

Tesis Doctoral

Optimización en el cálculo de perfiles de energía libre. Desarrollo y aplicación a reacciones de interés biológico

Ramírez, Claudia Lilián

2016-12-16

Este documento forma parte de la colección de tesis doctorales y de maestría de la Biblioteca Central Dr. Luis Federico Leloir, disponible en digital.bl.fcen.uba.ar. Su utilización debe ser acompañada por la cita bibliográfica con reconocimiento de la fuente.

This document is part of the doctoral theses collection of the Central Library Dr. Luis Federico Leloir, available in digital.bl.fcen.uba.ar. It should be used accompanied by the corresponding citation acknowledging the source.

Cita tipo APA:

Ramírez, Claudia Lilián. (2016-12-16). Optimización en el cálculo de perfiles de energía libre. Desarrollo y aplicación a reacciones de interés biológico. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires.

Cita tipo Chicago:

Ramírez, Claudia Lilián. "Optimización en el cálculo de perfiles de energía libre. Desarrollo y aplicación a reacciones de interés biológico". Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires. 2016-12-16.

EXACTAS UBA

Facultad de Ciencias Exactas y Naturales



UBA

Universidad de Buenos Aires



UNIVERSIDAD DE BUENOS AIRES
Facultad de Ciencias Exactas y Naturales
Departamento de Química Inorgánica, Analítica y Química Física

Optimización en el cálculo de perfiles de energía libre. Desarrollo y aplicación a reacciones de interés biológico

Tesis presentada para optar al título de Doctora de la Universidad de Buenos Aires
en el área de Química Inorgánica, Química Analítica y Química Física.

Autora

LIC. CLAUDIA LILIÁN RAMÍREZ

Director de Tesis

DR. MARCELO MARTÍ

Consejero de Estudios

DR. FABIO CUKIERNIK

Lugar de Trabajo

Instituto de Química Física de los Materiales, Medio Ambiente y Energía
y Departamento de Química Inorgánica, Analítica y Química Física
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

Ciudad Autónoma de Buenos Aires, 2016

Fecha de Defensa:

Optimización en el Cálculo de Perfiles de Energía Libre. Desarrollo y Aplicación a Reacciones de Interés Biológico

Uno de los objetivos principales de la química es comprender los principios que rigen las reacciones, tanto en términos de sus mecanismos como en la termodinámica que las gobierna. Utilizando dinámica molecular, en combinación con esquemas de muestreo sesgado, es posible simular y obtener perfiles de energía libre asociados a eventos que incluyen desde reacciones químicas en entornos complejos (en solución y/o enzimas) hasta movimientos conformacionales proteicos. Los primeros requieren un tratamiento cuántico, al menos del sitio activo, por lo que los métodos híbridos cuántico-clásicos son los escogidos. Mientras que para los segundos es posible emplear campos de fuerza clásicos, atomísticos o de grano grueso. Sin embargo, el costo computacional de la obtención de dichos perfiles es muy elevado para la gran mayoría de los sistemas biológicos, incluso imposibilitando en algunos casos su obtención. Siendo que en un perfil de energía libre se obtiene toda la termodinámica relevante para el sistema, y por ende todos los parámetros asociados de importancia, subsanar esta imposibilidad es de suma importancia. En esta tesis se realizaron desarrollos teóricos que permiten obtener perfiles de energía libre en menor tiempo computacional, para diversos niveles de teoría, permitiendo así el estudio termodinámico de procesos alostéricos y mecanismos de reacción de sistemas biológicos relevantes.

Palabras clave: Energía Libre, Mecanismo de Reacción, Alostерismo, Simulación Computacional

Optimization in Free Energy Profiles Calculation. Development and Application to Biological Relevant Reactions

One of the main goals of chemistry is to understand the underlying principles of chemical reactions, in terms of both its reaction mechanism and the thermodynamics that govern it. Using hybrid Molecular Dynamics in combination with a biased sampling schemes, it is possible to simulate and obtain the Free Energy Profiles associated with biological events, ranging from chemical reactions occurring inside complex environments (enzyme or aqueous solution) to protein conformational movements. The first requires a quantum treatment, at least for the active site, this is why Hybrid Quantum/Classical methods are chosen. While for the second, it is possible to only use classical force fields, either atomistic or coarse grain. However, the computational cost associated with the obtention of such Free Energy Profiles is enormous for most biological systems, making it even impracticable in some cases. Since the free energy profile offers all the relevant thermodynamics information for a system, and therefore all associated parameters of relevance, being able to obtain it, is of utmost importance. In this thesis we developed theoretical methodologies that allow to obtain Free Energy Profiles in less computational time, for different levels of theory, thus allowing the thermodynamic study of allosteric processes and reaction mechanisms of relevant biological systems.

Key Words: Free Energy, Reaction Mechanism, Allosterism, Computational Simulation

DEDICATORIA

A todas las feministas que lucharon y luchan por un mundo más libre, en el que todas podamos vivir como deseamos. A todos los feministas que entendieron que la mujer es un ser pensante y apoyaron el sufragio femenino. A todes las mujeres y hombres que no bajan los brazos y, pese a las muchas dificultades, consiguieron, entre muchos otros derechos, que hoy yo pueda estudiar y trabajar de lo que deseo. Mi eterna gratitud a todes los que hicieron, hacen y harán de este mundo un lugar más justo para todes, mi vida no sería la misma sin la de ustedes.

ÍNDICE GENERAL

	Página
1 Introducción	1
1.1 Las Proteínas	2
1.1.1 Reacciones Enzimáticas	4
1.1.2 Regulación de Proteínas: Alostерismo	4
1.2 Visión Fisicoquímica de los Fenómenos Biológicos	6
1.2.1 Termodinámica y Energía Libre	7
1.2.2 Simulación Molecular: Una respuesta a las preguntas correctas . . .	10
1.3 Objetivos y Lineamientos de esta Tesis	16
2 Metodología Computacional	17
2.1 Acerca de Simular	17
2.2 Métodos Cuánticos	19
2.2.1 El Teorema Variacional	21
2.2.2 Teoría del Funcional de la Densidad	22
2.2.3 Métodos Semiempíricos	25
2.3 Métodos Clásicos	27
2.3.1 Campo de Fuerzas de AMBER	28
2.3.2 Métodos de Grano Grueso	31
2.4 Métodos Híbridos Cuántico-Clásicos	33
2.5 Explorando la Superficie de Energía de un Sistema	35
2.5.1 Dinámica Molecular	35
2.5.2 El Perfil de Energía Libre	38
2.5.3 Igualdad de Jarzynski	39
3 HyDRA	45
3.1 El Esquema de Relajación Híbrido Diferencial	45
3.1.1 Fundamentos Teóricos	46
3.1.2 Implementación	47

3.1.3	HyDRA en AMBER	49
3.2	Validación del Esquema Propuesto	51
3.2.1	La Reacción Modelo: Corismato a Prefenato	52
3.2.2	Perfiles de Energía Libre de Referencia	54
3.2.3	Aplicación de HyDRA en Solución	55
3.2.4	Aplicación de HyDRA en la Enzima BsCM	61
3.3	Análisis del Costo Computacional	62
3.3.1	Poder Catalítico	65
3.4	Discusión	66
3.5	Conclusiones	67
4	Aplicación de HyDRA	69
4.1	Oxidación de Metanotiol por Peróxido de Hidrógeno	69
4.1.1	Condiciones de Simulación	70
4.1.2	Resultados	71
4.2	Hidrolasas de Zinc	71
4.2.1	Condiciones de Simulación	72
4.2.2	Resultados	73
4.3	Discusión	77
4.4	Conclusiones	79
5	Coordenadas de Reacción para Cambios Conformationales	81
5.1	La Problemática	81
5.2	La Desviación Cuadrática Media como coordenada de reacción	82
5.2.1	Implementación en AMBER	84
5.3	El Caso de la Hemoglobina	86
5.3.1	Validación del Modelo CG del Hemo	88
5.4	Resultados: Transición R-T en Hb	89
5.4.1	Modos Esenciales	93
5.4.2	Ángulo de Rotación	95
5.4.3	Switch de la Histidina	96
5.5	Discusión	97
5.6	Conclusiones	99
6	Conclusiones y Perspectivas Futuras	101
6.1	Conclusiones Generales	101
6.1.1	Esquema HyDRA	102
6.1.2	Coordenada de Reacción para Cambios Conformationales	104

6.2 Perspectivas Futuras	105
Bibliografía	109
A Publicaciones	119
B Formación Académica durante el Doctorado	121
C Agradecimientos	125

INTRODUCCIÓN

“Soy de las que piensan que la ciencia tiene una gran belleza. Un científico en su laboratorio no es sólo un técnico. También es un niño colocado ante fenómenos de la naturaleza que le impresionan como un cuento de hadas.”

- Marie Curie

La ciencia es el conjunto de conocimientos obtenidos mediante la observación y el razonamiento, sistemáticamente estructurados, y de los que se deducen principios y leyes generales con capacidad predictiva y comprobables experimentalmente. Cuando el objeto de estudio es la *naturaleza*, el campo se conoce como Ciencias Naturales y abarca cinco grandes disciplinas: física, química, biología, geología y astronomía. Estas divisiones no son tales, ya que ellas se relacionan, se comunican, se cruzan y se *necesitan* entre sí.

En este trabajo de tesis, son tres las ramas de la ciencia que se cruzan: la física, desde el estudio de las propiedades e interacciones de la materia y energía; la química, mediante el estudio de la estructura y de los cambios que experimenta la materia; y la biología, desde el estudio de los seres vivos. El trabajo aquí presentado tiene como objetivo último aportar herramientas a la comprensión de los *cómo* en la biología de proteínas, es por esto que se tomarán conceptos de la física, adecuados al contexto químico, para desarrollar metodologías que ayuden a explicar cómo las proteínas cumplen su función de catalizadoras. Debido a que, a pesar de que mucho se ha realizado en esta área, se mantiene aún vigente una pregunta: *cómo funcionan las proteínas?*.

1.1 Las Proteínas

Entre la gran variedad de biomoléculas presentes en los seres vivos, las proteínas juegan un rol principal, siendo las biomoléculas más versátiles y diversas, se encuentran en el centro de la acción en los procesos biológicos. Esto se debe en gran medida a las infinitas conformaciones posibles que pueden lograrse mediante la combinación de sus unidades estructurales simples repetitivas, denominadas aminoácidos, que se unen mediante enlaces peptídicos ¹. La secuencia que determina esta combinación para todos los seres vivos está determinada mayoritariamente por su genética (con excepción de algunos péptidos antimicrobianos de síntesis no ribosomal), es decir, que la información genética determina en gran medida qué proteínas tiene una célula, un tejido y un organismo. Debido a esto, las proteínas se sintetizarán dependiendo de cómo se encuentren regulados los genes (una porción de ADN) que las codifican, y por ende serán susceptibles a señales o factores externos, pudiendo también sufrir modificaciones químicas postraduccionales (antes de que la proteína sea funcional en la célula, o como parte de mecanismos de control).

Luego de la biosíntesis proteica (y las posibles modificaciones postraduccionales), las proteínas se pliegan espontáneamente, favorecidas por la termodinámica, formando su estructura tridimensional. La misma se encuentra determinada por la secuencia de aminoácidos y está directamente correlacionada con la función biológica, por tanto el plegamiento incorrecto de una proteína lleva a un funcionamiento incorrecto o nulo de la misma. El proceso de plegado se logra gracias a las interacciones de no enlace que se forman entre los diferentes residuos que componen la proteína ²: interacciones de Van der Waals, carga-carga, hidrofóbicas y enlaces hidrógeno. La gran mayoría de las proteínas se pliegan en tres estructuras dimensionales comunes, conocidas como *Hélices- α* , *Hojas- β* (paralelas y antiparalelas) y *Giros* (figura 1.1), o *loop*.

Las proteínas pueden además trabajar juntas para cumplir una función particular, a menudo asociándose para formar complejos proteicos estables, es decir, una proteína de muchos dominios (o subunidades). Debido a esto, las proteínas son componentes estructurales esenciales de las células y son capaces de realizar funciones de regulación, supervisión extracelular y de las condiciones intracelulares y transmisión de información a otros componentes celulares. Además, casi todas las transformaciones moleculares que definen el metabolismo celular están mediadas por proteínas catalizadoras ³ (enzimas). Un listado completo de las

¹El enlace peptídico es un enlace entre el grupo amino ($-\text{NH}_2$) de un aminoácido y el grupo carboxilo ($-\text{COOH}$) de otro aminoácido.

²Estas interacciones deben ser muy fuertes, de forma de incrementar la energía interna, para poder contrarrestar la disminución de entropía asociada a la pérdida de grados de libertad del sistema.

³Un catalizador es una sustancia que incrementa la velocidad de una reacción química y se recupera sin cambios esenciales al final de la reacción.

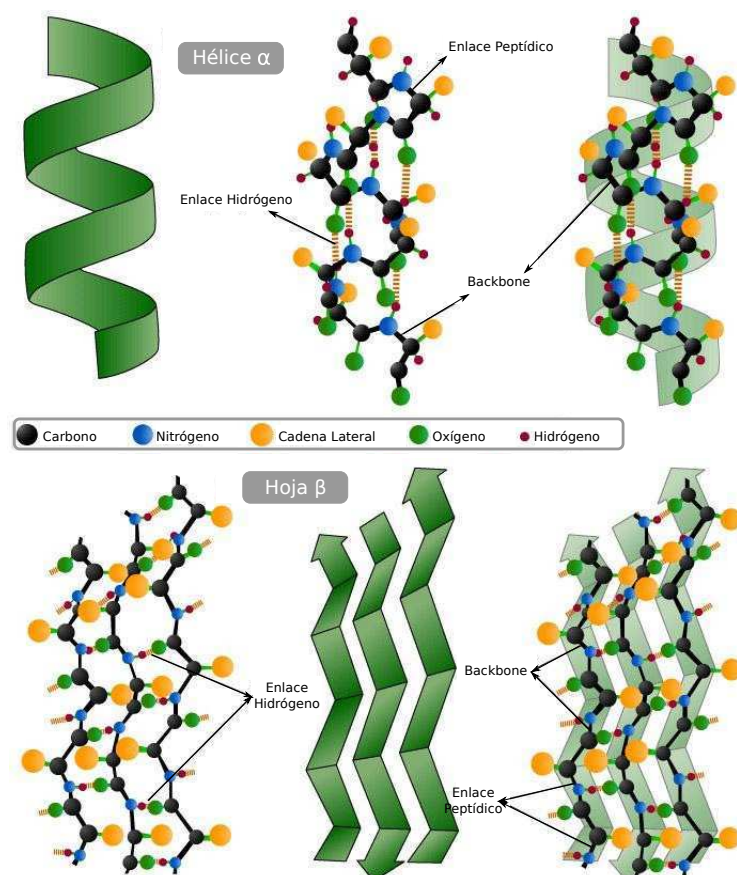


Figura 1.1: Representación de las estructuras secundarias posibles para una proteína. En la hélice alfa (figura superior), se forman puentes hidrógeno entre los átomos de oxígeno del grupo carbonilo de un aminoácido y el átomo de hidrógeno del grupo amino de otro aminoácido situado a cuatro aminoácidos de distancia en la cadena. Las cadenas laterales se extienden hacia afuera de la hélice. En la estructura de hoja beta (figura inferior), dos cadenas de aminoácidos se posicionan paralelas y los grupos amino de una forman enlaces hidrógeno con los grupos carbonilo de la opuesta, dejando las cadenas laterales sobre y bajo el plano de las láminas.

funciones de proteínas conocidas contendría muchos miles de entradas, incluyendo las proteínas que transportan otras moléculas y proteínas que generan fuerzas mecánicas y electroquímicas. Y tal lista no contendría las miles de proteínas cuyas funciones todavía no están totalmente caracterizadas o, en muchos casos, son completamente desconocidas. Es por tanto, ésta un área de continua investigación y desarrollo.

Como los eventos celulares que dependen del funcionamiento de proteínas son vastos, las líneas de investigación contenidas en este espacio multidisciplinario son muchas, incluyendo importantes temas como el desarrollo racional de drogas, el plegamiento de proteínas o la conversión biológica de la energía. En esta tesis se aborda otro tópico de gran importancia, como lo es la descripción de las bases moleculares que determinan la capacidad catalítica de una enzima, y cuáles son los mecanismos que subyacen este proceso, así como también los procesos que actúan en una enzima cuando está se encuentra sujeta a un cambio conformacional.

1.1.1 Reacciones Enzimáticas

Los organismos vivos realizan una enorme variedad de reacciones químicas. Inicialmente, debido a la incapacidad de reproducir en el laboratorio la mayor parte de las reacciones bioquímicas, Louis Pasteur y otros científicos asumieron que los sistemas vivos poseían una “fuerza vital” que les permitía violar las leyes naturales (cinéticas y termodinámicas). En 1878 se comenzaría a utilizar el término “enzima”, que proviene del griego *en-zyme*, para denotar que había algo *en* la levadura (*zyme*), no la levadura en sí, que catalizaba el proceso de fermentación. Sin embargo, no sería hasta 1926 que James Sumner establecería que las enzimas son proteínas, mediante la purificación y cristalización de la enzima ureasa. Las enzimas se definen entonces como moléculas de naturaleza proteica que catalizan reacciones químicas, conocidas como reacciones enzimáticas.

Lo que las vuelve de vital importancia para los seres vivos es que, si bien muchos de los componentes moleculares de un organismo se encuentran favorecidos termodinámicamente para reaccionar entre sí, sólo una fracción reducida de los mismos lo hace. Esto se debe a que la velocidad de cualquier reacción no depende de la diferencia de energía entre los estados final e inicial, sino del camino por el cual se llega de un estado a otro, es decir el esquema que describe la ruptura/formación de enlaces, incluyendo postulados de estados de transición, denominado *mecanismo de reacción*. De esta forma las enzimas, como todos los catalizadores, funcionan disminuyendo la energía de activación ($\Delta G^\ddagger$) de una reacción. Esto lo logran mediante la interacción física de los grupos funcionales presentes en el sitio activo de la enzima (región donde ocurre la reacción) con los reactivos o productos, de manera tal que proveen un mecanismo de reacción más favorable (con un estado de transición de menor energía). Son además extremadamente selectivas con sus sustratos, acelerando sólo ciertas reacciones, de esta forma el conjunto de enzimas presentes en una célula determina el tipo de metabolismo que tiene esa célula. Así, una gran variedad de enzimas actúan como mediadoras del flujo de energía en cada célula, reuniendo, acumulando o utilizando energía para producir trabajo dentro de la célula, o bien transfiriéndola a otras moléculas. Y si bien se podría pensar que la energía se acumula sólo en enlaces químicos, las enzimas son capaces de transformar la energía en calor, trabajo eléctrico e incluso mecánico, de acuerdo a las necesidades del organismo. Por estas razones estudiar, sistematizar y comprender los diferentes mecanismos enzimáticos es uno de los objetivos principales dentro de la biología.

1.1.2 Regulación de Proteínas: Alostерismo

Un organismo debe ser capaz de controlar la actividad catalítica de sus componentes enzimáticas, de manera de poder coordinar los múltiples procesos metabólicos, responder a

cambios en su entorno, y crecer y diferenciarse, todo de forma ordenada. Hay dos formas de que esto ocurra:

- I. Controlar la disponibilidad de la enzima. La cantidad de enzima presente en una célula depende de la relación entre las velocidades de síntesis y degradación. Éstas se encuentran directamente controladas por la célula y pueden sufrir grandes variaciones en cuestión de minutos (en bacterias) u horas (en organismos superiores).
- II. Controlar la actividad enzimática. La actividad de una enzima puede inhibirse mediante la acumulación de producto o, como es de esperarse, por la presencia de inhibidores. Puede además ser modulada, positiva o negativamente, a través de modificaciones estructurales que influyencien su afinidad por un sustrato. Cuando esto ocurre gracias a la unión de un ligando pequeño, el fenómeno se conoce como *alosterismo*.

