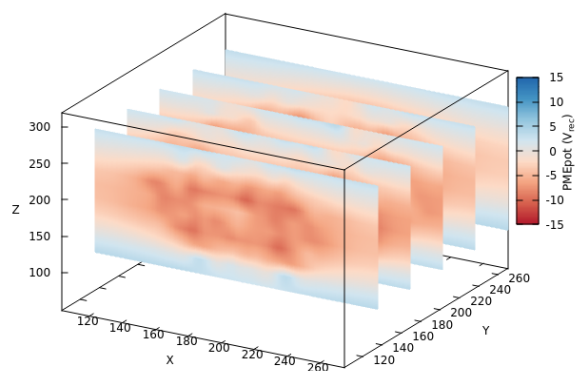


(c) Plano XY

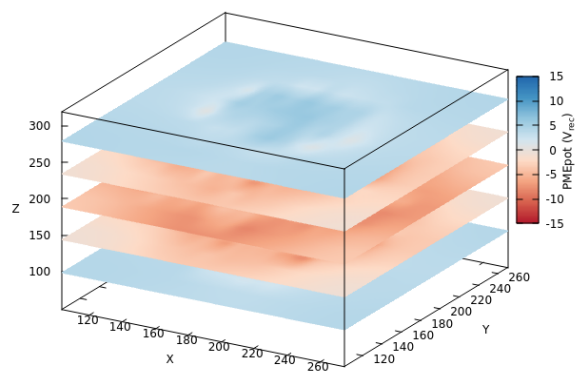
Figura 5.6: Mapas de la distribución de potencial electrostático de largo alcance promedio para el sistema HSP60 desde diferentes planos. La escala de valores para la distribución de potencial electrostático de largo alcance se encuentra representada en unidades de Volts (V).



(a) Plano YZ

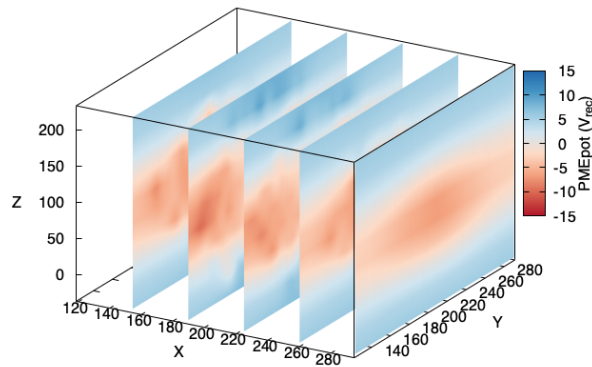


(b) Plano XZ

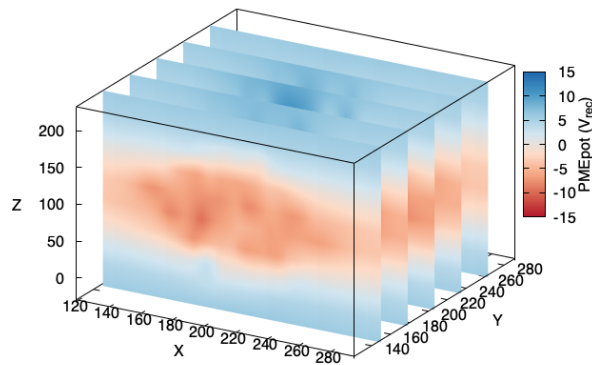


(c) Plano XY

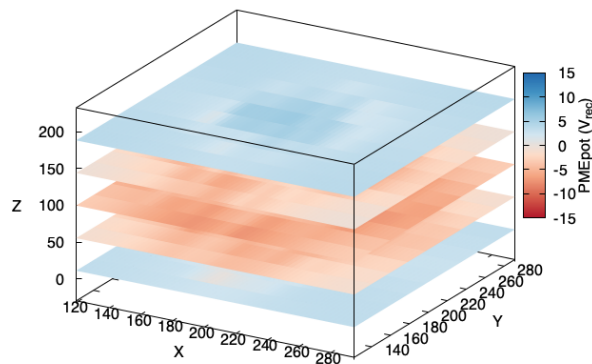
Figura 5.7: Mapas de la distribución de potencial electrostático de largo alcance promedio para el sistema HSP60-HSP10 desde diferentes planos. La escala de valores para la distribución de potencial electrostático de largo alcance se encuentra representada en unidades de Volts (V).



(a) Plano YZ



(b) Plano XZ



(c) Plano XY

Figura 5.8: Mapas de la distribución de potencial electrostático de largo alcance promedio para el sistema HSP60-HSP10-CitC desde diferentes planos. La escala de valores para la distribución de potencial electrostático de largo alcance se encuentra representada en unidades de Volts (V).