Podremos decir que una enzima alostérica cambia su afinidad de unión a un ligando en el sitio activo, cuando une un *efector* en un sitio diferente de la enzima, conocido como sitio alostérico. La regulación alostérica, además de controlar la actividad enzimática, interviene en el proceso de comunicación dentro de la célula gracias a la interacción de largo alcance que es posible realizar de esta forma. Esta “acción a distancia” es un concepto clave en el entendimiento del alosterismo y es la razón por la cual este fenómeno resulta de gran interés dentro del campo de la bioquímica.

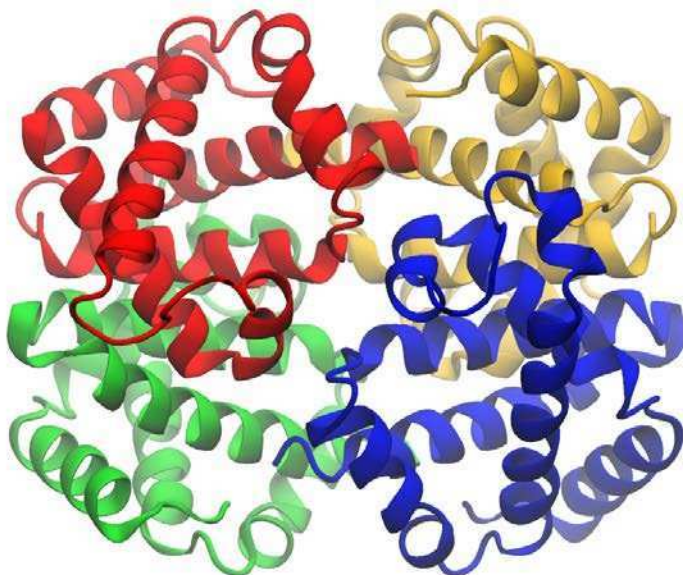


Figura 1.2: Estructura de la Hemoglobina, donde se representan con diferentes colores las 4 subunidades, o monómeros, que la componen. Esta enzima es un caso modelo de complejo proteico que sufre un cambio conformacional para poder llevar a cabo un proceso.

Los primeros estudios en describir estos procesos datan de los años 60, y se basaron en los resultados obtenidos para la Hemoglobina. En los mismos se postulaba que las subunidades de la enzima podían existir en dos conformaciones: Tensa (T) o Relajada (R), y que

las subunidades relajadas eran capaces de unir sustrato de forma más eficiente que las tensas. Se proponían además dos modelos diferentes para explicar el proceso de alosterismo, que continúan vigentes hoy en día. El primero, propuesto por Monod, Wyman y Changeux [53] en 1965 es el modelo *concertado*, o de equilibrio poblacional, en el cual se postula que las subunidades de la enzima están conectadas de forma tal que el cambio conformacional en una es necesariamente transferido a todas las subunidades. De esta forma las subunidades existen, obligatoriamente, en la misma conformación. Se plantea además que existe un equilibrio, en ausencia de ligando, que favorece uno de los estados conformacionales (sea R o T), y que se revierte cuando el mismo se une al sitio alostérico. Un año después, Koshland, Nemethy y Filmer [45] postularían el modelo secuencial que se opone al concertado, planteando que no existe en la enzima una conexión tal que induzca el cambio conformacional de una subunidad a otra. En vez de esto sugieren que el cambio conformacional se produce luego de la unión del sustrato, siguiendo un protocolo de *cambio inducido* (*induced fit*) en el que la enzima se amolda al sustrato disminuyendo la energía. Por esta razón proponen que los cambios conformacionales se producen en cada subunidad luego de la unión del ligando, y no se transmiten a las subunidades a través de la enzima. Es así que, en este modelo, pueden coexistir subunidades en diferentes estados. Ambos modelos continúan vigentes debido a que diferentes enzimas se comportan según un modelo u otro, por lo que es necesario estudiar cada caso en particular.

1.2 Visión Fisicoquímica de los Fenómenos Biológicos

Uno de los mayores retos de la biología del siglo XXI es alcanzar el nivel cuantitativo de la física y la química. Es decir, entender en términos cuantitativos problemas biológicos específicos a distintos niveles de complejidad, desde la escala de molécula única hasta la de ensamblajes macromoleculares subcelulares, con el fin de predecir, controlar y manipular funciones biológicas esenciales. Para esto es necesario que la física, la química y la biología trabajen en conjunto, conectando conceptos de estas disciplinas, para poder hallar respuestas más detalladas del comportamiento de los sistemas biológicos. Así, los conocimientos y enfoques acumulados en la física y la química básica pueden aplicarse al estudio de sistemas biológicos, y la resolución de problemas biológicos podrá estimular el desarrollo de nuevas teorías dentro de estas disciplinas. Un ejemplo de esto es el estudio de reacciones químicas catalizadas por enzimas, donde los cambios energéticos que ocurren requieren un enfoque físico teórico más profundo, que sería (en la gran mayoría de los casos) imposible llevar a cabo si no fuera por las simplificaciones que es posible realizar gracias al entendimiento bioquímico del sistema. Los cambios energéticos que ocurren durante una reacción química

catalizada por una enzima, o fenómenos como el acoplamiento químico-osmótico parecen requerir más de un enfoque físico teórico profundo que de una evaluación biológica. Es así que el estudio de las proteínas, con las herramientas y la visión de la fisicoquímica, ha constituido un gran aporte en el avance de la descripción de sistemas biomoleculares, y estos últimos han servido de inspiración para desarrollar nuevas teorías fisicoquímicas.

En cuanto a las perspectivas que esta disciplina despierta y afronta, existen muchos grupos de investigación en el mundo -y cada vez más en Latinoamérica- que buscan las respuestas vinculadas al comportamiento de proteínas en sistemas biológicos, en el mundo microscópico. Sin embargo, antes de esto, es necesario ir un paso atrás para entender los principios fundamentales de la termodinámica que rigen a los sistemas biológicos.

1.2.1 Termodinámica y Energía Libre

Las actividades normales de los organismos vivos (moverse, crecer, reproducirse) demandan un suministro de energía casi constante. Incluso estando en reposo, los organismos dedican una porción considerable de su aparato bioquímico a la adquisición y utilización de la energía. El estudio de la energía y sus efectos sobre la materia cae bajo el ámbito de la termodinámica y, a pesar de que los sistemas vivos presentan algunos desafíos prácticos para el análisis termodinámico, la vida obedece a las leyes de la misma. Por lo que comprenderlas es de suma importancia no sólo para describir un proceso particular en términos cuantitativos, tal como un reaccionaria bioquímica, sino también para predecir si ese proceso podrá o no ocurrir.

La primera ley de la termodinámica establece que la energía de un sistema ⁴ no puede crearse ni destruirse, sino que se transforma. Esto implica que la energía se conserva. Cuando un sistema sufre un cambio, por ejemplo, una parte de su energía se emplea para realizar trabajo. Por esta razón es usual escribir esta ley expresando que el cambio de energía (ΔU) en el sistema es igual a la diferencia entre el calor (Q) absorbido y el trabajo (W) realizado sobre el entorno,

$$(1.1) \quad \Delta U = Q - W$$

de esta forma el calor se asocia con el movimiento molecular aleatorio, mientras que el trabajo, se relaciona con el movimiento organizado. Es decir que, si un sistema sin posibilidad de intercambiar calor (cerrado adiabático) evoluciona de un estado inicial a uno final, el trabajo realizado no dependerá del proceso que lleva al sistema de un estado a otro. Es así que se suele emplear el primer principio de termodinámica para conectar energéticamente dos

⁴En termodinámica un sistema es la porción de universo de interés, mientras que el entorno es todo aquello que no es el sistema.

estados diferentes de un sistema. Debido a que la presión, P se mantiene constante en la mayoría de los procesos biológicos, resulta de utilidad definir la cantidad *Entalpía*, H como $U + PV$ (donde V es el volumen del sistema), de esta forma la variación de entalpía resulta igual a

$$(1.2) \quad \Delta H = \Delta U + P\Delta V$$

donde $P\Delta V$ es el trabajo realizado en la expansión de un gas. En numerosos casos, la variación de volumen es además muy pequeña, haciendo que la variación de entalpía sea prácticamente igual a la de energía interna.

Sin embargo, el primer principio de la termodinámica no puede determinar por sí mismo si un proceso será o no espontáneo. En el camino a predecir la espontaneidad de un dado proceso nos encontramos con la segunda ley de la termodinámica. La misma es probablemente la ley más importante en el campo de la física y, extrañamente, no posee una única definición sino que, dependiendo de las áreas, asume diferentes enunciados.

En química la definición más usual establece que:

Definición 1.1. Los procesos espontáneos producen estados de mayor desorden en el universo, mientras que los reversibles no modifican el desorden del universo.

En este contexto, el desorden se mide a través de una función de estado denominada *Entropía*, S . Esta magnitud se definirá en termodinámica clásica como el cociente entre el calor transmitido y la temperatura a la que se transmite. Mientras que a nivel atómico-molecular (campo de estudio de la mecánica estadística), el concepto de entropía está relacionado con la información disponible sobre el conjunto de partículas que componen el sistema. A esta escala, el concepto de *mayor desorden* (entropía elevada) es equivalente a pensar que se dispone de poca información sobre el estado individual de las partículas que integran el sistema. En otras palabras, un sistema con alta entropía posee un número elevado de microestados equivalentes posibles, perdiéndose de esta manera la información detallada sobre los elementos del conjunto (se conoce menos precisamente el estado de cada partícula). Por ejemplo, si consideramos un sistema gaseoso, la entropía es elevada debido a que se ha perdido información sobre los estados individuales de las partículas que lo componen. Por el contrario, cuando se trata de un sistema ordenado, de entropía baja, siempre se dispondrá de vasta información sobre el mismo. Este es el caso de un sólido, en el que la información microscópica disponible sobre el mismo es tal que permite conocer como están distribuidos espacialmente todos los átomos, a partir de la geometría de la red. Es así que la entropía se relaciona directamente con el número de microestados accesibles dentro de un sistema. Sería Ludwig Boltzmann en el siglo XIX el que desarrollaría los fundamentos teóricos para

finalmente expresar mediante una ecuación la relación entre ambas cantidades:

$$(1.3) \quad S = k_B \ln \Omega$$

donde Ω es el número de *arreglos* (o configuraciones) de microestados posibles y k_B es la constante de Boltzmann (igual al cociente entre la constante de los gases y el número de Avogadro).

Del segundo principio se desprenden una serie de corolarios que relacionan los microestados accesibles para un sistema y la entropía:

Corolario 1.1. *En un sistema aislado, ningún proceso puede ocurrir si a él se asocia una disminución de la entropía total del sistema.*

Corolario 1.2. *Ningún proceso cíclico es tal que el sistema en el que ocurre y su entorno puedan volver a la vez al mismo estado del que partieron.*

De esta forma el segundo principio marca la dirección en la que deben llevarse a cabo los procesos termodinámicos y, por lo tanto, la imposibilidad de que ocurran en el sentido contrario, imponiendo restricciones a las transferencias de energía que hipotéticamente pudieran llevarse a cabo teniendo en cuenta solo el primer principio. Por ejemplo, la sal una vez disuelta en agua no podrá volver a solidificarse espontáneamente sin alterar el sistema inicial “*sal + agua*”. Así podremos decir que el segundo principio constituye un criterio de espontaneidad y, si se conociera la variación de entropía del universo vinculada a cualquier proceso, se podría predecir si éste va a ocurrir o no. Sin embargo, el cambio de entropía del universo es un criterio de espontaneidad poco conveniente, debido a que no depende únicamente de variables termodinámicas del sistema, sino que también incluye cambios ocurridos en el ambiente

$$(1.4) \quad \Delta S_{\text{universo}} = \Delta S_{\text{sistema}} + \Delta S_{\text{medio}} \quad \text{con} \quad \Delta S = S_f - S_i$$

donde vale la pena remarcar que el cambio de entropía asociado a cualquier proceso se considera desde un estado inicial i hasta un estado final f ya que la entropía es función de estado. La forma entonces de fijar un criterio de espontaneidad consiste en combinar el primer principio, a temperatura y presión constante, con la definición clásica de entropía, es decir

$$(1.5) \quad \Delta S \geq \frac{Q}{T} = \frac{\Delta H}{T} \quad \Rightarrow \quad \boxed{\Delta H - T\Delta S \leq 0}$$

Este es el criterio de espontaneidad formulado por Josiah Gibbs en 1878, para el cual definió además la cantidad G , conocida como energía libre de Gibbs, según

$$(1.6) \quad G = H - TS$$

Entonces, debido a que es esta también una función de estado, para procesos a temperatura y presión constante, se podrá definir un criterio de espontaneidad acorde a la energía libre de la forma:

$$(1.7) \quad \Delta G = \Delta H - T\Delta S < 0$$

donde los procesos en estado de equilibrio tendrán un valor de ΔG igual a cero. De la definición de energía libre (de Gibbs), podemos notar que existen dos cantidades que determinarán la espontaneidad del proceso. Es así que en un proceso en el cual la entalpía aumenta ($\Delta H > 0$), desfavoreciendo la reacción, podrá ocurrir si el aumento de entropía del sistema es tal que compensa el de la entalpía. De manera equivalente, el sistema puede estar disminuyendo su entropía (nótese que hablamos de la entropía del sistema y no del universo), pero el proceso puede ocurrir gracias a la disminución de entalpía asociada. Por esta razón es de gran importancia determinar los valores de energía libre para poder predecir apropiadamente si una reacción ocurrirá o no espontáneamente y bajo que condiciones lo hará. Sin embargo, calcular estos valores no es sencillo debido a que, como mencionamos con anterioridad, la energía libre estará ligada no solo a la energía interna del sistema, sino también a todas las conformaciones posibles del mismo, complejizando enormemente el cálculo de esta magnitud.

1.2.2 Simulación Molecular: Una respuesta a las preguntas correctas

Entender los principios que rigen las reacciones químicas es uno de los objetivos principales dentro de las ciencias químicas. A un nivel atómico y molecular, una reacción es entendida en términos de la termodinámica que la gobierna y el camino de reacción que la describe. Es decir, que una reacción química será descrita en términos del movimiento atómico que conduce un sistema de reactivos a productos, y el perfil de energía libre que acompaña esta transición. Experimentalmente, la dinámica química será inferida de estudios de cinética química, que ayudan a definir el mecanismo de reacción y la energía libre relativa de los estados involucrados. Sin embargo, es usual proponer diversos mecanismos de reacción para un único par reactivo-producto y, salvo contadas excepciones, el detalle atomístico no podrá conocerse con precisión mediante ningún experimento. Para avanzar en este campo, y ser realmente capaces de comprender un proceso a nivel atómico, en las últimas décadas ha surgido una subdisciplina de la biofísicoquímica denominada simulación molecular, que intenta explicar las propiedades químicas de las biomoléculas a través de su estructura y sus propiedades dinámicas y de equilibrio, mediante el modelado computacional del sistema. Es decir, es el uso de matemáticas, física e informática para estudiar el comportamiento bioquímico de sistemas complejos mediante la simulación por computadora. Para que los

resultados sean relevantes, se debe ser capaz de realizar una correcta descripción del sistema de estudio mediante las leyes de la física, la química y la biología, y plasmar estas de manera apropiada en un modelo. La simulación molecular consiste, de este modo, en el empleo de modelos fisicoquímicos (que se valen de herramientas matemáticas) implementados en un código o programa de computadora, con el objeto de poder realizar predicciones acerca del comportamiento del sistema molecular que se esté modelando.

En la actualidad, la simulación molecular es una herramienta de gran importancia en el estudio de las reacciones químicas y bioquímicas y sus mecanismos subyacentes. Debido a esto existen, dentro de la química computacional, diversos modelos que permiten una correcta descripción del sistema bajo estudio, como ser los métodos cuánticos, clásicos y de grano grueso, en conjunto con los híbridos, que permiten combinar más de una representación para un dado sistema. La diversificación de los modelos tiene su origen en la diversidad misma de los sistemas de estudio: fenómenos en escalas temporales que van desde los pocos picosegundos hasta incluso segundos en sistemas en la escala de nanómetros o incluso milímetros, hacen que sea imposible abordar tan diversas problemáticas con un único modelo físico-químico. Ejemplos de esta diferenciación, y los modelos que se emplean en cada caso, son

- Espectros IR, donde la necesidad de describir las interacciones electrónicas apropiadamente hará necesario el empleo de modelos de naturaleza cuántica;
- Interacciones entre Dominios Proteicos, en los cuales la posibilidad de tratarlos a nivel cuántico es inexistente debido a su costo computacional. A su vez la contribución de los electrones a la descripción del sistema no es fundamental, por lo que una descripción clásica de los átomos permitirá su correcto estudio;
- Reacciones Enzimáticas, para las que será necesario combinar modelos cuánticos -para describir el sistema reactivo- y modelos clásicos -para ser capaces de describir apropiadamente el entorno proteico-, conocidos como métodos híbridos cuántico-clásicos;
- Proteínas de Membrana, en las que es necesario simular la proteína y la membrana en la cual se inserta, haciendo que el tamaño de estos sistemas imposibilite incluso un tratamiento a nivel clásico, debido a su costo computacional, por esta razón se emplean modelos de grano grueso, en los que un conjunto de átomos se representan con una única partícula;
- Plegamiento de Proteínas, que pueden requerir escalas temporales del orden de los milisegundos, haciendo necesaria una representación mínima del sistema -siendo

necesario incluso en muchos casos, que cada aminoácido sea representado sólo por una partícula, considerando únicamente las interacciones nativas (modelos tipo Go) - para poder observar el fenómeno en su totalidad.

Por esta razón, el modelo será escogido dependiendo del fenómeno de interés, haciendo incluso necesaria la combinación, directa o indirecta, de modelos para describir algunos fenómenos, como se verá con más detalle en el capítulo 2.

1.2.2.1 El Cálculo de Energía Libre

Para aquellas reacciones que ocurren en entornos complejos (solución o enzimas) el objetivo central en simulación molecular es determinar el perfil de energía libre (FEP) asociado a la reacción. Debido a que este puede relacionarse directamente con las propiedades experimentales, como ser velocidades de reacción y constantes de equilibrio, para así describir adecuadamente su comportamiento cinético y termodinámico. Sin embargo, y a pesar de que el cálculo de entalpía para un dado sistema es sencillo, la determinación de la energía libre requiere un muestreo extensivo de las posibles conformaciones del sistema, lo que dificulta considerablemente su obtención, en particular en entornos complejos como biomoléculas solvatadas.

Para describir correctamente la reactividad química, sin perder detalle del entorno proteico, es usual emplear métodos híbridos que combinan mecánica clásica y mecánica cuántica. En los mismos, el sistema reactivo es descripto empleando métodos cuánticos, mientras que el entorno es modelado con mecánica clásica. Esto se debe al aumento no lineal del costo computacional con el tamaño del sistema en los métodos cuánticos, que limita la aplicación de esta metodología a sistemas de unos pocos cientos de átomos, imposibilitando así tratar un sistema del tamaño de una proteína. Aunque existen diversos métodos para afrontar esta problemática [23, 34, 73, 74], la aplicación de métodos híbridos mencionada con anterioridad es probablemente la técnica más empleada en la actualidad [2, 79, 80]. A pesar de esto es importante tener en cuenta que, debido al extenso muestreo configuracional que es necesario para obtener valores de energía libre precisos y al alto costo computacional de las implementaciones QM/MM (requerido para calcular la energía cuántica y las fuerzas), el desarrollo y la validación de estrategias que permitan determinar la energía libre de forma exacta en tiempos de cómputo razonables, es un área de continua investigación.

Para determinar el FEP a lo largo de una coordenada de reacción seleccionada, ya sea utilizando un método cuántico, clásico o híbrido, es usual emplear estrategias de muestreo sesgado o mejorado. El método más usual entre estos es posiblemente la estrategia de muestreo llamada *umbrella sampling*, en la que la coordenada de reacción es dividida en segmentos y el muestreo en dicho segmento se ve favorecido por la adición de un potencial

externo (usualmente armónico). Si bien esta estrategia ha sido capaz de obtener perfiles de energía libre, presenta algunas dificultades técnicas difícilmente superables, como ser la imposibilidad de saber si el error cometido en la energía libre es por exceso o por defecto. Otras estrategias han sido desarrolladas en los últimos años, entre las que se pueden nombrar Metadinámica [47], *Adaptive Biasing Force* [35], *Free Energy Perturbation* [88] y *Orthogonal Space Random Walk* [87], entre otras. Una estrategia reciente, muy poderosa, se centra en realizar múltiples dinámicas moleculares guiadas (MDMG) que conducen al sistema a lo largo de la coordenada de reacción escogida, fuera del equilibrio, mediante una fuerza externa. En cada trayectoria, el trabajo realizado por la fuerza externa sobre el sistema ($W_i(\lambda)$) puede ser computado, y los trabajos de las múltiples trayectorias pueden ser combinados utilizando la ecuación de Jarzynski [40] (Ecuación 1.8) para determinar el perfil de energía libre correspondiente ($G(\lambda)$),

$$(1.8) \quad G(\lambda) = -\beta^{-1} \ln \langle e^{-\beta W_i(\lambda)} \rangle$$

donde $G(\lambda)$ representa el FEP en función de la coordenada de reacción λ , los *brackets* ($\langle \dots \rangle$) representan el promedio de la función que encierran para un ensamble, y $\beta = 1/k_b \cdot T$ donde k_b es la constante de Boltzmann y T la temperatura del sistema. De esta forma, la igualdad de Jarzynski logra relacionar los procesos de no equilibrio con una propiedad de equilibrio, como ser la energía libre.

Esta ecuación, demostrada teóricamente en 1997, fue también validada experimentalmente [52]. La misma requiere de algunas consideraciones prácticas para poder cumplirse, la primera de estas está relacionada con el promedio sobre el ensamble. Dicho promedio se realiza sobre un conjunto de trayectorias de dinámica molecular guiada, cada una de las cuales se realiza a partir de diferentes configuraciones iniciales, extraídas de dinámica de equilibrio. La igualdad expresada en la Ecuación 1.8 se cumple únicamente cuando este promedio se encuentra correctamente convergido, lo que es formalmente cierto sólo cuando se han realizado un número infinito de trayectorias (y computado sus trabajos). En la práctica se lleva a cabo un muestreo finito (que se define por el número de trayectorias realizadas independientes, N_{Tr}) y por lo tanto la energía libre resultante es en realidad un estimador estadístico. Sin embargo, una de las ventajas de este método radica en el hecho de que el estimador devuelto asumirá, en todos los casos, un valor mayor o igual al valor exacto de energía libre, debido a cómo fue construida la igualdad de Jarzynski.

Otro aspecto clave dentro de esta teoría está relacionada con la velocidad a la cual el sistema es llevado de reactivos a productos. Si bien en la teoría no existen restricciones al respecto, en la práctica una velocidad elevada puede conducir a valores erróneos, o puede requerir un número extremadamente elevado de trayectorias a ser considerado dentro del ensamble para arribar al resultado correcto. Para comprender con mayor claridad el concepto

de velocidad de la DMG -que no se encuentra relacionada con la velocidad de la reacción- detallaremos brevemente como se realiza una DMG (los detalles serán presentados más adelante en el Capítulo 2). En la implementación más sencilla que permite realizar una DMG se implementa una fuerza externa $\vec{F}$, que es aplicada a lo largo de la coordenada de reacción, λ . En la práctica es usual emplear una fuerza armónica de la forma

$$(1.9) \quad \vec{F} = -k[\lambda - (\lambda_0 + v \cdot dt)]$$

donde la expresión entre paréntesis representa el mínimo del potencial escogido, el cual recorrerá la coordenada de reacción, desde el valor inicial λ_0 , hasta su valor final, a una velocidad constante v . La velocidad, por tanto, determinará de forma directa la cantidad de pasos (y el costo computacional) necesarios para realizar cada trayectoria, e indirectamente influirá en el número de trayectorias requerido para converger el promedio exponencial. Por un lado, tendremos que si la velocidad es infinitamente lenta el proceso será reversible y sólo será necesaria una trayectoria, ya que el trabajo realizado en este caso será igual a la energía libre. Mientras que si la velocidad es muy elevada, el número de trayectorias necesario para que se cumpla la igualdad en la Ecuación 1.8 también lo será.

Otro aspecto importante de la distribución de trabajos al realizar MDMG es que, para cualquier velocidad, la distribución será más amplia cuanto más nos alejamos de las condiciones de partida. Por lo tanto, para un número fijo de trayectorias, las estimaciones más precisas para la energía libre se encontrarán cercanas a la conformación de partida. Por esta razón es usual [4, 17, 18] realizar la reacción en el sentido termodinámicamente favorable y en el inverso, de manera tal de poder combinar ambos resultados (donde cada uno presenta mayor exactitud) para obtener un perfil exacto y completo de energía libre para la reacción de interés. Las amplias distribuciones de trabajo al realizar DMG se deben a que el sistema se encuentra fuera del equilibrio. Por lo tanto, una estrategia capaz de mantener el sistema cercano al equilibrio, a lo largo de la coordenada de reacción, contribuirá para alcanzar resultados más precisos, con una distribución de trabajo estrecha, mejorando por lo tanto el estimador de energía libre.

En este trabajo de tesis se presentará el desarrollo y la aplicación de una estrategia computacional que permite alcanzar este objetivo. Los resultados de estas simulaciones contribuirán de manera significativa en la comprensión de un proceso de diversas maneras. Por un lado, el poder predecir el comportamiento del sistema real estudiado en respuesta a condiciones cambiantes, permite agilizar el estudio de un sistema gracias a la posibilidad de realizar miles de experimentos simulados por computadora. Esto permitirá luego identificar los experimentos físicos que con mayor probabilidad ayuden al investigador a encontrar la solución al problema bajo estudio, permitiéndole así descartar hipótesis erróneas con antelación, o bien comparar los resultados computacionales con los obtenidos de forma

experimental, de manera de poder demostrar fehacientemente un mecanismo de reacción propuesto. Por otro lado, contar con la información del comportamiento de los átomos durante un proceso de interés nos permite hacer inferencias acerca de la relación entre este último y las interacciones atomísticas, factor clave en numerosos procesos. Debido a que en la gran mayoría de los casos no es posible acceder a este tipo de información mediante experimentos físicos, la simulación molecular se convierte en la única forma de responder este tipo de preguntas, como ser, por ejemplo, las interacciones entre una enzima y su sustrato, o la verificación de un mecanismo de reacción propuesto.

1.2.2.2 Alosterismo en Simulación Molecular

Una de las preguntas más frecuentes en biología está relacionada con las proteínas que realizan cambios conformacionales para llevar a cabo su función. Sin embargo, debido al tiempo de simulación requerido para que el cambio ocurra, el mismo difícilmente es observado en una dinámica convencional. En los últimos años se ha realizado un esfuerzo conjunto dentro de la ciencia para ser capaces de modelar procesos cada vez más complejos y sistemas cada vez mayores. Uno de los enfoques más empleado en la actualidad para afrontar este tipo de problemáticas, tiene que ver con los modelos de grano grueso. En particular, cuando los sistemas de interés exceden los cientos de nanómetros, la reducción en la cantidad de partículas empleada por este modelo, permite estudiar sistemas de esta magnitud. A su vez, debido a que las interacciones se simplifican en este tipo de modelos, la superficie de energía potencial de un dado sistema se ve “suavizada”, es decir que muchos mínimos se ven reducidos o incluso cesan de existir. Esto permite estudiar fenómenos que ocurren en escalas temporales mayores, debido a que el sistema se mueve ahora en una superficie de energía potencial con menos mínimos locales, por lo que el riesgo de quedar atrapado en una de estas conformaciones disminuye. Más aún, la simplificación de las interacciones también permite aumentar el paso de integración (dt), debido a que las vibraciones características ocurren ahora en una mayor escala temporal. A pesar de esto, en muchos casos, el movimiento que realiza la proteína durante un fenómeno es de tal magnitud, que la barrera de activación de estos procesos continúa siendo alta, incluso para los modelos simplificados de grano grueso. Por esta razón estudiar este tipo de procesos continúa siendo, incluso hoy en día, un desafío dentro de la comunidad científica.

Como se mencionó en la Sección 1.1.2, el alosterismo es un fenómeno de gran importancia en el contexto de sistemas biológicos. Por esta razón, desarrollar estrategias que permitan estudiar los sistemas que sufren estas modificaciones conformacionales, permitiría ahondar en la comprensión de un gran número de problemáticas diversas. Es uno de los objetivos de esta tesis, desarrollar un esquema que permita estudiar el alosterismo mediante simulación

molecular, a un costo computacional accesible.

1.3 Objetivos y Lineamientos de esta Tesis

Como se mencionó previamente, este trabajo busca desarrollar esquemas que permitan estudiar diferentes fenómenos biológicos a nivel computacional. Específicamente, los objetivos generales son

- Desarrollar una metodología que permita mejorar la exactitud y precisión de los perfiles de energía libres, de reacciones químicas en entornos complejos, obtenidos mediante MDMG y la ecuación de Jarzynski.
- Reducir el costo computacional asociado al cálculo de perfiles de energía libre cuando se emplean esquemas híbridos cuántico-clásicos, sin perder exactitud.
- Desarrollar un esquema metodológico que permita realizar dinámicas moleculares guiadas capaces de conducir el movimiento conformacional de una proteína.

En este trabajo presentaremos la metodología, describiremos las condiciones de validez y argumentaremos sobre la eficacia de los esquemas desarrollados en el transcurso de esta tesis, todo esto sobre sistemas biológicos de relevancia.

Debido a que este es un trabajo fuertemente metodológico, en el Capítulo 2 sentaremos las bases teóricas necesarias para llevar a cabo la investigación realizada. A su vez, cada capítulo de resultados metodológicos tendrá el desarrollo teórico pertinente a cada esquema desarrollado. En el Capítulo 3 se muestra el desarrollo, validación y aplicación del Esquema de Relajación Híbrido Diferencial (HyDRA), que nos permitió llevar a cabo los dos primeros objetivos. Debido al éxito de esta metodología, se presentan además los resultados obtenidos para otras reacciones químicas de interés en el Capítulo 4. El tercer objetivo es llevado a cabo en el Capítulo 5, donde se presenta una coordenada de reacción capaz de guiar un cambio conformacional en una DMG, dentro de un modelo de grano grueso. Por último, en el Capítulo 6 se resumen las principales conclusiones de esta Tesis, junto a las perspectivas futuras de los temas investigados.

METODOLOGÍA COMPUTACIONAL

En este capítulo se presentarán de modo general y conceptualmente los métodos que serán empleados, y que a su vez representan las bases teóricas para el desarrollo metodológico realizado en la presente tesis. Los detalles relativos a la implementación de los mismos y los aspectos técnicos para la resolución de problemáticas inherentes a cada capítulo, serán detallados en los capítulos correspondientes.

2.1 Acerca de Simular

Desde la invención del ábaco, pasando por la máquina de Turing, hasta los microprocesadores actuales, la especie humana ha ido buscando herramientas que le permitan simplificar los cálculos que precisa realizar para comprender mejor el mundo que lo rodea. No es de extrañar, por tanto, que desde mediados del siglo XX se emplearan procesadores para desarrollar herramientas fisicoquímicas que permitan comprender los sistemas biológicos a nivel molecular. En este contexto, surge el modelado o simulación molecular, donde se toman modelos químico-físicos para representar la naturaleza de las interacciones en los sistemas biológicos. El modelado molecular consiste, de este modo, en una serie de modelos teóricos fisicoquímicos implementados a través de técnicas computacionales en un programa capaz de modelar, imitar y predecir el comportamiento de las moléculas del sistema de interés. La simulación computacional de sistemas moleculares es actualmente una herramienta muy valiosa y ubicua para el estudio de las reacciones químicas y bioquímicas y sus mecanismos subyacentes. Sin embargo, al utilizar métodos de simulación computacional es importante tener en cuenta que éstos se basan en modelos fisicoquímicos, es decir, en representacio-

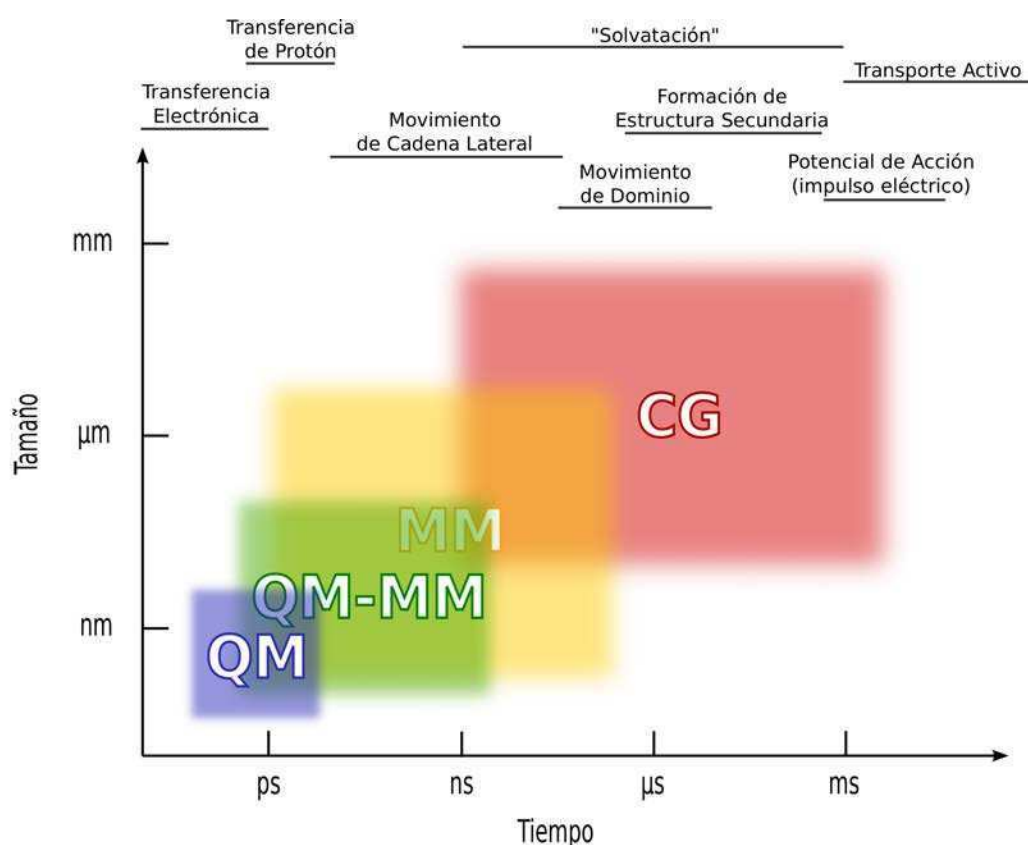


Figura 2.1: Escalas de tamaño y tiempo abordables por los diferentes métodos de simulación molecular.

nes del mundo real que tienen características y limitaciones que los definen y diferencian entre sí, debido al nivel de aproximación utilizado en cada caso. Es preciso entonces identificar primero cuál de estos métodos es el indicado para dar respuesta a los interrogantes fenomenológicos que se planteen (Figura 2.1).

Una categorización primaria de los diferentes métodos permite distinguir entre aquellos que se basan en los principios de la mecánica cuántica (QM, por sus siglas en inglés, *Quantum Mechanics*), y aquellos sustentados en la mecánica clásica. Los primeros son modelos más detallados que consideran de manera explícita la estructura electrónica de los átomos intervinientes en un sistema. Esto permite estudiar y resolver problemas que involucran movimiento de electrones explícitamente (reacciones químicas, transferencia de cargas, estados electrónicamente excitados). Sin embargo, alcanzar este nivel de detalle requiere la resolución de ecuaciones complejas para un gran número de partículas, cuyo tiempo de cómputo (habitualmente referido como costo computacional) resulta elevado, especialmente si el sistema bajo estudio es una biomolécula que, en conjunto con su entorno, puede contener del orden de 10^3 átomos. Para este tamaño de sistema son adecuados, en cam-

bio, los métodos basados en las leyes de la mecánica clásica, caracterizados por un nivel de detalle menor, donde una partícula de identidad parametrizada es descripta mediante un campo de fuerzas. Dentro de estos podemos hacer también una división, aquellos métodos denominados de mecánica molecular (MM, por sus siglas en inglés, *Molecular Mechanics*) que consideran al átomo como la partícula a parametrizar, y aquellos de grano grueso (CG, por sus siglas en inglés, *Coarse Grain*) donde la entidad parametrizable es un conjunto de átomos. Estos modelos son utilizados para resolver un amplio número de problemas referentes a la dinámica de sistemas biomoleculares aunque, evidentemente, presentan la limitación de que no son capaces de representar fenómenos reactivos. Resulta así que, en general, suelen estudiarse sistemas QM de un tamaño aproximado de entre 50 y 150 átomos con tiempos de simulación en la escala de decenas a cientos de picosegundos; en cuanto al número de partículas simuladas, los sistemas MM suelen abarcar hasta 3 órdenes de magnitud más, y tiempos de la escala de cientos de nanosegundos; y los sistemas CG alcanzan escalas de millones de átomos y microsegundos. En los últimos años se han desarrollado además metodologías híbridas, que se basan en la descripción de un sistema multi-escalarmente, es decir, con distinto grado de detalle para diferentes subsistemas del sistema global, intentando aprovechar las bondades de cada metodología acorde al problema de interés. En particular, en esta tesis trabajaremos sólo con el método híbrido cuántico-clásico (QM/MM, por sus siglas en inglés, *Quantum Mechanics/Molecular Mechanics*), en el cual una región del sistema es descripta con QM (generalmente el subsistema que contiene la parte reactiva), mientras que el entorno se describe con MM.

En las secciones siguientes se hará una breve descripción de las metodologías aquí presentadas, así como también las técnicas de muestreo necesarias para explorar la superficie de energía potencial.