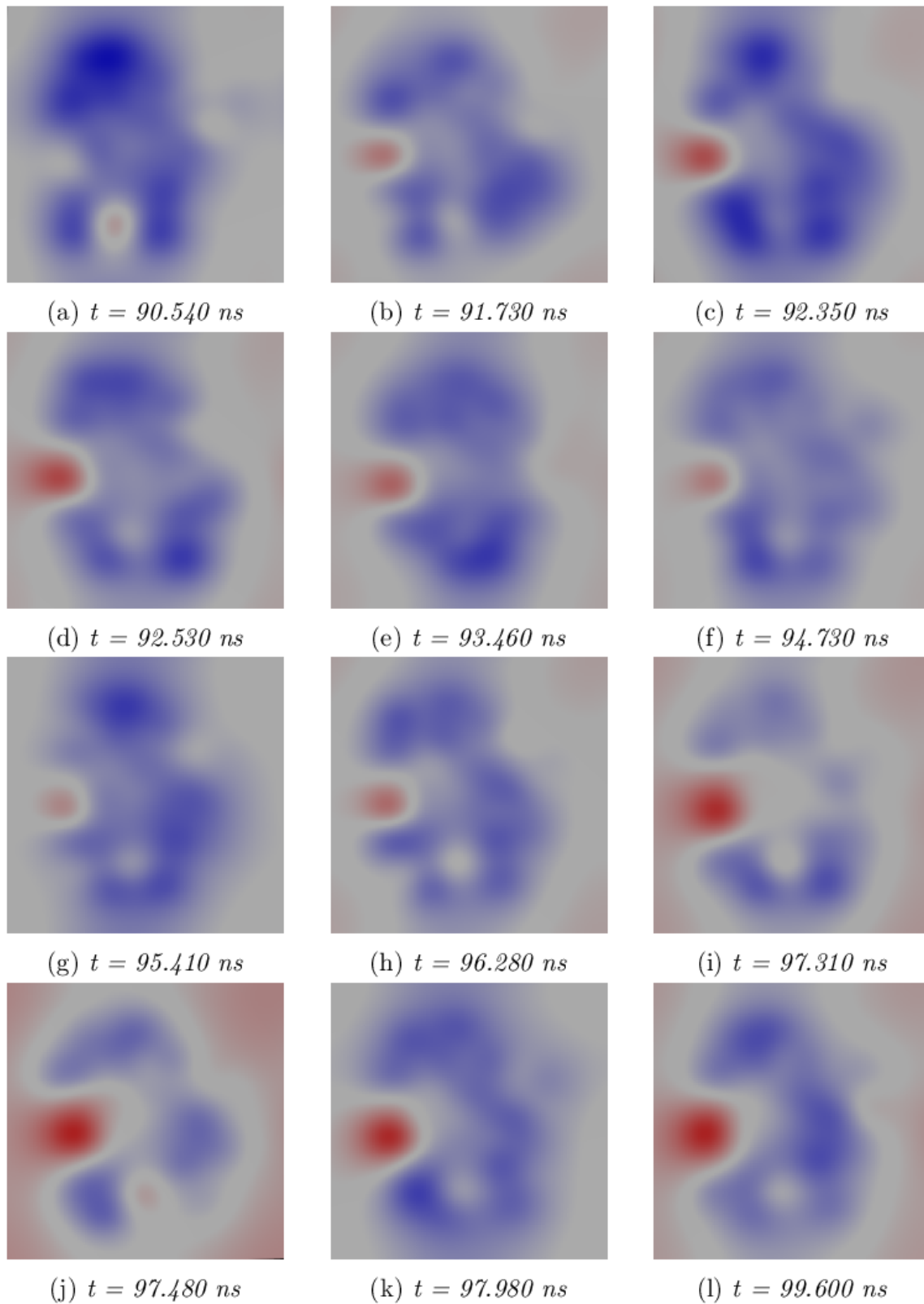


Figura 5.9: Mapas de la distribución del potencial electrostático de largo alcance instantánea de la proteína cytC en el plano YZ.

Adicionalmente, se analizó por separado la distribución de potencial electrostático instantánea y promedio de CitC. El resultado puede observarse en la Figuras 5.9 y 5.10. Para esta proteína, el patrón de comportamiento de la distribución de potencial electrostático de largo alcance no difiere del observado en el sistema HSP60-HSP10-cytC, esto es, la distribución de potencial electrostático es mayormente positiva (región de color azul), donde la región de potencial negativo (región de color rojo) corresponde al grupo carboxilo terminal ( $\text{COO}^-$ ).