2.2 Métodos Cuánticos

A principios del siglo XX la física clásica experimentaría una revolución originada por diversos experimentos, entre ellos aquellos que permitían estudiar la estructura de los átomos. Se hizo evidente que la validez de la física clásica cesa a nivel microscópico, y que eran necesarios nuevos conceptos para describir la estructura de los átomos y moléculas. Esto dio lugar a una serie de avances en el campo, introducidos por Planck, Einstein, Bohr y Compton¹, que conformaron las bases teóricas, así como también la confirmación experimental, de

¹En la búsqueda de describir la radiación de un cuerpo negro, Planck introdujo el concepto de cuantización de la energía en una onda electromagnética. Concepto que empleó Einstein para describir el efecto fotoeléctrico, y Bohr en su descripción del átomo. Sería Compton finalmente el que, al producir scattering de rayos X mediante electrones, confirmaría la teoría corpuscular de las ondas.

la naturaleza corpuscular de las ondas, es decir, el concepto de que las ondas exhiben el comportamiento de partículas a escala microscópica. En esta escala, la física clásica falla no sólo cuantitativa sino también cualitativa y conceptualmente. Como si las cosas no fueran suficientemente malas para la física clásica, de Broglie presentó en 1923 otra nueva y poderosa idea que la física clásica no podía subsanar: postuló que no sólo existe el comportamiento corpuscular de la radiación, sino que, de forma equivalente, las partículas se comportan como ondas. Concepto que fue demostrado experimentalmente por Davisson y Germer en 1927, cimentando así las bases de la mecánica cuántica. Esta teoría acabaría de formalizarse con dos formulaciones teóricas independientes: la mecánica matricial, desarrollada por Heisenberg en 1925, y la mecánica ondulatoria, formulada por Schrödinger en 1926.

La segunda metodología es más intuitiva, ya que describe la dinámica de una partícula microscópica a través de una función de onda, $\Psi(\mathbf{r}, t)$, la cual se obtiene resolviendo la ecuación de Schrödinger dependiente del tiempo para un potencial dado, $V(\mathbf{r}, t)$:

$$(2.1) \quad i\hbar \frac{\partial}{\partial t} \Psi(\mathbf{r}, t) = \left[\frac{-\hbar^2}{2\mu} \nabla^2 + V(\mathbf{r}, t) \right] \Psi(\mathbf{r}, t)$$

donde $\hbar$ es la constante de Planck sobre 2π , μ es la masa reducida del sistema y la expresión entre corchetes es el hamiltoniano del mismo. La resolución de esta ecuación conduce al espectro de energía, obtenido como autovalor, y a la función de onda del sistema de interés, la cual contiene toda la información necesaria para calcular las propiedades del sistema. En 1927 Max Born propuso una interpretación probabilística de la mecánica ondulatoria: el módulo cuadrado de la función de onda, que es solución a la ecuación de Schrödinger, es la densidad de probabilidad electrónica.

Debido a que esta ecuación sólo tiene solución exacta para átomos hidrogenoides, al momento de buscar una solución para esta ecuación en sistemas más complejos, como ser moléculas poliatómicas, es preciso realizar ciertas aproximaciones. Una de las más comunes es suponer que el potencial es independiente del tiempo, de manera que el hamiltoniano queda definido como:

$$(2.2) \quad \hat{H} = \frac{-\hbar^2}{2\mu} \nabla^2 + V$$

Por ende, la ecuación será ahora independiente del tiempo, y tanto la función de onda como los autovalores serán sólo función de las coordenadas espaciales del sistema:

$$(2.3) \quad \hat{H}\Psi(\mathbf{r}) = E\Psi(\mathbf{r})$$

Donde se ha representado de manera explícita la relación entre el hamiltoniano, los autovalores y la función de onda del sistema, descriptos con anterioridad.

Sin embargo, la expresión para el hamiltoniano, aunque independiente del tiempo, sigue siendo sumamente compleja, debido a que es necesario considerar no sólo el movimiento de los electrones sino también de las partículas en el núcleo. Es por esto que para sistemas moleculares se introduce adicionalmente la aproximación de Born-Oppenheimer. La misma utiliza la gran diferencia de masa (y por tanto de momento) entre los electrones y el núcleo para suponer que la función de onda del sistema completo es igual al producto de la función de onda electrónica y la nuclear, es decir, supone que es posible desacoplar el movimiento de los electrones y los núcleos. Es posible entonces dividir el problema en una contribución nuclear y en una electrónica, y resolver la ecuación de Schrödinger para un *sistema nuclear* y un *sistema electrónico*. Generalmente, en los programas de simulación molecular sólo se resuelve la ecuación de Schrödinger para la función de onda electrónica, donde el hamiltoniano electrónico adquiere la siguiente la forma:

$$(2.4) \quad \begin{aligned} \hat{H}_e &= \hat{T}_e + \hat{V}_{eN} + \hat{V}_{ee} \\ &= - \sum_{i=1}^N \frac{\hbar^2}{2m_e} \nabla_{\mathbf{r}_i}^2 - \sum_{i=1}^N \sum_{j=1}^M \frac{Z_j e^2}{4\pi\epsilon_0 |\mathbf{R}_j - \mathbf{r}_i|} - \sum_{i=1}^N \sum_{j>i}^N \frac{e^2}{4\pi\epsilon_0 |\mathbf{r}_i - \mathbf{r}_j|} \end{aligned}$$

donde M es el número de núcleos y N el número de electrones. En este operador se pueden distinguir tres términos, el primero representa la energía cinética electrónica, el segundo es el potencial asociado a la interacción núcleo-electrón y el tercero es el potencial de la interacción electrón-electrón. Será este entonces el hamiltoniano que buscarán resolver los diferentes métodos cuánticos, para diversos sistemas moleculares.

2.2.1 El Teorema Variacional

Aún con las simplificaciones realizadas, obtener la energía de un sistema complejo continúa siendo inviable con las computadoras actuales. Por esta razón se hace uso del principio o teorema variacional, el cual postula que al evaluar la energía con una función de prueba $\varphi(\mathbf{r})$, diferente de aquella que es solución del hamiltoniano electrónico ($\psi(\mathbf{r})$), la energía que se obtendrá será siempre mayor al autovalor que se obtendría con la función de onda correcta, es decir

$$(2.5) \quad \langle \varphi(\mathbf{r}) | \hat{H} | \varphi(\mathbf{r}) \rangle \geq \langle \psi(\mathbf{r}) | \hat{H} | \psi(\mathbf{r}) \rangle = E_0$$

donde E_0 es la energía del estado fundamental. De esta forma la función de onda que corresponde al sistema de interés puede aproximarse mediante la búsqueda de funciones que minimicen la energía del sistema. En la práctica entonces, un tratamiento completo de una molécula poliatómica incluye el cálculo (mediante la ecuación de Schrödinger) de la función de onda electrónica para una dada configuración atómica (una estructura o geometría), que se obtiene empleando el teorema variacional.

Las tres aproximaciones más importantes para calcular las propiedades moleculares son: los métodos *ab initio* (incluyendo el método del funcional de la densidad)², los métodos semiempíricos y el método de mecánica molecular. Los métodos *ab initio* (o de primeros principios) usan el hamiltoniano molecular exacto, es decir no desprecian ni aproximan términos del mismo, y la resolución se logra resolviendo las ecuaciones sin emplear otros datos experimentales que no sean los valores de las constantes físicas fundamentales. Una de las formas de resolución más conocida en esta área se denomina cálculo SCF Hartree-Fock. Sin embargo, no debe interpretarse a los cálculos *ab initio* como “cien por ciento correctos”, ya que en estos muchas veces es también necesario realizar otro tipo de aproximaciones³. Si bien este método es utilizado en algunos sistemas, no es el caso de este trabajo, por lo que esta metodología no se desarrollará con más detalle. Sin embargo el método del funcional de la densidad, los métodos semiempíricos y el método de mecánica molecular serán ampliamente utilizados en esta tesis, por lo que los mismos serán desarrollados en las secciones siguientes.

2.2.2 Teoría del Funcional de la Densidad

Este formalismo, que busca describir un sistema molecular mediante la mecánica cuántica, se ha popularizado en los últimos años debido a su eficiencia computacional (en relación a otros métodos previos, como ser Hartree-Fock) y a la calidad de los resultados obtenidos. El mismo consiste no en obtener la función de onda molecular sino en determinar las propiedades del sistema por medio de la densidad electrónica. De esta forma, la teoría del funcional⁴ de la densidad (DFT, por sus siglas en inglés *density functional theory*) basa su formulación en las coordenadas espaciales y en los valores de espín de los electrones. Actualmente DFT se sitúa como uno de los métodos más populares y versátiles disponibles dentro de los campos de materia condensada, física y química computacional.

La teoría tiene sus raíces teóricas en los teoremas de Hohenberg–Kohn [37] conceptualizados en 1964 para sistemas en ausencia de campos magnéticos y con estados fundamentales no degenerados. El primero de estos teoremas establece que la densidad del estado fundamental determina, salvo por una constante aditiva, el potencial externo, y por tanto todas las propiedades del sistema, incluyendo la función de onda de un sistema de muchos cuerpos. Se dice entonces que la energía del estado fundamental, E_0 , es un funcional de la densidad

²Algunos autores no consideran que este método sea también de primeros principios, o *ab initio*. Sin embargo en este trabajo de tesis se lo enmarca aquí debido a que se considera que el desarrollo teórico es exacto, y que las aproximaciones realizadas en su ejecución no son inherentes a la teoría.

³Como por ejemplo en el caso mencionado de Hartree-Fock, se emplea la aproximación de tomar ψ como un producto antisimetrizado de espín-orbitales de un electrón, donde se emplea además una base finita (y, por tanto, incompleta).

⁴Un funcional es una función que toma funciones como su argumento; es decir, una función cuyo dominio es un conjunto de funciones.

electrónica de este mismo estado, ρ_0 . Se escribe así, por ende, la energía como un funcional de la densidad:

$$(2.6) \quad E_0[\rho_0] = T[\rho_0] + V_{ee}[\rho_0] + V_{ne}[\rho_0]$$

donde los términos corresponden a la energía cinética (T), la interacción electrón-electrón (V_{ee}) y la interacción núcleo-electrón (V_{ne}), que en el contexto de DFT se denomina *potencial externo* ya que se produce por las cargas externas al sistema de electrones.

El segundo de ellos es conocido como teorema variacional de Hohenberg–Kohn, debido a que es análogo al teorema variacional mencionado anteriormente. Esto implica que existe una densidad electrónica $\rho_0(\mathbf{r})$ que minimiza el funcional energía, al igual que la función de onda del estado fundamental minimiza la energía del sistema. Por lo tanto, para toda densidad de prueba $\rho_p(\mathbf{r})$ que satisface $\rho_p(\mathbf{r}) \geq 0 \forall \mathbf{r}$ y que $\int \rho_p(\mathbf{r}) = n$, siendo n el número de electrones totales, se cumple que

$$(2.7) \quad E_0[\rho_p] \geq E_0$$

Los teoremas de Hohenber-Kohn nos dicen que es posible, en principio, calcular todas las propiedades moleculares del estado fundamental a partir de ρ_0 . Sin embargo, no nos dicen *cómo* calcular E_0 a partir de ρ_0 (ya que los funcionales en 2.6 son desconocidos), ni *cómo* obtener ρ_0 sin obtener primeramente la función de onda. Un paso clave hacia estos objetivos lo darían un año después, Walter Kohn y Lu Jeu Sham al desarrollar las ecuaciones que permiten reducir el problema de n electrones interactuantes en un potencial externo, a uno de n electrones no interactuantes en un potencial efectivo, V^{KS} . Dicho potencial es de una forma tal que permite que la densidad de probabilidad electrónica del estado fundamental del sistema no interactuante, ρ_f (en adelante se empleará el subíndice f para denotar dicho sistema ficticio), sea igual a la del estado fundamental exacto del sistema de interés, ρ_0 . Debido a que la función de densidad de probabilidad del estado fundamental determina el potencial externo, una vez que ρ_f está definido por el sistema de referencia, el potencial externo estará también unívocamente determinado. Como los electrones del sistema de referencia no interactúan entre sí, el hamiltoniano de dicho sistema será:

$$(2.8) \quad \hat{H}_f = \sum_{i=1}^n \left[-\frac{1}{2} \nabla_i^2 + V^{KS}(\mathbf{r}_i) \right] \equiv \sum_{i=1}^n \hat{h}_i^{KS}$$

donde $\hat{h}_i^{KS}$ es el hamiltoniano de Kohn-Sham (KS) para un electrón. Debido a que el sistema ficticio se encuentra conformado por partículas no interactuantes, la función de onda del estado fundamental, $\psi_{f,0}$, vendrá dada por el determinante de Slater⁵ de los espín-orbitales

⁵El determinante de Slater es una expresión que se emplea en mecánica cuántica para describir la función

de Kohn-Sham de más baja energía, $\Phi(\mathbf{r}, s)$, esto es

$$(2.9) \quad \psi_{f,0} = \begin{vmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_N \end{vmatrix} \quad \text{donde} \quad \Phi_i(\mathbf{r}_i, s_i) = \varphi_i(\mathbf{r}_i) \chi_i(s_i)$$

$$(2.10) \quad \hat{h}_i^{KS} \varphi_i(\mathbf{r}_i) = \varepsilon_i^{KS} \varphi_i(\mathbf{r}_i)$$

siendo $\chi_i(s_i)$ una función de espín que se introduce manualmente⁶, y $\varphi_i(\mathbf{r}_i)$ es la autofunción del operador unielectrónico $\hat{h}_i^{KS}$ con energía ε_i^{KS} . Con esto podemos ahora expresar la Ecuación 2.6 de manera tal que los términos puedan evaluarse a partir de ρ (donde se omite el subíndice cero para mayor claridad):

$$(2.11) \quad E[\rho] = \bar{T}_f[\rho] + \bar{V}_{ne}[\rho] + \frac{1}{2} \iint \frac{\rho(r_1)\rho(r_2)}{r_{12}} dr_1 dr_2 + E_{xc}[\rho]$$

donde los tres primeros términos incluyen las principales contribuciones a la energía del estado fundamental, correspondientes a la energía cinética electrónica promedio (la sobre- raya denota valores promedio), la energía potencial promedio de atracción nuclear para un electrón y a la energía de repulsión electroestática interelectrónica clásica (en unidades atómicas) para electrones dispersos en una distribución continua de densidad electrónica ρ , respectivamente. El cuarto término, E_{xc} , se denomina funcional de intercambio y correlación, y se define como

$$(2.12) \quad E_{xc}[\rho] \equiv \Delta \bar{T}[\rho] - \Delta \bar{V}_e[\rho]$$

donde $\Delta \bar{T}[\rho]$ es la diferencia en la energía cinética electrónica media del estado fundamental entre el sistema real y el ficticio de electrones no interactuantes, ambos con la misma densidad electrónica (del sistema real), y $\Delta \bar{V}_e[\rho]$ es la diferencia en la energía de repulsión electroestática interelectrónica real y la calculada con la expresión clásica. Evidentemente, esta contribución no es sencilla de calcular con precisión, y si bien es un término relativamente pequeño será la clave para efectuar un cálculo de KS-DFT con precisión.

de onda de un sistema de múltiples fermiones, $\psi(\mathbf{r}, s)$, de manera tal que se cumpla el principio de exclusión de Pauli y la antisimetría requerida para la función de onda de este tipo de sistemas. Para un sistema de N electrones el determinante de Slater se definirá como un determinante normalizado de las funciones espín-orbitales, $\Phi(\mathbf{r}, s)$:

$$\psi(\mathbf{r}_1, s_1; \mathbf{r}_2, s_2; \dots; \mathbf{r}_N, s_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \Phi_1(\mathbf{r}_1, s_1) & \Phi_2(\mathbf{r}_1, s_1) & \cdots & \Phi_N(\mathbf{r}_1, s_1) \\ \Phi_1(\mathbf{r}_2, s_2) & \Phi_2(\mathbf{r}_2, s_2) & \cdots & \Phi_N(\mathbf{r}_2, s_2) \\ \vdots & \vdots & \ddots & \vdots \\ \Phi_1(\mathbf{r}_N, s_N) & \Phi_2(\mathbf{r}_N, s_N) & \cdots & \Phi_N(\mathbf{r}_N, s_N) \end{vmatrix} \equiv \begin{vmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_N \end{vmatrix}$$

⁶En la resolución de la ecuación de Schrödinger la función espín debe ser introducida manualmente, ya que la misma no se obtiene del desarrollo teórico, como sucede con la ecuación de Dirac. Dicha función suele escogerse tal que sea autoestado del operador momento angular de espín en el eje z , $\hat{S}_z$.

Resumiendo, KS DFT es un método práctico que permite obtener la densidad electrónica, y con este la energía, del estado fundamental y es, en principio, exacto. Pero debido a que las ecuaciones de este método contienen un funcional desconocido (de intercambio y correlación) que debe aproximarse, esta formulación conduce a resultados aproximados y su precisión dependerá de la aproximación que se utilice para dicho funcional. En la actualidad existen diversos modelos para aproximar este funcional, siendo algunos de los más utilizados en sistemas bioquímicos BLYP, B3LYP, PBE [3, 50, 60]. En este trabajo de tesis los cálculos de DFT se realizaron con el programa LIO [56], que emplea el funcional PBE.

2.2.3 Métodos Semiempíricos

Debido a las dificultades de aplicar los métodos *ab initio* a moléculas de tamaño mediano y grande, se han desarrollado numerosos métodos mecanocuánticos moleculares simplificados, denominados semiempíricos. Los mismos emplean un hamiltoniano más simple, donde una parte del mismo se encuentra parametrizada para concordar con los datos experimentales o con los resultados de cálculos *ab initio*. En esta tesis se trabajó con el método *Density Functional Tight-Binding* (DFTB) [26], que es una aproximación metodológica basada en la de DFT, por lo que se realizará a continuación una breve descripción del mismo.

DFTB se puede derivar de un desarrollo en serie de Taylor de la energía total de KS-DFT (ecuación 2.10) en torno a una densidad de referencia adecuadamente escogida, $\rho(\mathbf{r})$. En lugar de hallar la densidad electrónica $\rho(\mathbf{r})$ que minimiza la energía, se escoge una densidad de referencia ρ^0 que se perturba con una fluctuación en la densidad,

$$(2.13) \quad \rho(\mathbf{r}) = \rho^0(\mathbf{r}) + \delta\rho(\mathbf{r})$$

El funcional de energía es ahora expandido en una serie de Taylor y dependiendo de la inclusión de los términos de la serie, se considerarán diferentes modelos teóricos, que son, evidentemente, contruidos uno por encima de otro. Si se considera sólo la expansión a primer orden se obtiene el modelo DFTB1 (o simplemente DFTB), incluyendo segundo orden emerge el modelo DFTB2, originalmente conocido como *Self-Consistent Charge, Density Functional Tight-Binding* (SCC-DFTB) y el más reciente incluyendo la expansión de tercer orden, DFTB3 [19, 22, 26].

En la aproximación SCC-DFTB, el nivel teórico empleado en este trabajo, la energía total se describe como [26]:

$$(2.14) \quad E^{SCC-DFTB} = \sum_i \langle \varphi_i | \hat{H}^0 | \varphi_i \rangle + E_{rep} + \frac{1}{2} \sum_{a,b} \Delta q_a \Delta q_b \gamma_{ab}$$

donde los dos primeros términos son computados para la densidad ρ^0 y corresponden al modelo DFTB1 [26]. El primero de ellos es el “término de la estructura de banda” (terminología

derivada de física de sólidos, donde la aproximación DFTB se origina), y se resuelve mediante una combinación lineal de orbitales atómicos (CLOA) [27]. Para esto, los orbitales atómicos ortogonales⁷, ϕ_μ se obtienen de cálculos de KS-DFT para los átomos correspondientes en presencia de un potencial de confinamiento [63]:

$$(2.15) \quad \varphi_i = \sum_{\mu} c_{\mu i} \phi_{\mu}$$

Así, los elementos de matriz del Hamiltoniano se pueden calcular como:

$$(2.16) \quad H_{\mu\nu}^0 = \begin{cases} \varepsilon_{\mu}^{at. libre} & , \quad \text{si } \phi_{\mu} = \phi_{\nu} ; \\ \langle \phi_{\mu} | \hat{H}^0(\rho_a^0 + \rho_b^0) | \phi_{\nu} \rangle & , \quad \phi_{\mu} \in a \text{ y } \phi_{\nu} \in b ; \\ 0 & , \quad \text{en todos los otros casos} \end{cases}$$

donde $\varepsilon_{\mu}^{at. libre}$ es el autovalor de KS para el orbital ϕ_{μ} en el átomo sin confinamiento, y los términos de tres centros han sido descartados. El Hamiltoniano y los elementos matriciales de solapamiento son pre-calculados y tabulados empleando la aproximación de dos centros para distancias interatómicas en una escala relevante, los elementos matriciales a distancias arbitrarias podrán obtenerse mediante la interpolación de los valores tabulados.