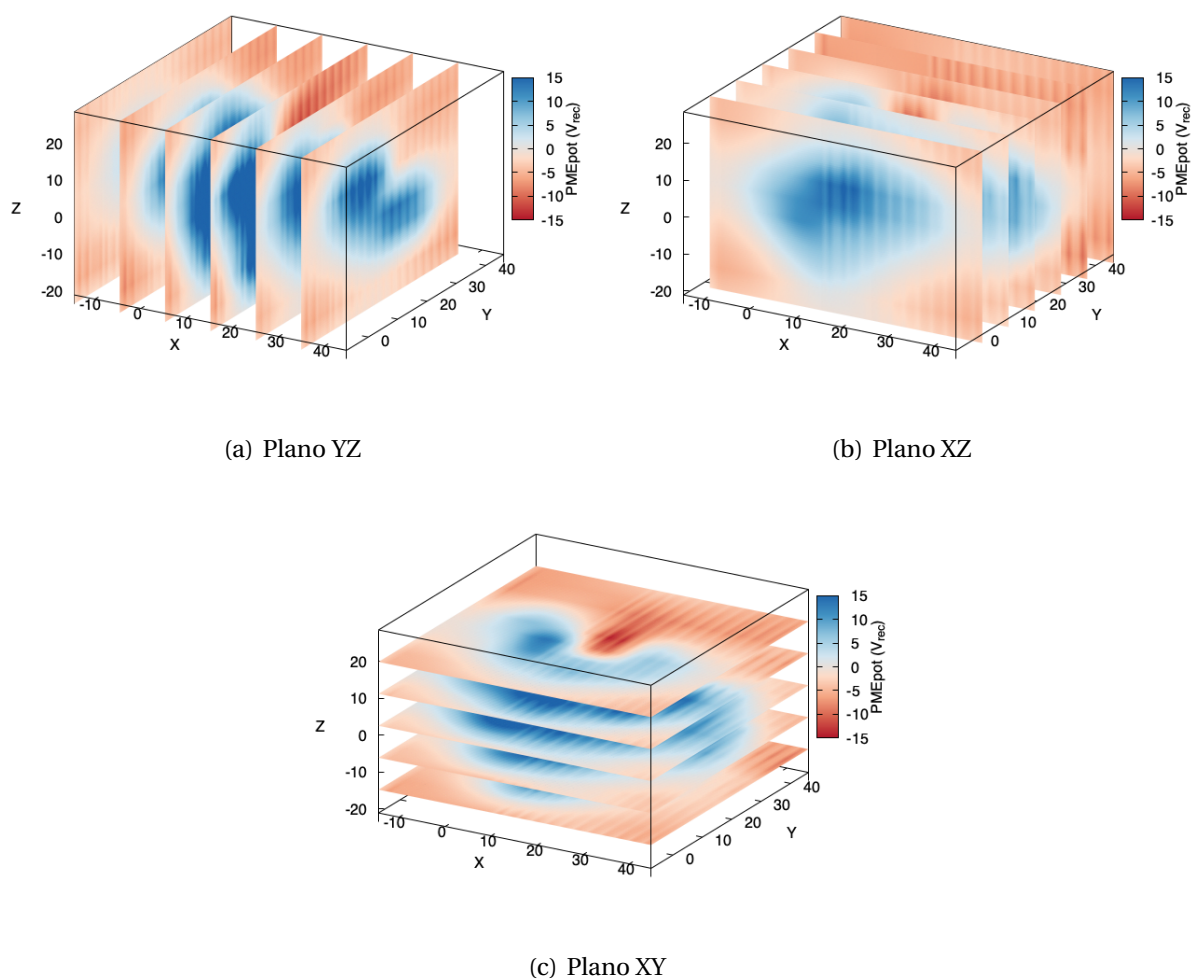


Figura 5.10: **Mapas de la distribución de potencial electrostático de largo alcance promedio para CitC desde diferentes planos.** La escala de valores para la distribución de potencial electrostático de largo alcance se encuentra representada en unidades de Volts (V).

## Conclusiones

Las conclusiones más importantes que se derivan de esta tesis pueden resumirse en los siguientes puntos:

- Se logró construir sistemas moleculares de estudio útiles para representar los efectos electrostáticos de largo alcance en el interior de la chaperonina HSP60.
- Se encontró que la estructura de chaperona HSP60 sola y la HSP60 en complejo con HSP10 y la proteína del citocromo C alcanzan estados estructuralmente estables, sin evidencia de que éstas últimas alteren significativamente el movimiento relativo de los átomos de la HSP60.
- Y, lo más relevante, se encontró que la distribución espacial de  $V_{rec}$  en el interior de la chaperonina HSP60-HSP10 es esencialmente invariante en el tiempo y, espacialmente, no es aleatoria. Lo que parece indicar que la contribución de la chaperonina es ejercer un campo electrostático de largo alcance aproximadamente independiente en el tiempo. En consecuencia, parece factible simular el efecto de una chaperona a través de un campo de fuerzas externo que reproduzca el patrón observado.

# Perspectivas

Una vez que tenemos la prueba de concepto de que el campo electrostático de largo alcance,  $V_{\text{rec}}$  (en el interior de la cavidad de la chaperona) tiene un patrón que no es aleatorio y que es invariante en el tiempo, planteamos las siguientes líneas posibles de investigación.

- La forma de  $V_{\text{rec}}$ , como función de las coordenadas espaciales, se podría representar mediante alguna expansión de algún conjunto de polinomios ortogonales. Para este tópico, se puede utilizar la malla de puntos que encontramos en esta tesis como puntos de control para hallar los coeficientes de la expansión.
- Una vez se encuentre la expansión de  $V_{\text{rec}}$ , ésta se puede añadir como un campo de fuerzas externo a las simulaciones de dinámica molecular (DM). Al ser un campo externo, se espera que el costo computacional sea de  $O(N)$ , donde  $N$  es el número de átomos de la proteína que se pliega. Nótese que no habría necesidad de incluir la chaperona en sí (lo cual nos ahorraría más de cien mil átomos en las simulaciones). Esto permitiría incrementar el tiempo de observación, lo que a su vez, permitiría explorar una región mayor de la superficie de energía de Gibbs de la proteína que se pliega.
- Si bien en este trabajo no se discutió explícitamente el hecho de que las proteínas se pliegan dentro de la cavidad de la chaperona, esto se puede incluir en simulaciones de DM removiendo las condiciones de frontera periódicas y reemplazándolas por una pared confinante. Esto nos ahorraría una cantidad importantísima de moléculas de solvente (del orden de varios cientos de miles de átomos), ya que sólo se requeriría añadir

las moléculas de solvente que cupiesen dentro de la pared confinante. Es altamente posible que esta pared confinante ejercería un efecto importante sobre la proteína que se pliega.

- La forma geométrica de la pared confinante podría ser simplemente un elipsoide, una esfera o un cilindro. Sin embargo, idealmente podríamos ajustarla de manera que coincidiera con la forma de la cavidad de la chaperona. De ser así, se podría utilizar esta superficie para ajustar mejor  $V_{\text{rec}}$ .
- Los puntos anteriores pueden implementarse en algún programa de DM, como Gromacs o LAMMPS, por lo que serían retos interesantes tanto corroborar el modelo que proponemos como contribuir al desarrollo del software en sí.

# Apéndices

## Apéndice A. Cálculo, generación de datos y representación de la electrostática de largo alcance

Abajo se describe el conjunto de herramientas (*genAveragePME* y *manipdx*), diseñadas a propósito, para el cálculo, elaboración de datos y representación de la electrostática en una molécula de interés, que tienen como objetivo facilitar la configuración y ejecución de cálculos electrostáticos y, así, ampliar la accesibilidad de su análisis. El esquema general de trabajo adoptado puede verse representado en la Figura 6. Primero, *genAveragePME* llama a PDB2PQR [60], un programa que lee un archivo pdb y asocia un valor de carga a cada átomo, basado en una biblioteca interna y un conjunto de parámetros seleccionados por el usuario y especificados a través de la interfaz de *genAveragePME*. La salida de PDB2PQR consta de un solo archivo: el archivo PQR, que es básicamente una versión modificada de pdb, y el cual será utilizado por *plug-in* PMEpot de VMD [58]. En el siguiente paso, el *plug-in* PMEpot, evalúa la distribución de potencial electrostático en el espacio de la molécula, resolviendo el término recíproco del método PME, en el que se integran los valores de las cargas parciales asociadas con los átomos, y los valores del potencial se asignan a puntos de la cuadrícula. Este programa crea un archivo dx que tiene la información del potencial electrostático de largo alcance en los puntos de la cuadrícula; por defecto, el tamaño de la cuadrícula está determinado por el tamaño de la molécula, mientras que su densidad puede ser definida

por el usuario: una cuadrícula más fina permite una mejor descripción. En este punto, dado que nos interesa obtener un campo promedio válido de los valores de potenciales calculados por *PMEpot* sobre una muestra representativa, *genAveragePME*, a través de una serie de instrucciones especificadas, lee un conjunto de archivos *dx*, suma los valores del potencial en cada punto de la cuadrícula y los multiplica por el recíproco del número de archivos *dx* considerados, a través de trazar un misma cuadrícula con respecto al centro geométrico de la molécula considerada (ver Figura 6.1). La salida de *genAveragePME* consta de un solo archivo: el archivo *averagePME.dx* que contiene el campo promedio fijado sobre el centro geométrico de la molécula de interés. Finalmente, este nuevo conjunto de datos (archivo *averagePME.dx*) puede pasarse a *manipdx*, un programa, también diseñado a propósito, que permite manipular estos valores mediante un conjunto de operaciones aritméticas básicas o bien, interconvertir el formato del archivo dado para utilizarse en herramientas de gráficos, como *GNUPLOT*, que permita obtener una visualización interactiva y agradable al espectador.