El segundo término en la ecuación 2.14 corresponde a la interacción repulsiva de corto alcance entre pares, y es representado como la suma de potenciales de dos cuerpos:

$$(2.17) \quad E_{rep} = \sum_{a,b} U_{ab}$$

En la práctica, este término se ajusta para corresponderse a la diferencia en la energía total entre un cálculo completo de DFT y la parte electrónica del cálculo de DFTB (incluyendo todos los términos, salvo E_{rep}), para una distancia de enlace de interés para un conjunto de moléculas de referencia escogido [25].

La última contribución a la energía total (ecuación 2.14) es la que viene dada por la expansión de Taylor en el término de fluctuación de la densidad, y es la que define la aproximación SCC de DFTB. La misma introduce un término de corrección para las interacciones coulómbicas de largo alcance a la energía total, que se debe a la fluctuación de la carga, $\Delta q_a = q_a - Z_a$, centrada en el átomo (en R_a). Se emplea además una expresión analítica, γ_{ab} , que representa el comportamiento real del sistema: a distancias largas describe las interacciones electrostáticas de largo alcance entre dos cargas puntuales localizadas en R_a y R_b , y además incluye términos de auto-interacción para cuando $R_a = R_b$

⁷La base suele restringirse a los electrones de valencia, en estos casos será necesaria una ortogonalización de la base, lo que se logra mediante una ortogonalización de Schmidt:

$$|\phi_{\mu}\rangle = |\phi_{\mu}\rangle - \sum_{b \neq a} \sum_{\kappa} |\phi_{\mu}^{\kappa}\rangle \langle \phi_{\mu}^{\kappa} | \phi_{\mu} \rangle, \quad \mu \in a$$

Finalmente, aplicando el principio variacional, se obtiene la ecuación de Kohn-Sham como:

$$(2.18) \quad \sum_{\nu} c_{\nu i} (H_{\nu\mu} - \epsilon_i S_{\nu\mu}) = 0 \quad \text{con} \quad \nu \in b \text{ y } \forall a, \mu \in a$$

$$(2.19) \quad \text{donde} \quad H_{\nu\mu} = H_{\nu\mu}^0 + \frac{1}{2} S_{\nu\mu} \sum_c \Delta q_c (\gamma_{ac} + \gamma_{bc})$$

donde $S_{\nu\mu}$ es la matriz de solapamiento. Los elementos matriciales del hamiltoniano dependen de las cargas de Mulliken q_a ($\Delta q_a = q_a - Z_a$), que a su vez dependen de los coeficientes del orbital molecular, $c_{\mu i}$, por ende estas ecuaciones deben ser resueltas de forma autoconsistente.

En este trabajo los cálculos con esta metodología fueron empleados en la mayor parte de los casos de prueba, debido al bajo costo computacional de los mismos. En particular la implementación empleada es la que se encuentra en el paquete de simulación computacional AMBER [22].

En general, los métodos semiempíricos, como es el caso de DFTB, suponen una disminución notoria en el costo computacional total, incluso alcanzando en algunos casos escalas temporales equivalentes a las logradas con modelos clásicos. Sin embargo, no es siempre posible describir un sistema mediante este tipo de metodología, debido a que los fenómenos cuánticos son mucho más complejos que los clásicos, por lo que las parametrizaciones no siempre logran ajustarse a la realidad. Es por esto que aún se siguen empleando métodos *ab initio* dentro de la química computacional, debido a que el inmenso poder predictivo ofrecido por estas metodologías continúa siendo en muchos casos la única alternativa para estudiar un sistema, incluso presentando un costo computacional muy elevado.

2.3 Métodos Clásicos

En química, los métodos cuánticos fueron una extensión natural de la física atómica de principios del siglo XX, avanzando hacia moléculas de mayor tamaño gracias al nacimiento de la tecnología moderna y al desarrollo de métodos como Hartree-Fock (HF), post HF, y DFT. A pesar de que el poder de cómputo se ha incrementado en el último siglo, este tipo de cálculos continua siendo muy demandante en cuanto al costo computacional, permitiendo sólo el estudio de sistemas de cientos de átomos y dinámicas de pocos picosegundos (ver Figura 2.1). Sin embargo, el estudio de sistemas biológicos requiere la inclusión de un gran número de átomos (del orden de miles) y dinámicas de cientos a miles de nanosegundos. Para este tipo de sistemas resulta, aún con el poder computacional existente hoy en día,

prácticamente imposible su tratamiento completo con mecánica cuántica. Es por esto que en gran cantidad de problemas en los que no se requiere un detalle de la distribución electrónica, se utilizan métodos basados en la mecánica clásica, comúnmente denominados de “mecánica molecular” (MM). En estos, se ignora el movimiento de los electrones, y se calcula la energía exclusivamente en función de las posiciones de los núcleos. Sin embargo, la mecánica cuántica se utiliza para el desarrollo de los distintos parámetros requeridos para el cálculo, como ser los enlaces, ángulos y diedros que forma con otros átomos, así como también las interacciones electrostáticas y de Van der Waals. Todas estos parámetros, para diversas combinaciones de átomos, se agrupan en lo que hoy en día se conoce como un *campo de fuerza*. El mismo es, por tanto, una expresión de la energía potencial del sistema, que dependerá en cada caso de las coordenadas de los núcleos y de la identidad (o tipo) de átomo. Evidentemente, los resultados que se obtengan en el estudio de los diferentes sistemas dependerán por completo de la precisión de esta parametrización ⁸. Uno de los campos de fuerzas más utilizados en literatura es el de AMBER [11]. Es también el empleado en esta tesis, por lo que en la siguiente sección se realizará una descripción detallada del mismo.

2.3.1 Campo de Fuerzas de AMBER

La expresión del campo de fuerza AMBER 99SB [38] contiene dos clases de contribuciones: las de unión y las de no unión. Las primeras se calculan para átomos contiguos y hasta terceros vecinos, e incluyen las contribuciones correspondientes a enlaces, ángulos y ángulos diedros. Las contribuciones de no unión, se computan entre átomos que están en diferentes moléculas o a más de 4 enlaces de distancia, y se dividen en interacciones electrostáticas (basadas en la ley de Coulomb) y en interacciones de Van der Waals y de repulsión por el principio de exclusión de Pauli (a distancias muy cortas). La energía potencial de cualquier sistema, dentro del campo de fuerzas de AMBER, se describe como:

$$(2.20) \quad E(r^N) = \sum_{\text{enlaces}} \frac{k_{b,i}}{2} (r_i - r_{0,i})^2 + \sum_{\text{ángulos}} \frac{k_{\theta,i}}{2} (\theta_i - \theta_{0,i})^2 + \sum_{\text{torsiones}} \frac{V_n}{2} \cos(n\varphi - \gamma) + \sum_{i=1}^{N-1} \sum_{j=i+1}^N \left\{ \epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\}$$

donde cada uno de los términos se corresponde con un tipo de interacción entre átomos. El primer término modela un enlace covalente, el cual se describe a través de un potencial armónico, donde k_b se relaciona con la frecuencia de vibración del enlace y r_0 a la distancia de equilibrio de la unión. Si bien una mejor descripción de un enlace químico, que permita la

⁸Dado que la energía cinética vendrá dada por la temperatura del sistema, y, en el caso de los campos de fuerzas, este valor no afecta la energía potencial en la que se mueve el sistema.

representación de lo que ocurre en el sistema fuera de la distancia de equilibrio, se obtendría con un potencial de Morse, éste supone un costo computacional y de parametrización adicional que en la mayoría de los casos no sería empleado, ya que la mayor parte de los enlaces simulados en un sistema bioquímico se mantienen. Por otra parte, el potencial armónico resulta una buena y sencilla aproximación al potencial real en la zona cercana al mínimo, donde no hay ruptura de enlaces. El segundo término corresponde a las contribuciones dadas por las flexiones angulares. Estas también se representan con un potencial armónico de constante k_θ y valor de equilibrio θ_0 . Normalmente, la energía requerida para modificar un ángulo es menor a la requerida para variar la distancia entre dos átomos, por lo que los valores de k_θ suelen ser menores que los de las constantes k_b . El término asociado a las torsiones o ángulos diedros presenta tres parámetros: V_n corresponde a la barrera energética asociada a la rotación entre dos mínimos sucesivos, n representa la multiplicidad del potencial, es decir indica el número de mínimos encontrados entre 0° y 360° , y por último γ , que corresponde a la fase de la función sinusoidal.

El último término de la ecuación 2.20 contiene las contribuciones de no unión, y depende de la distancia entre pares de átomos, r_{ij} y de parámetros específicos asociados a cada tipo de átomo. La primera parte viene dada por el potencial de Lennard-Jones, que representa dos fuerzas opuestas y de distinto origen: una fuerza atractiva de dispersión que actúa a grandes distancias (fuerza de Van Der Waals) y una repulsiva que actúa a distancias cortas representando la sobreposición de los orbitales electrónicos (repulsión de Pauli). Este potencial contiene los parámetros ϵ_{ij} , asociado a la profundidad del pozo de energía, y σ_{ij} , igual a la distancia (finita) en la que el potencial entre los átomos es cero. La última expresión se corresponde con la contribución electrostática clásica, por lo que sus parámetros son las cargas (parciales) sobre los átomos q .

En el campo de fuerzas de AMBER, la distribución de cargas es representada mediante cargas puntuales ubicadas en el centro de las coordenadas de cada átomo. Las mismas se obtienen de modo tal que sean capaces de reproducir el potencial electrostático de la molécula, a partir de cálculos cuánticos. Los términos de no unión suelen arraigar una complicación antes no mencionada: debido a que es usual emplear condiciones periódicas de contorno durante una simulación (para evitar efectos de borde), el costo computacional asociado a este tipo de interacciones se ve notoriamente incrementado. Para reducirlo, se emplea un radio de corte, es decir que, para los átomos que se encuentren a una distancia mayor a este radio de corte, dichas interacciones no se computarán. Sin embargo, al lidiar con las interacciones electrostáticas el radio de corte puede llevar a una acumulación de carga ficticia y por tanto a efectos indeseados. Para evitar esta problemática es recomendado emplear sumas de Ewald, un método originalmente desarrollado para el estudio de cristales iónicos en su implementación denominada Particle Mesh Ewald (PME)[20, 69]. Esta metodología tiene

como ventaja una convergencia mucha más rápida que utilizando cálculos de interacciones tradicionales. De esta manera se logra tener en cuenta las interacciones electrostáticas de largo alcance más allá del radio de corte de las interacciones de no unión, en forma altamente eficiente.

Puede observarse que la mera expresión de la energía potencial en la Ecuación 2.20 no es suficiente para definir un campo de fuerza, sino que se necesitan además una gran cantidad de parámetros asociados a cada una de las interacciones (como ser las constantes de fuerza para los enlaces y ángulos, las cargas parciales, etcétera). Para el caso de simulación de proteínas, el campo de fuerza de AMBER contiene los parámetros asociados a los aminoácidos naturales presentes en la mayoría de las proteínas, en sus distintos estados de oxidación o protonación, cuando corresponda. En este contexto, puede hablarse de la transferibilidad del campo de fuerza, es decir que el mismo conjunto de parámetros puede utilizarse para diversas moléculas, sin necesidad de parametrizar todas las veces todos los residuos de la proteína que se desee estudiar. En el campo de fuerza de AMBER, los átomos dentro de cada residuo reciben un nombre de átomo y un tipo de átomo. Los tipos de átomo permiten que no sea necesario incluir parámetros para cada átomo incluido en cada uno de los aminoácidos. Distintas especies pueden poseer el mismo tipo de átomo, y los parámetros asociados a las interacciones de unión (uniones, ángulos y diedros) se dan por tipo de átomo, en lugar de para cada átomo de cada residuo en particular. Si bien en el campo de fuerza se incluyen parámetros específicos para todos los aminoácidos naturales y otras moléculas comúnmente encontradas en los sistemas biológicos, en algunos casos será necesario obtener parámetros para algún residuo o sustrato que no se encuentre en la base de datos y que sea necesario para realizar la simulación. En este trabajo utilizamos la estrategia desarrollada en el *General AMBER Force Field* (GAFF) [77], habiendo sido necesario sólo parametrizar el sustrato de la reacción modelo del Capítulo 3 y el sitio activo de las Hidrolasas de Zn en el Capítulo 4.

Finalmente, en los sistemas bioquímicos no son sólo los parámetros de residuos proteicos los que determinarán la precisión del cálculo. La forma en la que se representan las moléculas de agua, y las interacciones de las mismas, son también de crucial importancia debido a la gran cantidad de moléculas de agua presentes típicamente en un sistema. Existen en literatura modelos de agua simples, como ser SPC [8], SPC/E [5], TIP3P, TIP4P [42], en donde cada molécula de agua se mantiene en una geometría rígida, y la interacción entre moléculas de agua se describe a través de interacciones Coulómbicas y de Lennard-Jones. Los valores de los parámetros asociados a los modelos de agua simples se obtienen de forma tal de reproducir las propiedades del agua medidas experimentalmente como la densidad, la función de distribución radial, y otras propiedades fisicoquímicas. Cómo se tratan de modelos sencillos, existen gran cantidad de propiedades que no pueden ser descriptas. Existen modelos más complejos, como así también más costosos, que incluyen efectos de

polarización, lo que resulta de importancia en sistemas donde se espera que el solvente experimente un efecto de polarización significativo por parte del resto del sistema. En los cálculos realizados en este trabajo, se utilizó el modelo TIP3P, en el cual se ubican tres cargas puntuales, una sobre cada átomo de la molécula, y la carga negativa ubicada en el átomo de oxígeno se compensa por las cargas positivas en los átomos de hidrógeno. Además, las distancias entre el oxígeno y los átomos de hidrógeno, así como el ángulo H-O-H se mantuvieron fijos (mediante el uso del algoritmo de SHAKE [70]).

2.3.2 Métodos de Grano Grueso

Los sistemas biológicos son increíblemente amplios y diversos, tanto que en muchos casos el interés está puesto en la interacción de escenarios sumamente complejos, como ser el movimiento de dominios dentro de una proteína o la interacción de una proteína con una membrana. Los experimentos no invasivos que se realizan para entender este tipo de procesos se encuentran en escalas temporales del orden de $ms - \mu s$, por lo que el entendimiento de los mismos a nivel atómico experimental es aún inviable. Por otra parte, los modelos teóricos y computacionales vistos hasta el momento, que permiten una descripción de los sistemas biológicos con todos sus átomos, no pueden alcanzar las escalas temporales y espaciales de este tipo de fenómenos (ver Figura 2.1). Por esta razón, en las últimas décadas se han desarrollado campos de fuerzas ⁹ de *grano grueso* (o CG, por sus siglas en inglés, *Coarse Grain*), en los cuales la parametrización no se realiza ya sobre el átomo en sí, sino sobre un conjunto de átomos. Debido a la reducción en los grados de libertad, y a la eliminación de los detalles finos de interacción, las simulaciones de grano grueso requieren menos poder de cómputo que las MM o QM, permitiendo así incrementar las escalas temporales y de tamaño del sistema a simular, en varios ordenes de magnitud.

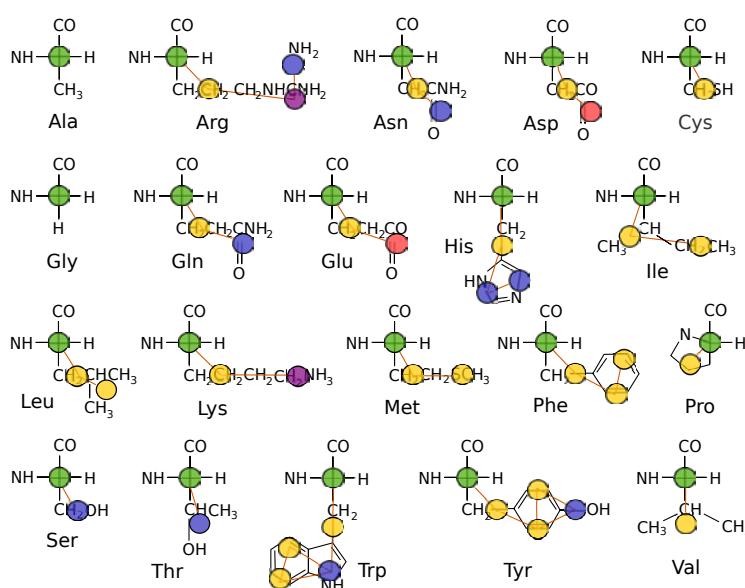
En los últimos años, se ha buscado activamente desarrollar modelos de CG que permitan una representación precisa de la dinámica de proteínas en su estado nativo [21, 36, 54]. Inicialmente, la atención se centró en los procesos de plegado, con el modelo de tipo Go siendo el método de referencia para este tipo de estudios [30], en el mismo, cada aminoácido se sustituye por un centro o “grano” y la dinámica se basa exclusivamente en los contactos nativos. Sin embargo, a menudo es deseable retener información sobre las posiciones de la cadena lateral y las interacciones entre los aminoácidos. Por esta razón se desarrollaron campos de fuerzas específicos para realizar simulaciones CG en proteínas, como la extensión del campo de fuerzas Martini [54], el campo de fuerzas Sirah [21] o el modelo CG para proteínas desarrollado por Hills y colaboradores [36], conocido como modelo de Voth en la

⁹La parametrización de estos modelos se realiza de forma equivalente a la descrita con anterioridad, es decir, buscando reproducir datos experimentales, o de simulación atomística.

literatura. Este último fue el empleado en este trabajo, por lo que a continuación se realizará una descripción más detallada del mismo.

En el modelo de Voth cada aminoácido se representa con una a cinco partículas (o granos), dependiendo de la complejidad de los parámetros del residuo [36] (ver Figura 2.2). Los

Figura 2.2: Representación de grano grueso de los aminoácidos en el esquema de Voth, donde el *backbone* está representado por una sola partícula en el sitio del carbono alfa (verde). Los sitios de las cadenas laterales se asignan según ciertas características en el centro de masa del grupo funcional: polar (azul), apolar (amarillo), positivo (violeta) y negativo (rojo), limitando el tamaño a dos o tres átomos pesados. Los sitios de CG están conectados por enlaces (líneas naranja).



parámetros para este campo de fuerzas se obtuvieron utilizando un enfoque *bottom-up*, donde se buscó ajustar los parámetros de forma tal de reproducir los resultados obtenidos con simulaciones atomísticas. Las interacciones de no unión fueron representadas empleando 5 tipos diferentes de átomos (polar, apolar, positivo, negativo y carbono alfa), donde los 15 pares de potenciales asociados fueron ajustados empleando una aproximación que permitiera igualar las fuerzas. Las interacciones de enlace se parametrizaron para reproducir la distribución de Boltzmann de las simulaciones atomísticas, con la excepción de las torsiones del *backbone*, que fueron tomadas de la base de datos PDB. Este modelo CG fue aplicado con éxito en diversos sistemas, entre estos para el estudio de plegamiento de la *Jaula de Trp* (Trp cage), las transiciones conformacionales de la Adenilato Kinasa [36] y el transporte de MsbA ABC en una bicapa mixta DOPC/DOPE [78]. Considerando las ventajas mencionadas de esta metodología, como ser la posibilidad de obtener los parámetros de simulaciones atomísticas, y la relevancia de las hemoproteínas, en este trabajo se empleó además un modelo de CG del grupo hemo, obtenido mediante un protocolo similar al empleado por Hills *et al.* [65].

2.4 Métodos Híbridos Cuántico-Clásicos

La restricción en cuanto al tamaño del sistema y a las escalas de tiempo accesibles en los métodos cuánticos, y la imposibilidad de representar la dinámica electrónica en los métodos clásicos, representan una barrera hacia el conocimiento, debido a que las reacciones químicas más interesantes ocurren en fases condensadas, como ser solventes acuosos o en el interior de enzimas. Las reacciones en enzimas son particularmente interesantes, debido a que las enzimas no sólo son catalizadoras excepcionales, incrementando la velocidad de ciertas reacciones hasta 10^{15} veces, sino que además son sumamente específicas en cuanto a la selectividad de sustratos y reactivos/productos. Para comprender realmente un mecanismo enzimático, necesitamos describir no sólo cómo se modifica la energía libre con el movimiento de los átomos a lo largo de la reacción (mecanismo de reacción), sino también cómo el entorno proteico disminuye la energía libre del estado de transición (TS), con respecto al mismo mecanismo de reacción en solución. En otras palabras, ¿cómo, estructural y energéticamente hablando, logran las enzimas alcanzar su tasas de eficiencia catalítica?. La respuesta a este interrogante requiere por un lado el tratamiento del sistema reactivo a nivel cuántico, y por otro la inclusión de un gran número de átomos (usualmente miles) de la enzima, y en algunos casos del solvente, para representar adecuadamente el entorno de la reacción. Una forma posible de superar este obstáculo es el modelado multiescala, en particular el método híbrido de mecánica cuántica / mecánica molecular (QM/MM, por sus siglas en inglés *Quantum Mechanics/ Molecular Mechanics*), que fue originalmente propuesto en la década de 1970 por los premios Nobel en química Warshel, Levitt y Karplus [2, 79, 80]. Este tipo de metodología consiste en dividir al sistema en dos: una región QM que abarca el subsistema reactivo (el soluto y/o el sitio activo de la enzima) donde se emplea un nivel cuántico de teoría en la descripción del mismo, y una región MM que incluye todos los átomos del entorno de la región anterior (es decir la enzima y/o el solvente), tratada con potenciales clásicos. Esto constituye un avance significativo respecto de la utilización de sistemas modelo para tratar este tipo de problemáticas, donde solo se incluía el sitio activo y algunos grupos pequeños relevantes. En las metodologías híbridas se considera el efecto de toda la proteína, crucial en el desarrollo del fenómeno reactivo en la mayoría de los casos. Para aplicar una metodología QM/MM es clave no sólo la definición de cada región, sino también el acoplamiento entre ambas, para describir correctamente la electrostática del sistema, así como los efectos estéricos del entorno en el subsistema de reactivo.

Dentro de los métodos QM/MM, pueden distinguirse dos clases de metodologías: los esquemas aditivos y los esquemas sustractivos, siendo el primero el elegido en este trabajo de tesis. En los métodos QM/MM aditivos, el Hamiltoniano consiste en la suma de la contribución del subsistema cuántico (H_{QM}), el subsistema clásico (H_{MM}) y un término de

acoplamiento entre ambas regiones ($H_{QM/MM}$):

$$(2.21) \quad H = H_{QM} + H_{MM} + H_{QM/MM}$$

En este trabajo de tesis, el cálculo de la energía E_{QM} se realizó a nivel DFT [56] y DFTB [22], empleando la interfaz implementada en AMBER [11] en ambos casos. La energía E_{MM} fue calculada con la misma interfaz, empleando el campo de fuerza AMBER99SB [38]. Estas contribuciones a la energía total fueron explicados en las secciones 2.2.2, 2.2.3 y 2.3.1, respectivamente, por lo que a continuación describiremos sólo el término del acoplamiento cuántico-clásico, $H_{QM/MM}$. Este hamiltoniano está compuesto por un término electrostático, que calcula la interacción entre el potencial QM electrostático y las cargas puntuales clásicas, y un término que calcula la interacción estérica entre los átomos de la región QM y la MM, de esta forma:

$$(2.22) \quad E_{QM/MM} = \sum_{i=1}^C q_i \int \frac{\rho(r)}{|r - \tau_i|} dr + \sum_{i=1}^C \sum_{j=1}^Q \frac{q_i Z_j}{|R_j - \tau_i|} + E_{QM/MM}^{LJ}$$

donde la suma en i es para todas las partículas clásicas (C) y en j es para todas las cuánticas (Q), representando q_i a las cargas puntuales de la región MM localizadas en τ_i y Z_j a las cargas de los núcleos cuánticos localizados en R_j , mientras que ρ es la densidad de carga electrónica. Donde el primer término representa la interacción entre la densidad electrónica cuántica y las cargas parciales de los átomos clásicos, y el segundo la interacción entre los núcleos de los átomos cuánticos y las cargas parciales clásicas, de esta forma estos dos términos representan la interacción electrostática entre los dos subsistemas. Es importante notar también que en el H_{QM} , el potencial ejercido sobre el subsistema cuántico por las cargas clásicas (puntuales) es sumado de forma auto-consistente al potencial de los núcleos, de esta forma la estructura electrónica obtenida se corresponde con el potencial híbrido QM/MM calculado. El tercer término representa la interacción de Van Der Waals y repulsiva a corto alcance entre los átomos de ambos sistemas y se calcula usando un potencial de Lennard-Jones, idéntico al usado en el campo de fuerzas clásico. Esto implica que es necesario conocer los parámetros asociados al potencial de Lennard-Jones ϵ y σ para los átomos del subsistema QM, los cuales se obtienen, usualmente, del campo de fuerza utilizado para representar el subsistema MM.

Aunque los métodos QM/MM permiten una descripción de la dinámica de sistemas moleculares grandes y el estudio de las reacciones químicas, en las simulaciones simples a temperatura ambiente, el sistema no es capaz de cruzar barreras energéticas moderadamente altas ($>kT$), tales como las presentes en reacciones enzimáticas, quedando así atrapadas en el estado inicial (reactivo). Por lo tanto, para simular la dinámica de reacciones químicas es necesario guiar al sistema a lo largo de la reacción, como se verá en la siguiente sección.

2.5 Explorando la Superficie de Energía de un Sistema

La superficie de energía potencial (SEP) es una función escalar que define el valor de la energía potencial del sistema en función de las coordenadas de los núcleos, es decir, para cada una de las posibles configuraciones del sistema. El conocimiento de esta superficie puede brindarnos información cinética y termodinámica de un sistema, como por ejemplo los mínimos de la misma nos permitirán encontrar las configuraciones estables del sistema, mientras que los puntos silla se corresponderán con los estados inestables. En la SEP una reacción se verá representada por lo que se denomina un camino de reacción, que consiste en la secuencia de valores que minimizan la energía al pasar de un mínimo a otro, atravesando una barrera de energía.

Para sistemas de gran tamaño, como son las proteínas, la SEP resulta muy compleja y es imposible conocerla en su totalidad. En esta sección se describirán distintos métodos desarrollados para explorarla de forma eficiente, usualmente concentrándonos en la región de la misma que se corresponde con el fenómeno de interés. Si bien existen diversas metodologías para realizar esta tarea, se desarrollarán solamente aquellas aplicadas en los cálculos realizados en este trabajo.

2.5.1 Dinámica Molecular

En sistemas biológicos, la complejidad de la superficie de energía potencial involucra la existencia de gran cantidad de mínimos locales. Esto implica que no existe un solo mínimo representativo de cada sistema, sino que existen una serie de conformaciones, que estarán presentes en mayor o menor proporción según sus energías relativas (de acuerdo a la ley de Boltzmann).

Cuando se realiza una medida experimental, todas las conformaciones relevantes estarán representadas en el ensamble de moléculas presente en el sistema. Una manera de representar este ensamble es a través de la reconstrucción de la evolución temporal de la dinámica del sistema, asumiendo que el sistema ha evolucionado un tiempo suficiente para que haya podido explorar las distintas conformaciones disponibles (convergencia de la simulación). La hipótesis ergódica [48] postula que el promedio temporal puede reemplazar el promedio sobre las configuraciones de un sistema dado, dando origen así a las simulaciones basadas en evolución temporal, diferenciándose de aquellas que buscan explorar el colectivo de configuraciones espaciales, como el método de Montecarlo.

En una simulación de dinámica molecular (DM), se realiza la reconstrucción de la evolución temporal de la dinámica de un sistema a través de la resolución de las ecuaciones de

movimiento de Newton,

$$(2.23) \quad a_{\chi_i}(t) = \frac{d^2 \chi_i}{dt^2} = \frac{F_{\chi_i}}{m_i}$$

donde se describe el movimiento de una partícula de masa m_i en la coordenada χ_i , con la fuerza en esa dirección, F_{χ_i} . La fuerza sobre cada una de las partículas cambia cuando la partícula modifica su posición, y es función del movimiento de todas las otras partículas. La resolución analítica del problema no es posible, por lo que debe recurrirse a métodos numéricos. El método implementado en el programa AMBER se denomina algoritmo Velocity-Verlet [33] que fue desarrollado para resolver las ecuaciones de movimiento de Newton de forma eficiente permitiendo calcular las velocidades (necesarias para determinar la energía cinética y, por lo tanto, la temperatura) y posición de los átomos en el mismo paso de integración, Δt . En el mismo las posiciones de los átomos del sistema en un tiempo $t + \Delta t$ se calculan según

$$(2.24) \quad \vec{x}(t + \Delta t) = \vec{x}(t) + \vec{v}(t) \Delta t + \frac{1}{2} \vec{a}(t) \Delta t^2$$

Con las nuevas coordenadas de los átomos es posible calcular las fuerzas resultantes que actúan sobre éstos (obtenidas con el nivel de teoría escogido), y de esa forma calcular $\vec{a}(t + \Delta t)$, con lo que se obtiene

$$(2.25) \quad \vec{v}(t + \Delta t) = \vec{v}(t) + \frac{\vec{a}(t) + \vec{a}(t + \Delta t)}{2} \Delta t$$

De esta forma, la posición de los átomos y su velocidad sólo dependen de la aceleración $\vec{a}$, es decir de la fuerza resultante sobre cada átomo. El sistema se propaga iterando las ecuaciones 2.24 y 2.25 por un número de pasos fijos.

Debe prestarse especial atención a la elección del intervalo de tiempo (de integración) Δt . Cuanto menor sea éste, mayor será la cantidad de pasos necesarios para realizar una simulación de una dada cantidad de tiempo, con lo que incrementará el costo computacional. Sin embargo, un Δt demasiado grande dará origen a una integración incorrecta, haciendo que el resultado diverga. Por otra parte, no sólo debe escogerse un valor de Δt que permita una correcta integración del algoritmo, sino que también es necesario tener en cuenta la naturaleza del sistema, de manera que el intervalo de tiempo escogido permita describir el movimiento de las partículas en la molécula (electrones, átomos o grupos de átomos, según el nivel de teoría empleado). En la práctica se utiliza un intervalo de tiempo que sea aproximadamente un décimo del movimiento característico más rápido del sistema. En el caso de un sistema de átomos simulado de manera clásica, los movimientos más rápidos están asociados a las vibraciones de los enlaces, y las más rápidas están asociadas a los enlaces en los que uno de los átomos involucrados en la unión es un átomo de hidrógeno.

Una estrategia para aumentar el Δt y así acelerar los cálculos consiste en utilizar el algoritmo SHAKE[67], que permite fijar estas uniones en sus posiciones de equilibrio.

Normalmente, es deseable realizar las simulaciones de DM a temperatura constante. De acuerdo al principio de equipartición de la energía, la temperatura está relacionada con la energía cinética (K) de la siguiente manera:

$$(2.26) \quad K = \sum_{i=1}^N \frac{m_i \vec{v}_i^2}{2} = \frac{3k_B T}{2} N$$

donde m_i y v_i corresponden a la velocidad y la masa de la partícula i , y N es el número de partículas del sistema. Es de esperar entonces que para mantener la temperatura constante en una simulación sea necesario escalar las velocidades de forma tal que 2.26 se ajuste a la temperatura deseada.

Uno de los termostatos más empleados en la literatura, y el utilizado también en este trabajo, es el termostado de Langevin. En este, todas las partículas del sistema reciben una fuerza aleatoria y reducen su velocidad mediante una constante de fricción en todos los pasos. Ambas fuerzas están relacionadas de manera tal que la magnitud promedio garantice el teorema de “fluctuación-disipación” del ensamble canónico (NVT, número de partículas, volumen y temperatura constantes). En este formalismo, se añade entonces una fuerza, $F_{\chi_i}^{LT}$ sobre el átomo i en la coordenada χ , con dos componentes, que actúa como termostato:

$$(2.27) \quad F_{\chi_i}^{LT} = -\gamma v_{\chi_i} + \gamma f_{\chi_i}(T)$$

donde f_{χ_i} es una función aleatoria que depende de la temperatura externa y γ es la frecuencia de colisión, que se ajusta como un parámetro externamente. Cuanto menor sea este parámetro, más débil será el acoplamiento, por lo que la fluctuación en la temperatura será mayor, un factor de colisión de $1,0 \text{ ps}^{-1}$ otorga buenos resultados para los sistemas similares a los estudiados en esta tesis.

Por otra parte, es también común realizar simulaciones de DM en condiciones de presión constante, por lo que el sistema debe variar el volumen de la celda en la que se realiza la simulación. Uno de los esquemas más empleados es el de Berendsen [6] en el cual se propone acoplar el sistema a un *baño* equilibrado a una cierta presión, que actúa de forma equivalente a como lo hace un baño térmico al regular la temperatura de un sistema. El volumen de la celda se escala entonces en un factor γ , lo que equivale a escalar las coordenadas atómicas en $\gamma^{1/3}$. Resulta entonces:

$$(2.28) \quad \gamma = 1 + \kappa \frac{\delta t}{\tau_P} [P - P_b]$$

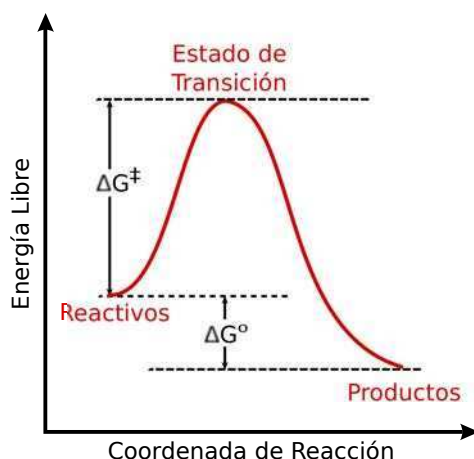
donde $\kappa = -\frac{1}{V} \left(\frac{\partial V}{\partial P} \right)_T$ es la compresibilidad isotérmica, y τ_P corresponde al factor de acoplamiento del baño al sistema.

Luego de realizar una simulación de DM, el resultado principal que se obtiene es la trayectoria de cada uno de los átomos que conforman el sistema a lo largo del tiempo simulado. En teoría, diferentes procesos biológicos que involucren diferentes conformaciones del mismo sistema podrían ocurrir durante una simulación, sin embargo el tiempo de simulación necesario para que esto ocurra está fuera del alcance de la tecnología actual. Debido a que la mayor parte de los fenómenos de interés en sistemas biológicos no puede ser estudiado en los tiempos de dinámica molecular tradicionales, se emplean esquemas de Dinámica Molecular Guiada (DMG). En los mismos se emplea una fuerza externa que provoca un cambio en la estructura del sistema, de manera que, a lo largo de la dinámica, los átomos se ajustan al cambio forzado en la estructura y diferentes conformaciones pueden ser muestreadas a lo largo de un camino particular. A medida que la fuerza actúa y el movimiento ocurre, la energética del sistema puede ser calculada de forma tal que el proceso de interés pueda ser estudiado termodinámicamente.

2.5.2 El Perfil de Energía Libre

Para sistemas complejos es necesaria más información que la SEP para predecir si una reacción ocurrirá, para esto se emplea el concepto de Energía Libre. Como se mencionó en el

Figura 2.3: Esquema que muestra la variación de Energía Libre de una reacción elemental como función de la Coordenada de Reacción.



Capítulo 1, ésta es una función de estado, por lo que el valor neto de la misma no es relevante, sino que lo es la diferencia de energía libre entre estados, $\Delta G = G_{final} - G_{inicial}$. Esta diferencia nos dará información acerca de la direccionalidad de una reacción, asumiendo valores negativos cuando la reacción es espontánea en el sentido en el que ha sido planteada, y valores positivos en caso contrario. De esta forma, existe para cualquier reacción una hipersuperficie de energía libre asociada, que contiene toda la información, cinética y termodinámica, del sistema. Como ocurre con la SEP, esta hipersuperficie es extremadamente compleja, y en la mayoría de los casos bastará con obtener el perfil de energía libre asociado con la reacción

de interés. En el mismo se grafica la energía libre como función de los diferentes valores que asume la *coordenada de reacción* (CR) a lo largo del mecanismo de reacción (ver Figura 2.3). Donde definiremos una CR como una combinación de parámetros estructurales, usualmente distancias y combinaciones de distancias (o ángulos) entre átomos, que proveen una descripción del progreso de la reacción, pudiendo relacionarse a un mecanismo de reacción completo, o a una etapa del mismo.

Es usual que los fenómenos biológicos de interés se encuentren caracterizados por estados diferentes (reactivos/productos), los cuales se encuentran separados por una barrera de energía libre. En una simulación estándar, es poco probable que el sistema cruce esta barrera de modo espontáneo, salvo que la misma sea muy pequeña. Por esta razón es necesario, en la mayoría de los casos, conducir al sistema a lo largo de un camino escogido que adopte los valores de la CR necesarios para representar el fenómeno (como por ejemplo una reacción o un cambio conformacional) en su totalidad.

Un muestreo apropiado de este proceso permitirá la obtención del Perfil de Energía Libre correspondiente (FEP, por sus siglas en inglés *Free Energy profile*, o en algunos casos también conocido como *Potencial of Mean Force*, PMF), el cual describirá los cambios de energía libre a lo largo de la reacción, proporcionando la barrera correspondiente ($\Delta G^\ddagger$) y la energía libre de reacción (ΔG°), pudiéndose comparar ambos valores con los datos experimentales de forma directa. Más aún, la obtención de estas cantidades permite además analizar diferentes propuestas para el mecanismo de reacción mediante la comparación de los FEP obtenidos con coordenadas de reacción diferentes.

Es por tanto clave en la obtención de FEPs precisos, y por lo tanto significativos, muestrear adecuadamente el espacio configuracional a lo largo de la CR. Existen para esto diversos métodos, el primero y uno de los más comunmente utilizados es conocido con el nombre de *Umbrella Sampling* [46], en el mismo, un potencial armónico es incluido en el hamiltoniano del sistema. La particularidad del método reside en que la posición de equilibrio de dicho potencial es modificada a lo largo de la CR en “ventanas” discretas, hasta muestrear el mecanismo completo. El FEP resultante se obtiene empleando una distribución de Boltzmann para incorporar peso estadístico a las configuraciones que fueron muestreadas. Existen en la actualidad diversas metodologías que permiten calcular la energía libre de una transición. Entre estas se encuentra el trabajo excepcional de Christopher Jarzynski [40], el mismo será empleado en este trabajo de tesis, por lo que será descrito en detalle en la próxima sección.