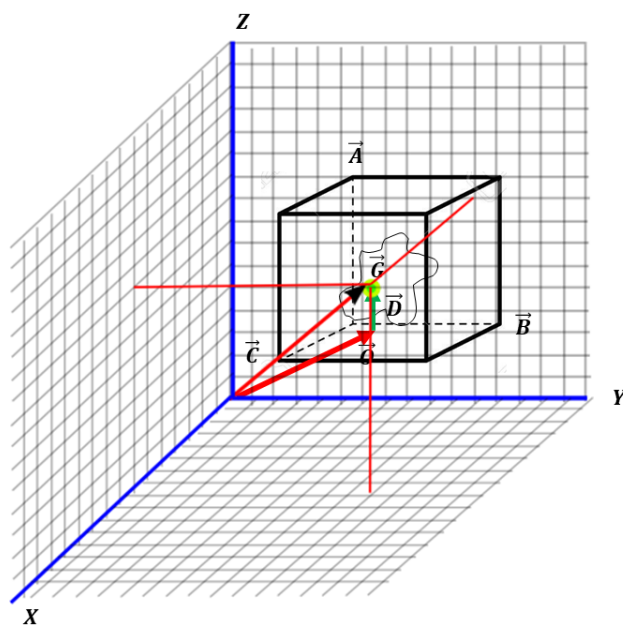


Figura 6.1: Ejemplo de cuadrícula de potencial para calcular el campo promedio de la distribución de largo alcance de una molécula de interés en *genAveragePME*. Los vectores  $\vec{A}$ ,  $\vec{B}$ ,  $\vec{C}$  y  $\vec{O}$  definen el largo, ancho y alto y el origen de la cuadrícula de potencial, respectivamente. El vector  $\vec{G}$  representa el centro geométrico de la molécula considerada y  $\vec{D}$ , la diferencia entre el vector  $\vec{O}$  y  $\vec{G}$  ( $\vec{O} - \vec{G}$ ), que permite fijar la cuadrícula de potencial a la misma distancia del centro geométrico de la molécula independientemente del fotograma analizado.

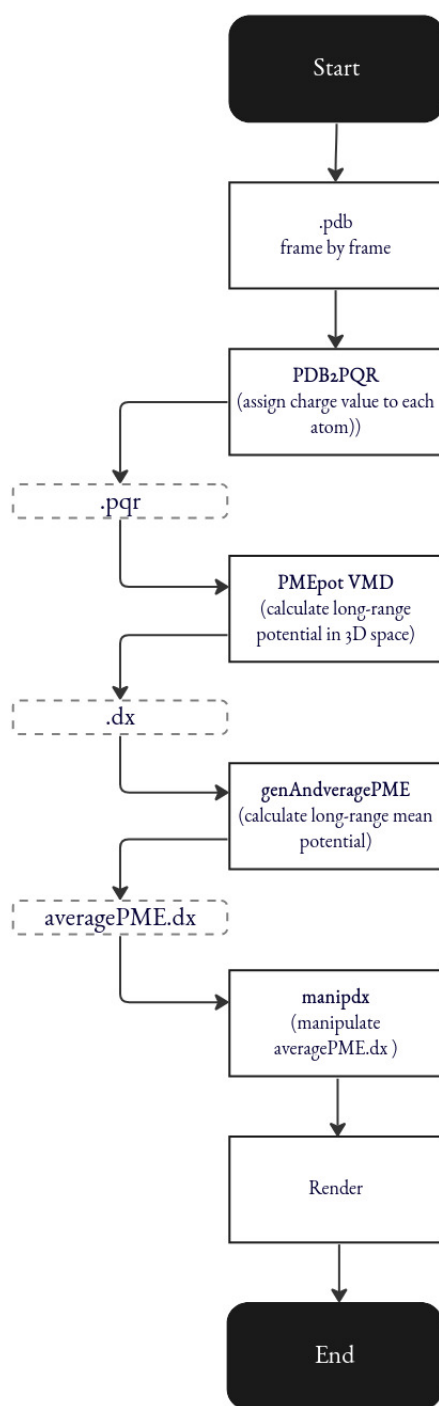


Figura 6.2: Esquema general del proceso para el cálculo, generación de datos y representación de la distribución de potencial electrostático de largo alcance con ayuda de los programas *genAndveragePME* y *manipdx* diseñados a propósito para realizar esta tarea (basado en [61]). Cada uno de estas herramientas se encuentran en el servidor de cómputo TRANTOR disponible en el Laboratorio de Físicoquímica Teórica de la Benemérita Universidad Autónoma de Puebla.