2.5.3 Igualdad de Jarzynski

La Igualdad de Jarzynski (IJ) es una ecuación de mecánica estadística que relaciona la diferencia de Energía Libre entre dos estados de equilibrio, y el trabajo irreversible de un

ensamble de trayectorias que unen los dos estados [40]. En termodinámica, la diferencia de energía libre, $\Delta G = G_B - G_A$ entre dos estados A y B se relaciona con el trabajo W realizado sobre el sistema mediante la siguiente ecuación:

$$(2.29) \quad \Delta G \leq W$$

donde la igualdad es válida solo en el caso de procesos reversibles o *cuasi estáticos*, es decir, cuando el sistema es conducido del estado A al estado B infinitamente lento, de manera que no exista calor disipado.

Hasta 1997, esta era la única relación existente entre ambas cantidades, y sólo podría alcanzarse la igualdad en procesos infinitamente lentos. Sin embargo, la Igualdad de Jarzynski, válida independientemente de la velocidad del proceso $A \rightarrow B$, establece que

$$(2.30) \quad \langle e^{-\beta W_{irr}} \rangle = e^{-\beta \Delta G}$$

$$\Delta G = -\frac{1}{\beta} \ln \langle e^{-\beta W_{irr}} \rangle$$

donde, ΔG es la variación de energía libre entre dos puntos, $\beta = (k_b T)^{-1}$ siendo k_b la constante de Boltzmann y T la temperatura del sistema, y $\langle e^{(\dots)} \rangle$ representa el promedio exponencial de los trabajos irreversibles individuales (W_i) punto a punto a lo largo de la coordenada de reacción r . En otras palabras, la IJ muestra que, para cualquier sistema que parte de un estado inicial de equilibrio y es conducido externamente a un estado final, también de equilibrio, a cualquier velocidad, la energía libre entre ellos puede obtenerse si el trabajo realizado durante este proceso es exponencialmente promediado para un número elevado de trayectorias individuales [17, 18].

En la práctica, la IJ puede implementarse para obtener un FEP a través de simulaciones de Dinámica Molecular Guiada. Usualmente, la fuerza externa aplicada con la que se guía al sistema se escoge como un potencial armónico, cuyo mínimo se mueve a lo largo de la CR a una determinada velocidad, acorde a la ecuación 2.31.

$$(2.31) \quad \vec{F}_{ext}(t) = -k[\vec{r} - (\vec{r}_o + \vec{v}t)]$$

De esta manera la fuerza aplicada guía al sistema de reactivos a productos a lo largo de la coordenada de reacción, al mismo tiempo que nos permite obtener el trabajo realizado sobre el sistema (indispensable para el cálculo de energía libre):

$$W = - \int \vec{F}_{ext} d\vec{r}$$

Para poder obtener el perfil de energía libre, se realizan N_t trayectorias guiadas individuales, partiendo de un conjunto diferente de configuraciones iniciales¹⁰ en equilibrio, y para cada

¹⁰Las configuraciones iniciales pueden pertenecer al reactivo o bien al producto. Es usual, al realizar una DMG en combinación con la IJ, realizar el camino de reacción de reactivos a productos, y viceversa, para obtener un mejor muestreo configuracional.

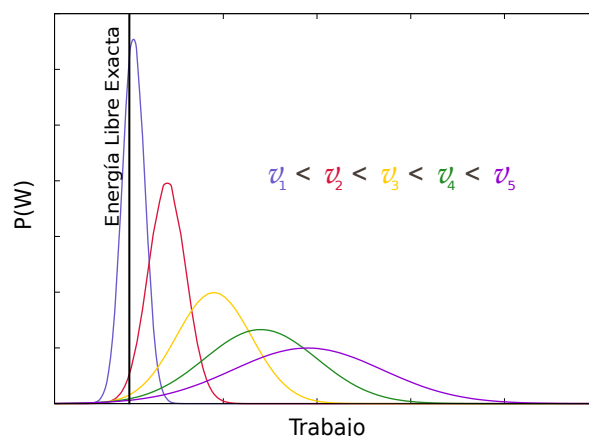
una de ellas se computa el perfil de trabajo a lo largo de la CR. Finalmente, los múltiples trabajos son exponencialmente promediados empleando la ecuación 2.30 para obtener el perfil de energía libre correspondiente. Esta metodología, que usualmente se conoce como Múltiples Dinámicas Moleculares Guiadas (MDMG o MSMD, por sus siglas en inglés) fue inicialmente empleado por Shulten y colaboradores [59] y comparado con el método de Umbrella Sampling, demostrando que se podía obtener igual precisión con una reducción del costo computacional. Desde su derivación original, la igualdad Jarzynski ha sido aplicada a una inmensa variedad de sistemas, y ha recibido validación computacional y experimental [14, 15, 52, 68, 83], y ha sido la base para nuevas derivaciones teóricas [17, 18], lo que refuerza la confianza en su generalidad.

Existen algunas consideraciones a tener en cuenta cuando se emplea MDMG con la IJ para obtener perfiles de energía libre. El primer factor importante es que el promedio exponencial en la Ecuación 2.30 debe estar convergido para que la ecuación valga, lo que es formalmente cierto sólo cuando se han realizado un número infinito de trayectorias. Evidentemente esto no es posible y en la práctica se lleva a cabo un muestreo finito (que se define por el número de trayectorias independientes realizadas, N_{Tr}) y por lo tanto la energía libre es un estimador estadístico. Sin embargo, una de las grandes ventajas de este método radica en que, debido a cómo se calcula el FEP, la Ecuación 2.29 continúa siendo válida y por tanto cualquier estimador que se obtenga cumplirá siempre ser menor o igual a la energía libre real (característica que no pueden asegurar otras metodologías, como sucede con el método de *Umbrella Sampling*). El otro aspecto clave de la IJ es que, si bien no hay ningún requisito en cuanto a la velocidad en la cual el sistema es llevado de reactivos a productos, dicha velocidad tendrá un efecto notorio en la rapidez con la que se obtenga la convergencia del perfil y en la precisión del mismo. Teniendo en cuenta la forma escogida del potencial externo (Ecuación 2.31), la velocidad, o *pulling speed*¹¹, con la cual el mínimo del potencial recorre la coordenada de reacción, definirá el número de pasos de dinámica molecular, y por ende también el costo computacional, necesarios para llevar a cabo la reacción.

Para comprender mejor cómo se relacionan la velocidad y el número de trayectorias con la convergencia del promedio exponencial y con la energía libre, es útil considerar algunos casos extremos. Supongamos primero que la distribución de trabajos que se obtiene es aproximadamente gaussiana (este requerimiento no es necesario, pero facilita la explicación), y supongamos que realizamos el mismo experimento de MDMG para diferentes velocidades,

¹¹Esta velocidad no está relacionada con la velocidad de la reacción, se relaciona simplemente con la distancia recorrida sobre la CR y el número de pasos de DMG en los que se realizó la dinámica, es por esto que se suele emplear el término inglés *pulling speed* para denominar la velocidad de la ecuación 2.31 y diferenciarla así de la de reacción. En este trabajo emplearemos la expresión *velocidad* para referirnos a lo que en literatura se conoce como *pulling speed*, y haremos explícita la mención cuando se trata de considerar la velocidad de la reacción.

Figura 2.4: Representación de la distribución de probabilidad para los trabajos colectados a diferentes velocidades.



como se muestra en la Figura 2.4. En condiciones cercanas a las de equilibrio, es decir empleando velocidades cercanas a cero, la distribución de trabajos se vuelve muy angosta y el promedio de los trabajos externos se iguala a la energía libre, en concordancia con la segunda Ley de la Termodinámica. Sin embargo, si la velocidad aumenta, la distribución de trabajos se ensancha, haciendo que el trabajo promedio asuma valores mayores. Si bien la obtención de la energía libre por medio de la IJ otorga más peso estadístico a los valores de trabajo menores (por tratarse de un promedio exponencial), mientras más rápida sea la velocidad de la DMG menos probabilidades habrá de obtener una trayectoria que contenga efectivamente el valor de energía libre correcto. Por ende, para obtener un muestreo correcto del proceso será necesario realizar DMG sobre un mayor número de trayectorias (N_{Tr}). Es así que cuando hablamos de promedio convergido, nos referimos al muestreo completo de la distribución de trabajos, sólo de esta forma podremos garantizar la igualdad en la Ecuación 2.30. Sin embargo, debido a que la distribución de trabajos no es conocida, es necesario estimar apropiadamente los valores de velocidad y N_{Tr} que realizaremos: en un extremo tendríamos entonces una velocidad infinitamente lenta para la cual sólo requeriríamos una trayectoria, mientras que en el extremo opuesto realizaríamos infinitas trayectorias a una velocidad muy alta.

Otro aspecto importante cuando se realizan MDMG es que, para cualquier velocidad, la distribución de trabajos es más amplia conforme nos alejamos del punto de partida, es decir que a medida que la CR avanza, nos movemos fuera del equilibrio y por ende el trabajo aumenta. Por esta razón es usual realizar dos tipos simulaciones para la misma reacción: una en el sentido en el que la reacción ocurre (*forward*), y otra en sentido inverso (*backward*). Esto se hace debido a que los valores de energía libre más precisos se encontrarán cercanos a las conformaciones iniciales, de esta forma, combinando los dos resultados obtenidos, podrán obtenerse perfiles de energía libre precisos a lo largo de toda la coordenada de reacción[4, 18].

Una estrategia diferente para subsanar las distribuciones amplias de trabajo, originadas

en mover la CR fuera del equilibrio, fue realizada durante este trabajo de tesis y se presentará en el capítulo 3.

HyDRA: UN ESQUEMA PARA OPTIMIZAR LOS CÁLCULOS DE ENERGÍA LIBRE HÍBRIDOS

En el capítulo 2 se introdujo la problemática del cálculo preciso de perfiles de energía libre, en tiempos computacionales accesibles, para diferentes procesos biológicos. En este capítulo se presentará el Esquema de Relajación Diferencial Híbrido (HyDRA, por sus siglas en inglés para *Hybrid Differential Relaxation Algorithm*) desarrollado durante este trabajo de tesis que permite obtener perfiles de energía libres precisos de modo más eficiente. El esquema desarrollado permite además reducir el tiempo de simulación computacional necesario para obtener un FEP preciso en los casos en los que el proceso biológico de interés incluye reactividad química, y el sistema es estudiado con metodologías híbridas cuántico-clásicas.

3.1 El Esquema de Relajación Híbrido Diferencial

En la última década ha habido un intenso debate acerca de si la estrategia de MDMG, en combinación con la IJ, otorga alguna ventaja, en términos de costo computacional, frente a otros métodos disponibles. Si bien es sabido que es posible obtener FEP precisos de esta forma, en muchos casos esta estrategia es descartada. Una de las causas más comunes es la imposibilidad de realizar la reacción en sentido inverso (*backward*), sabiendo que, por el error acumulado al final de la DMG, la velocidad de la misma deberá ser muy lenta (ver sección 2.5.3), incrementando así el número de pasos en los que se lleva a cabo la reacción y finalmente produciendo también un aumento en el costo computacional. Este efecto suele agravarse cuando además el fenómeno de interés es una reacción enzimática (en las que

es usual emplear dinámicas híbridas QM/MM) donde el costo computacional asociado al subsistema reactivo puede incluso hacer inviable la obtención de un perfil de energía libre. El trabajo presentado en esta sección nace como una alternativa a esta problemática. Se basa en el hecho de que, en los perfiles de energía libre de reacción de entornos complejos (solución, enzimas) obtenidos mediante MDMG y la ecuación de Jarzynski, el error acumulado en el trabajo al final de la reacción se debe mayoritariamente a la relajación inadecuada del entorno, que produce una acumulación de trabajo no disipado y que, por lo tanto, genera una sobreestimación de la energía libre. Es decir, que el exceso de trabajo acumulado al final de la DMG podría reducirse efectuando una relajación más adecuada del entorno, y no necesariamente disminuyendo la velocidad (o *pulling speed*) actuante en la fuerza externa. Por lo tanto es posible obtener FEPs precisos a un menor costo computacional, permitiendo un muestreo más eficiente del entorno, para lo que se propone un esquema de relajación diferencial dentro de la metodología QM/MM.

3.1.1 Fundamentos Teóricos

En el año 2002 Woo *et al.* introdujeron un esquema de tiempos múltiples [81] para explorar el espacio de las fases en sistemas de gran tamaño, empleando Dinámica Molecular *ab initio* QM/MM de Car-Parrinello. El mismo busca optimizar el muestreo al reducir la masa del subsistema con mayor variabilidad conformacional, aludiendo al hecho de que el promedio configuracional no depende de las masas de los núcleos. Esto permite muestrear las configuraciones de ambas regiones diferencialmente. Sin embargo, al reducir las masas, es necesario emplear intervalos del tiempo de integración (dt) menores para poder evaluar apropiadamente las ecuaciones de movimiento de un sistema ahora más liviano, y por lo tanto más *veloz*. Emplear el mismo dt para ambas regiones (QM y MM) implicaría realizar pasos de dinámica cuánticos innecesarios. Debido a que los cálculos cuánticos son los más costosos computacionalmente, los autores buscaron reducir este incremento empleando escalas temporales diferenciadas por región, dt mayores en las regiones de mayor masa (subsistema QM) y dt menores para las regiones de masas reducida (subsistema MM). Este esquema, con valores de masa y pasos temporales diferenciados por región, permitió obtener un mayor muestreo del sistema global, sin que esto implicara un costo computacional elevado. Si bien demostraron que el esquema de tiempos múltiples híbridos podía acelerar los procesos de equilibración y de muestreo conformacional en el espacio de las fases, no era posible hacer inferencias acerca de que ocurriría con la evolución de las trayectorias en el espacio real, y por tanto no se podrían calcular propiedades que surgieran de la misma. Por esta razón los autores no fueron capaces de determinar un perfil de energía libre con el esquema mencionado.

Más recientemente, y en relación con el cálculo de perfiles de energía libre obtenidos mediante la IJ, Ozer y colaboradores [57, 58] desarrollaron un esquema que permite mejorar la convergencia del FEP obtenido a través de DMG. En el mismo, conocido como Dinámica Molecular Guiada Adaptativa (o ASMD por sus siglas en inglés, *Adaptive Steered Molecular Dynamics*), la coordenada de reacción predeterminada se divide en segmentos, comúnmente referidos como etapas (no relacionadas con los pasos de un mecanismo de reacción), dentro de los cuales se realiza DMG y se calcula el estimador de Jarzynski a lo largo de cada etapa. Entre etapas, se selecciona una única trayectoria, con el criterio de que el perfil de trabajo obtenido durante la misma sea el más cercano al estimador de Jarzynski. Las coordenadas del final de la etapa para la trayectoria seleccionada se utilizan para inicializar la siguiente etapa de trayectorias de DMG. Los autores demuestran que, eliminando las trayectorias que se alejan del valor significativo de energía libre, se puede disminuir el número de trayectorias que deben realizarse, sin necesidad de realizar aproximaciones de ningún tipo. Este enfoque teórico, de gran relevancia en la aplicación de esta técnica, fue aplicado para estudiar los procesos de plegado proteico, únicamente desde la mecánica molecular. Si bien estas teorías sentaron un marco de referencia, hasta el momento no se habían realizado desarrollos semejantes que permitieran realizar cálculos de energía libre de reacciones químicas en entornos complejos.

3.1.2 Implementación

Tomando como base los esquemas desarrollados por Woo *et al.* y Ozer *et al.*, se buscó combinar ambos enfoques: Idear un esquema de tiempos múltiples que permitiera un muestreo más amplio de la región clásica, y que además empleará la configuración del *paso* anterior para calcular la configuración actual. Con este fin se desarrolló un esquema de relajación diferencial que permite llevar a cabo una relajación del sistema MM por un número fijo de pasos, mientras la región QM permanece estática (o congelada), para luego realizar un paso conjunto, es decir, mover ambos sistemas en simultáneo. El cociente entre el número de pasos puramente clásicos y los pasos conjuntos (o QM/MM) recibe el nombre de DRAr, por sus siglas en inglés para *Differential Relaxation Algorithm ratio*, es decir *cociente del Algoritmo de Relajación Diferencial*. Bajo esta nomenclatura nos referiremos a la dinámica QM/MM convencional como dinámica con DRAr de 1. En la Figura 3.1 se esquematiza el accionar del algoritmo para un DRAr de 4, lo cual implica que se realiza un paso de dinámica cuántica cada cuatro pasos clásicos (o un paso de dinámica QM/MM y tres puramente MM).

En este tipo de esquemas multiescala es indispensable la correcta evaluación de las fuerzas, ya que la misma diferirá cuando se realiza un paso conjunto o un paso de relajación MM. Las fuerzas que actúan sobre el sistema QM, en una dada conformación, tienen contri-

buciones de los átomos en la región cuántica, así como de los átomos en la región clásica. Notaremos esta componente de la fuerza total como $F_{QM}(QM, MM)$. De manera análoga, en la región clásica actúan fuerzas que son contribución de la región QM y de la MM, es decir, $F_{MM}(QM, MM)$. Para ejemplificar el algoritmo emplearemos el esquema de la Figura 3.1.

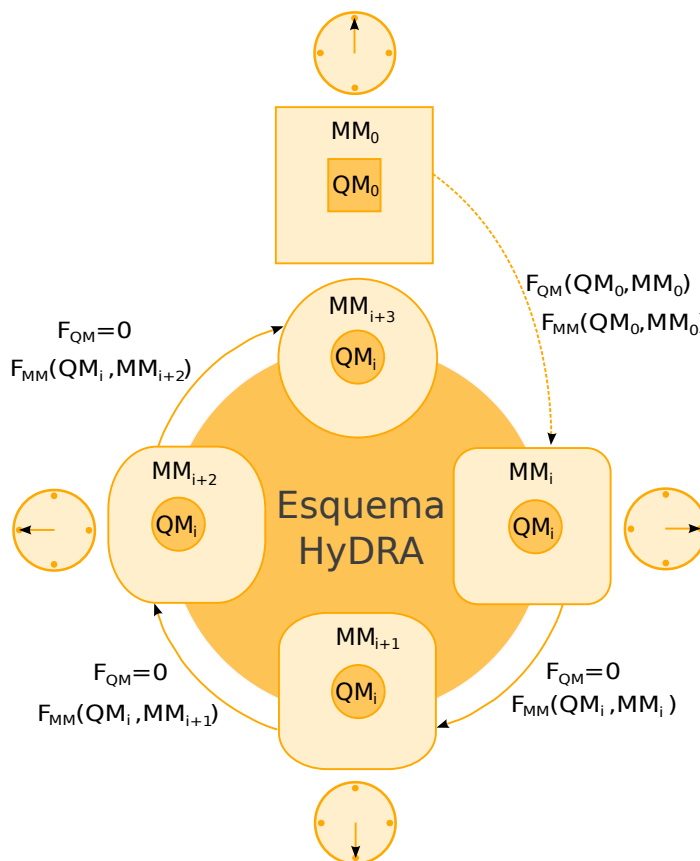


Figura 3.1: Esquema del Algoritmo de Relajación Híbrida Diferencial (*Hybrid Differential Relaxation Algorithm*) para un cociente entre el número de pasos puramente clásicos y los QM/MM (DRAr) de 4.