# Bibliografía

- [1] A. Martínez del Pozo, “Estaba christian anfinen en lo cierto,” *E-Print Complutense*, pp. 1–8, 2010.
- [2] S. Romero Romero, D. A. Fernández Velasco, and M. Costas, “Estabilidad termodinámica de proteínas,” *Educación Química*, vol. 29, no. 3, pp. 1–3, 2018.
- [3] A. Dwevedi, *Protein folding examining the challenges from synthesis to folded form*. SpringerBriefs in Biochemistry and Molecular Biology, 2015.
- [4] W. A. Houry, *The molecular chaperones interaction networks in protein folding and degradation*. Springer New York Heidelberg Dordrecht London, 2014.
- [5] O. Llorca, J. L. Carrascosa, and J. M. Valpuesta, “Las chaperoninas: Máquinas plegadoras de proteínas,” *Ciencia al Día Internacional*, vol. 4, no. 2, 2001.
- [6] T. González García, “Diseño de potenciales externos para inducir el plegamiento de proteínas en dinámica molecula,” Master’s thesis, Universidad Nacional Autónoma de México, 2017.
- [7] S. Nisemblat, O. Yaniv, A. Parnas, F. Frolov, and A. Azem, “Crystal structure of the human mitochondrial chaperonin symmetrical football complex,” *Proceedings of the National Academy of Sciences*, vol. 112, no. 19, pp. 6044–6049, 2015.

- [8] Y. Gomez-Llorente, F. Jebara, M. Patra, R. Malik, S. Nisemblat, O. Chomsky-Hecht, A. Parnas, A. Azem, J. A. Hirsch, and I. Ubarretxena-Belandia, “Structural basis for active single and double ring complexes in human mitochondrial Hsp60-Hsp10 chaperonin,” *Nature Communications*, vol. 11, no. 1, p. 1916, 2020.
- [9] C. Hernández-Chinea, “REVISIÓN DEL DOGMA ESTRUCTURA-FUNCIÓN EN LAS PROTEÍNAS,” vol. LXXV, pp. 11–15, 2015.
- [10] S. Salamanca Seguí, *Plegamiento y funcionalidad biológica del inhibidor de metalocarboxipeptidasas LCI (Leech Carboxypeptidase Inhibitor)*. PhD thesis, Universidad Autónoma de Madrid, 2008.
- [11] L. Olivares-Quiroz and L. G.-C. Scherer, “Plegamiento de las proteínas: Un problema interdisciplinario,” *Rev. Soc. Quím. Méx*, vol. 48, 2004.
- [12] J. Macošek, G. Mas, and S. Hiller, “Redefining Molecular Chaperones as Chaotropes,” *Frontiers in Molecular Biosciences*, vol. 8, pp. 1–11, 2021.
- [13] R. J. Ellis and S. M. Hemmingsen, “Molecular chaperones: proteins essential for the biogenesis of some macromolecular structures,” *Trends in Biochemical Sciences*, vol. 14, no. 8, pp. 339–342, 1989.
- [14] M. M. Altamirano, R. Golbik, R. Zahn, A. M. Buckle, and A. R. Fersht, “Refolding chromatography with immobilized mini-chaperones,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 8, pp. 3576–3578, 1997.
- [15] K. Gulukota and P. G. Wolynes, “Statistical mechanics of kinetic proofreading in protein folding in vivo,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 91, no. 20, pp. 9292–9296, 1994.
- [16] C. D. Sfatos, A. M. Gutin, V. I. Abkevich, and E. I. Shakhnovich, “Simulations of chaperone-assisted folding,” *Biochemistry*, vol. 35, no. 1, pp. 334–339, 1996.
- [17] H. Fan, “Mimicking the action of GroEL in molecular dynamics simulations: Application to the refinement of protein structures,” *Protein Science*, vol. 15, no. 3, pp. 441–448, 2006.

## BIBLIOGRAFÍA

---

- [18] J. L. Carrascosa, O. Llorca, and J. M. Valpuesta, “Structural comparison of prokaryotic and eukaryotic chaperonins,” *Micron*, vol. 32, no. 1, pp. 43–50, 2001.
- [19] A. M. Roseman, S. Chen, H. White, K. Braig, and H. R. Saibil, “The chaperonin ATPase cycle: Mechanism of allosteric switching and movements of substrate-binding domains in GroEL,” *Cell*, vol. 87, no. 2, pp. 241–251, 1996.
- [20] M. Klumpp and W. Baumeister, “The thermosome: Archetype of group II chaperonins,” *FEBS Letters*, vol. 430, no. 1-2, pp. 73–77, 1998.
- [21] N. R. Douglas, S. Reissmann, J. Zhang, B. Chen, J. Jakana, R. Kumar, W. Chiu, and J. Frydman, “Dual action of ATP hydrolysis couples lid closure to substrate release into the group II chaperonin chamber,” *Molecular and Cellular Biochemistry*, vol. 144, pp. 240–252, 2011.
- [22] P. B. Sigler, Z. Xu, H. S. Hrye, S. G. Burston, W. A. Fenton, and A. L. Horwich, “Structure and function in groel-mediated protein folding,” *Annual review of biochemistry*, vol. 67, pp. 581–608, 1998.
- [23] A. White, A. Nowinski, W. Huang, A. Keefe, F. Sun, and S. Jiang, “Decoding nonspecific interactions from nature,” *Chemical science (Royal Society of Chemistry)*, vol. 3, p. 3488, 09 2012.
- [24] M. Shtilerman, G. H. Lorimer, and S. W. Englander, “Chaperonin function: Folding by forced unfolding,” *Science*, vol. 284, no. 5415, pp. 822–825, 1999.
- [25] A. Delgado, “DeepMind y el plegamiento de las proteínas at <https://pixelabs.es/como-la-ia-ayuda-a-predecir-el-plegamiento-de-las-proteinas/>,” 2021.
- [26] “La suma de matemáticas y bioinformática desentraña las piezas básicas de los seres humanos at <https://www.csic.es/es>,” *Campillo, N. and Fernández, C.*, 2022.
- [27] M. I. Romero Hermida, “Efecto de la composición de los gases de combustión en la captura de CO<sub>2</sub> por nanotubos de carbono: un estudio de simulación,” Master’s thesis, Uni-

- versidad de Sevilla, Departamento de Ingeniería Química y Ambiental, Escuela Técnica Superior de Ingeniería, 2012.
- [28] B. A. Stickler and E. Schachinger, *Molecular Dynamics*, pp. 97–109. Cham: Springer International Publishing, 2014.
- [29] H. Hajiabadi, “Getting Started With Molecular Dynamics Simulation at <https://insilicosci.com/molecular-dynamics-simulation-concepts>,” *Insilicosci-computational chemistry, physics, biology, drug design, and materials science*, 2022.
- [30] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen, “Evaluation and Re-parametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides,” *The Journal of Physical Chemistry B*, vol. 105, no. 28, pp. 6474–6487, 2001.
- [31] W. Damm, A. Frontera, J. Tirado-Rives, and W. L. Jorgensen, “OPLS all-atom force field for carbohydrates,” *Journal of Computational Chemistry*, vol. 18, no. 16, pp. 1955–1970, 1997.
- [32] D. Quiñonero, S. Tomàs, A. Frontera, C. Garau, P. Ballester, A. Costa, and P. M. Deyà, “OPLS all-atom force field for squaramides and squaric acid,” *Chemical Physics Letters*, vol. 350, no. 3-4, pp. 331–338, 2001.
- [33] K. Murzyn, M. Bratek, and M. Pasenkiewicz-Gierula, “Refined OPLS all-atom force field parameters for n-pentadecane, methyl acetate, and dimethyl phosphate,” *Journal of Physical Chemistry B*, vol. 117, no. 51, pp. 16388–16396, 2013.
- [34] R. Dunbrack, “Opls,” *Dictionary of Bioinformatics and Computational Biology*, 2004.
- [35] W. Jorgensen, D. Maxwell, and J. Tirado-Rives, “Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids,” *J. Am. Chem. Soc.*, vol. 118, pp. 11225–11236, 01 1996.
- [36] M. Abraham, D. van der Spoel, E. Lindahl, B. Hess, G. development team, D. van der Spoel, E. L. V. 2016.3, and 2016, “Gromacs 2018,” *GROMACS User Manual*, 2018.

- [37] R. A. Carraro Pacheco, *Modelado y estudio de complejos de Ciclosporina A y compuestos relacionados con una ciclofilina de Trypanosoma cruzi*. PhD thesis, Centro de Bioinformática Estructural DETEMA, Facultad de Química-Universidad de la República, 2013.
- [38] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. Mark, and H. Berendsen, “GROMACS: fast, flexible, and free,” *Journal of computational chemistry*, vol. 26, pp. 1701–18, 2005.
- [39] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, “Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers.,” *SoftwareX*, 2015.
- [40] A. R. Panchenko, Z. Luthey-Schulten, and P. G. Wolynes, “Foldons, protein structural modules, and exons,” *Proceedings of the National Academy of Sciences*, vol. 93, pp. 2008–2013, 1996.
- [41] A. Panchenko, Z. Luthey-Schulten, R. Cole, and P. Wolynes, “The foldon universe: a survey of structural similarity and self-recognition of independently folding units,” *Journal of molecular biology*, vol. 272, pp. 95–105, 1997.
- [42] Y. Bai, T. Sosnick, L. Mayne, and S. Englander, “Protein folding intermediates: native-state hydrogen exchange,” *Sciences*, vol. 14, pp. 192–197, 1995.
- [43] Y. Xu, L. Mayne, and S. W. Englander, “Evidence for an unfolding and refolding pathway in cytochrome c,” *Nature Structural Biology*, vol. 5, pp. 774–778, 1998.
- [44] R. L. Baldwin, “Clash between energy landscape theory and foldon-dependent protein folding,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 114, pp. 8442–8443, 2017.
- [45] S. W. Englander and L. Mayne, “The case for defined protein folding pathways,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 114, pp. 8253–8258, 2017.

- [46] H. Maity, M. Maity, M. G. Krishna Mallela, L. Mayne, and S. W. Englander, “Protein folding: The stepwise assembly of foldon units,” *Proceedings of the National Academy of Sciences*, vol. 102, pp. 4741–4746, 2005.
- [47] B. S. Rajagopal, A. N. Edzuma, M. A. Hough, K. L. I. M. Blundell, V. E. Kagan, A. A. Kapralov, L. A. Fraser, J. N. Butt, G. G. Silkstone, M. T. Wilson, D. A. Svistunenko, and J. A. R. Worrall, “The hydrogen-peroxide-induced radical behaviour in human cytochrome c–phospholipid complexes: implications for the enhanced pro-apoptotic activity of the G41S mutant,” *Biochemical Journal*, vol. 456, no. 3, pp. 441–452, 2013.
- [48] L. Peña Ortiz, “Technical report: Configuración de un sistema de simulación que contenga a las chaperonas hsp60 y hsp10 en complejo con la proteína del citocromo c, benemérita universidad autónoma de puebla, laboratorio de software y modelado matemático unpublished,” 2022.
- [49] J. A. Lemkul, “From Proteins to Perturbed Hamiltonians: A Suite of Tutorials for the GROMACS-2018 Molecular Simulation Package [Article v1.0],” *Living Journal of Computational Molecular Science*, vol. 1, no. 1, p. 5068, 2018.
- [50] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, *Interaction Models for Water in Relation to Protein Hydration*, pp. 331–342. Springer Netherlands, 1981.
- [51] G. Bussi, D. Donadio, and M. Parrinello, “Canonical sampling through velocity rescaling,” *Journal of Chemical Physics*, vol. 126, no. 1, 2007.
- [52] M. Parrinello and A. Rahman, “Polymorphic transitions in single crystals: A new molecular dynamics method,” *Journal of Applied Physics*, vol. 52, no. 12, pp. 7182–7190, 1981.
- [53] W. F. Van Gunsteren and H. J. Berendsen, “A leap-frog algorithm for stochastic dynamics,” *Molecular Simulation*, vol. 1, no. 3, pp. 173–185, 1988.
- [54] L. E. Kavradi, *Molecular Distance Measures*. Connexions, 2009.

- 
- [55] T. Darden and L. Perderson, "Particle mesh ewald: An  $n\log(n)$  method for ewald sums in large systems," *J. Chem. Phys.*, vol. 98, pp. 8577–8592, 1993.
- [56] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, L. Hising, and et. al, "A smooth particle mesh ewald method," *J. Chem. Phys.*, vol. 108, pp. 8577–8592, 1995.
- [57] A. Toukmaji and J. Board, "Technical report no. 95-004, duke university unpublished. 30 l. greengard," *The Rapid Evaluation of Potential Fields in Particle Systems*, 1988.
- [58] A. Aksimentiev and K. Schulten, "Imaging  $\alpha$ -Hemolysin with Molecular Dynamics: Ionic Conductance, Osmotic Permeability, and the Electrostatic Potential Map," *Biophysical Journal*, vol. 88, no. 6, pp. 3745–3761, 2005.
- [59] E. Furey, "Random Number Generator at <https://www.calculatorsoup.com/calculators/statistics/random-number-generator.php> from CalculatorSoup <https://www.calculatorsoup.com-Onlinecalculators>," 2006-2022. All rights reserved.
- [60] T. J. Dolinsky, P. Czodrowski, H. Li, J. E. Nielsen, J. H. Jensen, G. Klebe, and N. A. Baker, "PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations," *Nucleic acids research*, vol. 35, pp. 522–525, 2007.
- [61] N. Approaches and I. B. Macromolecules, *Computational Electrostatics for Biological Applications*. 2015.