Al inicio de la dinámica, los sistemas cuántico y clásico se encuentran sincronizados en la conformación inicial (subíndice 0, arriba en la figura). Para mover el sistema en un paso de tiempo (dt) se considerarán todas las contribuciones a la fuerza en ambos sistemas, como en un paso de dinámica QM/MM convencional. Es decir que calcularemos las fuerzas $F_{QM}(QM_0, MM_0)$ y $F_{MM}(QM_0, MM_0)$, y con estas realizaremos el primer paso de dinámica, alcanzando la conformación i (a la derecha en la figura). Ahora debemos comenzar con los pasos de relajación puramente clásicos. Para ello las fuerzas sobre el sistema clásico son computadas como es usual, pero considerando las coordenadas fijas en la conformación i del sistema QM. Por ende, podríamos considerar que el sistema MM se mueve en un campo de fuerzas QM fijo para llegar al estado $i + 1$. Durante este paso las fuerzas sobre el sistema cuántico, $F_{QM}(QM, MM)$, se hacen iguales a cero, para que este subsistema permanezca *congelado*. Una vez que el sistema MM ha evolucionado al estado $i + 1$ (mientras el sistema QM permanece fijo en la configuración i), se realizarán otros dos pasos puramente clásicos, arribando finalmente a la conformación MM_{i+3} (Figura circular superior en el

esquema). Luego de que el sistema clásico a evolucionado por DRAr pasos, el ciclo volvería a comenzar, efectuando un paso conjunto. Las fuerzas que actuarán para este paso serán aquellas que surjan de considerar la estructura $i + 3$ para el sistema clásico, y la i para el sistema cuántico. De esta manera, cada estructura (o conformación), obtenida mediante este algoritmo, luego de un paso conjunto será equivalente a una obtenida en una DM QM/MM convencional, mientras que las estructuras obtenidas en los pasos puramente clásicos representan conformación que se obtendrían al permitir que el sistema MM se relaje alrededor de un área cuántica inmóvil.

Debido a que el algoritmo implementado es una aproximación teórica, es necesario discutir los alcances de la misma, así como también los casos en los que resultará benéfica su aplicación. En el desarrollo original de Jarzynski, se establece que sólo es necesario calcular el trabajo realizado sobre el sistema por la fuerza externa para obtener la energía libre, mientras que el calor disipado no se considera. Por ende, sólo las instancias en las cuales la fuerza externa es aplicada sobre la coordenada elegida deberían ser consideradas en el cálculo de los trabajos, necesario para obtener la energía libre. Es decir que, si la coordenada de reacción permanece fija mientras el entorno se mueve, no se habrá realizado ningún trabajo sobre el sistema de interés¹, sólo habremos disipado calor, por ende estos pasos no perjudicarían el cálculo de energía libre mediante la ecuación de Jarzynski. Esto ha sido demostrado teóricamente en el trabajo presentado por Ozer *et al.* en 2010. Esto implica que si el esquema de HyDRA fuera aplicado sólo sobre los átomos que conforman la coordenada de reacción, el resultado sería exacto y no se estaría realizando ninguna aproximación. Pero en la realidad, el subsistema cuántico no incluye solo a los átomos implicados en la coordenada de reacción, haciendo que este tipo de enfoques siga siendo costoso. Una aproximación posible es despreciar los grados de libertad que aporta el subsistema cuántico, debido a que los que aporta el subsistema clásico son mucho más elevados, fijando las coordenadas de todos los átomos de la región QM, y no sólo aquellos involucrados en la coordenada de reacción. Es de vital importancia, entonces, que el esquema de HyDRA sea aplicado a sistemas en los cuales la región QM elegida sea significativamente menor que el entorno MM simulado. La validez de esta aproximación se demostrará empíricamente y se discutirá como parte de los resultados de esta tesis.

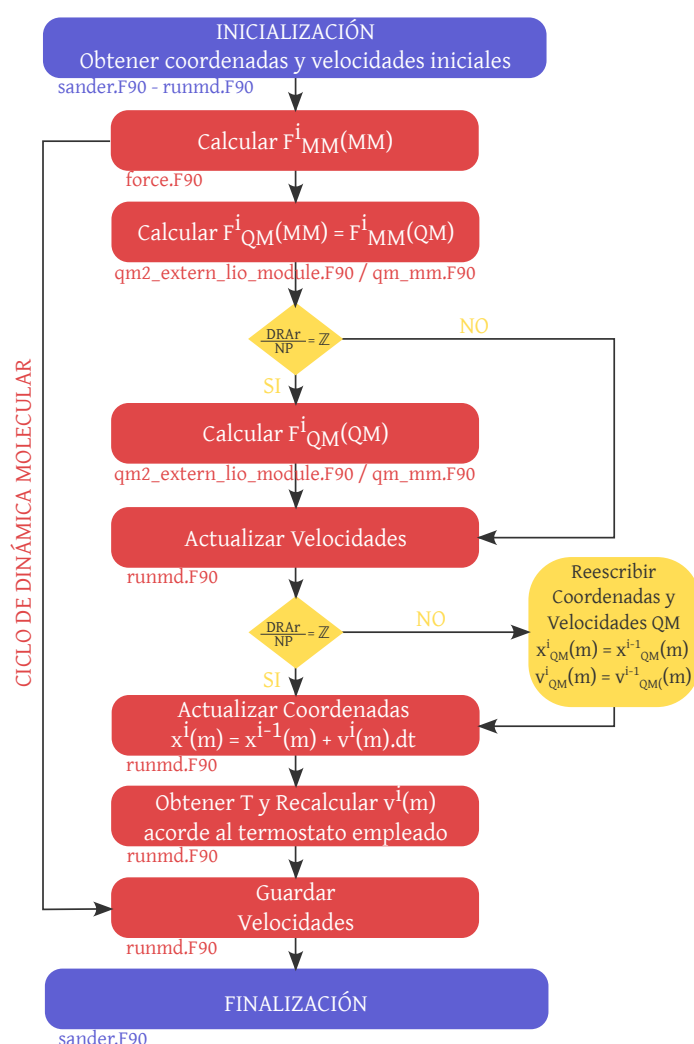
3.1.3 HyDRA en AMBER

El esquema HyDRA fue implementado originalmente dentro del módulo SANDER del programa AMBER, versión 12 [11]. Actualmente el esquema será lanzado dentro del paquete de programas de simulación computacional molecular de AmberTools versión 17 (*release*

¹Nos referimos al trabajo computado que permite el cálculo de la energía libre de la reacción estudiada.

Abril de 2017), conteniendo los manuales y tutoriales correspondientes para que se pueda utilizar en la comunidad científica. Dentro del código se realizaron diversas modificaciones, esquematizadas en la Figura 3.2, para permitir que las fuerzas cuánticas sólo se calculen cada DRAr pasos. Esto se logra empleando el siguiente criterio de selección: si el resultado del cociente entre el valor de DRAr y el número de paso actual (que varía desde 1 al número de pasos total de la dinámica molecular, que se asigna en el archivo de entrada con el nombre de NTS LIM) es un número entero, entonces se calcula la contribución cuántica, mientras que en el caso contrario, no.

Figura 3.2: Diagrama de flujo que representa el esquema general de realización de una dinámica molecular dentro del paquete AMBER y el módulo SANDER. Los pasos de inicialización y finalización (en azul) han sido sobresimplificados para mayor claridad de forma que el ciclo de DM (en rojo) pueda ser descripto con más detalle. En amarillo se incluyen las modificaciones implementadas para poder realizar el esquema HyDRA. Debajo de cada recuadro se nota además la subrutina en la que se encuentra cada uno de los pasos descriptos, donde se ha puesto más detalle en los pasos que generan la dinámica. Para la inicialización y finalización sólo se incluyen las que contienen los llamados a subrutinas particulares. En el caso de las fuerzas sobre el subsistema cuántico, y las de interacción QM/MM, como estas dependen del nivel de teoría elegido (DFTB o DFT en este caso) se incluyen las dos subrutinas en las que se realizaron modificaciones.



Se han realizado además modificaciones que no han sido representadas en la Figura 3.2, éstas afectan la interfaz con el usuario y le permiten al mismo activar el esquema HyDRA desde el archivo de entrada. En la Figura 3.3 se ejemplifican dos archivos de entrada, el de la izquierda que representa una dinámica tradicional híbrida, con nivel de teoría cuántico DFTB, y a la derecha se muestra la modificación que es necesario realizar para aplicar el

esquema HyDRA, donde han sido remarcadas las diferencias. Una de ellas es incluir, dentro de la sección "&qmmm" la variable `drar`, que activa el esquema siempre que el valor de ésta sea mayor a uno, y que guarda el valor de DRAr en una variable interna.

1ps de DM convencional	1ps de DM con HyDRA (DRAr=10)
<pre>&cntrl nstlim=1000, dt = 0.001, [...] ifqnt=1 / &qmmm qmmask=":1", qm_theory="DFTB"</pre>	<pre>&cntrl nstlim=10000, dt = 0.001, [...] ifqnt=1 / &qmmm qmmask=":1", qm_theory="DFTB", drar=10</pre>

Figura 3.3: Ejemplos de archivos de entrada para el programa SANDER dentro del paquete de simulación AMBER. En ambos casos se muestra un extracto del archivo necesario para realizar una dinámica molecular híbrida cuántico-clásica (`ifqnt=1`) de 1 ps con un intervalo de tiempo de 1fs (`dt=0.001`). En el caso de la derecha se emplea el esquema HyDRA con un DRAr de 10, por lo que el número de pasos (`nstlim`) aumenta en 10 para alcanzar 1ps de DM QM/MM.

Además de esto, es necesario modificar el valor de "`nstlim`" para que coincida con el número de pasos **clásicos** que queremos realizar. Recordemos que durante una DM con HyDRA tendremos diferentes números de pasos para el subsistema clásico y el cuántico, donde el primero será DRAr veces mayor que el segundo. El número de pasos cuánticos, igual al cociente entre `nstlim` y `drar`, es de suma importancia, ya que determinará la velocidad de la reacción cuando se realiza una DMG. En el ejemplo del archivo de entrada presentado en la Figura 3.2 a la derecha, el número de pasos cuánticos será igual a 1000, ya que `nstlim=10000` y `drar=10`.

Siendo que el esquema de HyDRA surge como una solución a un problema práctico de la aplicación computacional de la ecuación de Jarzynski, sólo podrá ser verificado empíricamente. La implementación de dicho esquema en el paquete de simulación computacional AMBER nos permitió realizar diversos experimentos computacionales con el fin de corroborar la hipótesis planteada.

3.2 Validación del Esquema Propuesto

Debido a que la aproximación que realizamos con HyDRA no puede ser confirmada teóricamente, resulta indispensable tomar un caso como modelo y contrastar los resultados

obtenidos con HyDRA contra valores más precisos que pueden ser tomados como de referencia. En este contexto, la reacción de Corismato a Prefenato resulta útil ya que se cuenta con numerosos estudios teóricos [15, 16, 49, 76] y experimentales [1, 43, 44] con los que contrastar los resultados.

3.2.1 La Reacción Modelo: Corismato a Prefenato

La reacción de Corismato a Prefenato está involucrada en la biosíntesis de aminoácidos aromáticos en bacterias, hongos y plantas, es por esto que ha sido objeto de extensos estudios experimentales y teóricos. La enzima Corismato Mutasa (CM) cataliza una transposición de Claisen que, según la literatura, ocurre con idéntico mecanismo de reacción en la enzima y en solución. En el año 1997 Kast y colaboradores [44] reportaron los valores de activación experimental para la reacción catalizada por la enzima: $\Delta G^\ddagger = 15,4$ kcal/mol, $\Delta H^\ddagger = 12,7$ kcal/mol y $\Delta S^\ddagger = -9,1$ kcal/mol. El grupo en el que realicé mi trabajo de tesis se encontró entre los primeros en implementar la técnica de MDMG junto con la ecuación de Jarzynski en simulaciones QM/MM para estudiar una reacción enzimática: la conversión de Corismato a Prefenato en la enzima *Bacillus subtilis* Corismato Mutasa (BsCM) en 2005 [15]. Además de este, otros grupos - como el de Mulholland [13] en 2004 y el de Lee Woodcock [49, 82] en 2003 y 2007- han reportado también estudios teóricos de los parámetros de activación para la CM, realizados con diferentes niveles de teoría y estrategias de simulación.

El estudio de la reacción que se presenta a continuación fue realizado tanto en solución como en la enzima BsCM. La coordenada de reacción elegida fue una combinación de distancias, que involucra la distancia entre los carbonos que pasarán a formar un enlace, d_{CC} , y la del enlace carbono-oxígeno que se rompe, d_{CO} , como se muestra en la Figura 3.4. Se ha demostrado en trabajos previos [13, 15, 16, 43, 44] que esta CR puede reproducir apropiadamente el mecanismo de reacción. En nuestro caso, la misma fue variada entre los valores $-3,3$ y $+2,0$ Å, valores que permiten una descripción completa de la reacción.

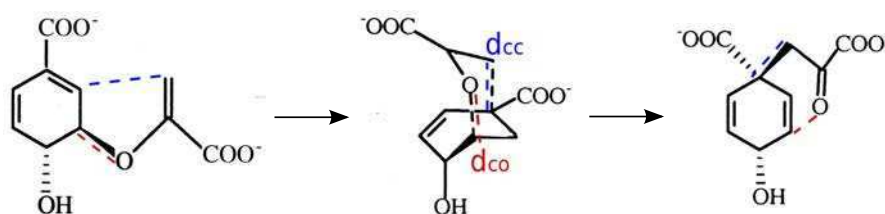


Figura 3.4: Esquema que describe la reacción de Corismato a Prefenato, en solución y en la enzima BsCM. La coordena de reacción empleada consiste en una resta entre las distancias entre dos carbonos, d_{CC} en azul, y la distancia carbono-oxígeno, d_{CO} mostrada en rojo.

3.2.1.1 Condiciones de Simulación

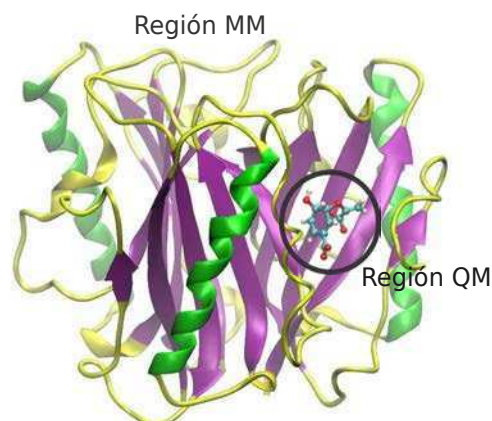
Para todas las simulaciones clásicas se empleó el campo de fuerzas ff99SB [51] para la descripción de los 20 aminoácidos del sistema biológico y las moléculas de agua fueron representadas con el modelo TIP3P [42]. Los parámetros clásicos del sistema reactivo (Corismato y Prefenato) fueron tomados de trabajos anteriores del grupo [15, 16]. Estos mismos parámetros se emplearon para la parte clásica de las simulaciones híbridas QM/MM. Para la reacción se empleó un esquema híbrido QM/MM, tratando el sistema cuántico con la teoría de *Density Functional Tight Binding* (DFTB) [19, 22]. Las simulaciones se realizaron empleando el termostato de Berendsen, tal como se encuentra implementado en el módulo de SANDER, usando condiciones periódicas de contorno, y volumen o presión constantes. Se emplearon, además, sumas de Ewald para tratar las interacciones electrostáticas de largo alcance.

Corismato en solución. El sistema consta de una molécula de sustrato (corismato) solvatada con una caja octaédrica de alrededor de 1500 moléculas de agua. En primer lugar, se realizó una optimización de la geometría completa usando el algoritmo predeterminado de AMBER seguido de una simulación de dinámica molecular clásica para equilibrar la densidad del sistema a la temperatura deseada (300K). Esto se logró aplicando primero una rampa de temperatura hasta los 300K durante $1ns$ a volumen constante, seguido de $1ns$ a presión constante, empleando un paso temporal de $1fs$.

Estudios previos demuestran que para que la reacción ocurra es necesario que la molécula de corismato se encuentre en la conformación diaxial, donde la distancia d_{CC} asume un valor cercano a $4,7 \text{ \AA}$. Por esta razón se realizó primeramente una DM híbrida QM/MM de $200ps$ con un $dt = 0,5fs$, considerando al sustrato como cuántico, de la cual se seleccionó una estructura diaxial para realizar la corrida de producción (restringida a dicha conformación) de $2ns$. A continuación se seleccionaron 20 estructuras de partida diferentes para realizar las MDMG, necesarias para calcular la energía libre con el estimador de Jarzynski.

Corismato en BsCM. Para estudiar la reacción de corismato a prefenato en BsCM, se escogió el cristal correspondiente a la enzima *wild type* unida a prefenato (PDB id 1COM [12]) utilizado en el estudio anterior de esta reacción [15, 16]. Los átomos de hidrógeno fueron añadidos empleando el módulo *tleap* del paquete AMBER, y estados estándar de protonación fueron escogidos para los residuos titulables (Aspártico y Glutámico cargados negativamente, y Lisina y Arginina negativamente). Mientras que la protonación de la histidina fue escogida de manera tal que favoreciera los enlaces hidrógeno en la estructura cristalina. Uno de los prefenatos presentes en el trímero de BsCM, fue reemplazado por una molécula de corismato, para poder llevar a cabo la reacción.

Figura 3.5: Representación de cintas de la enzima BsCM empleada para estudiar la reacción de Corismato a Prefenato. En la misma se ignoran las aguas y el sustrato se encuentra en la representación de pelotas y palos, para mayor claridad. De los tres sustratos posibles, se escogió sólo uno para ser tratado cuánticamente, mientras que los otros se representan de forma clásica.



El complejo proteico resultante fue solvatado con una caja octaédrica de alrededor de 15000 moléculas de agua TIP3P. En la Figura 3.5 se encuentra representado, con pelotas y palos, el sustrato escogido para ser tratado cuánticamente, llevando a cabo la reacción. El sistema se optimizó primero durante 2000 pasos utilizando un algoritmo de gradiente conjugado, seguido de una DM clásica de 200ps a volumen constante donde la temperatura del sistema se elevó lentamente hasta 300 K y de una DM de 200ps a presión constante. Durante las dinámicas, los carbonos alfa de la proteína fueron restringidos con una constante de fuerza de $1\text{kcal/mol}\text{\AA}^2$. Una vez equilibrado el sistema se realizó una dinámica molecular clásica de 5ns sin restricciones. El procedimiento posterior es similar al realizado para la reacción en solución, con la excepción de que, debido a las restricciones geométricas naturales de la enzima, no fue necesario fijar la molécula de corismato en la posición diaxial (debido a que esta es la única conformación posible para la molécula dentro de la enzima).

3.2.2 Perfiles de Energía Libre de Referencia

Si bien existen en literatura numerosos estudios teóricos que caracterizan esta reacción, los mismos fueron realizados empleando diferentes niveles de teoría. Debido a que sería incorrecto comparar valores numéricos para diferentes niveles de teoría, fue necesario determinar perfiles de energía libre de referencia propios, de manera tal que se pudieran emplear para validar el esquema propuesto. La metodología que se empleará en la presente sección, para validar el esquema de HyDRA, es la de QM(DFTB)/MM tal como se encuentra implementada en el modulo de SANDER. Comenzamos nuestro análisis determinando un perfil de referencia para la conversión de Corismato a Prefenato en solución. Se realizaron tres SMD a una velocidad de $0,02\text{\AA}/\text{ps}$, considerando que este valor se aproxima considerablemente a las condiciones de reversibilidad, dado que las curvas de trabajo difieren en menos de 1kcal/mol . Esto se puede observar en la Figura 3.6 (en azul para la reacción en

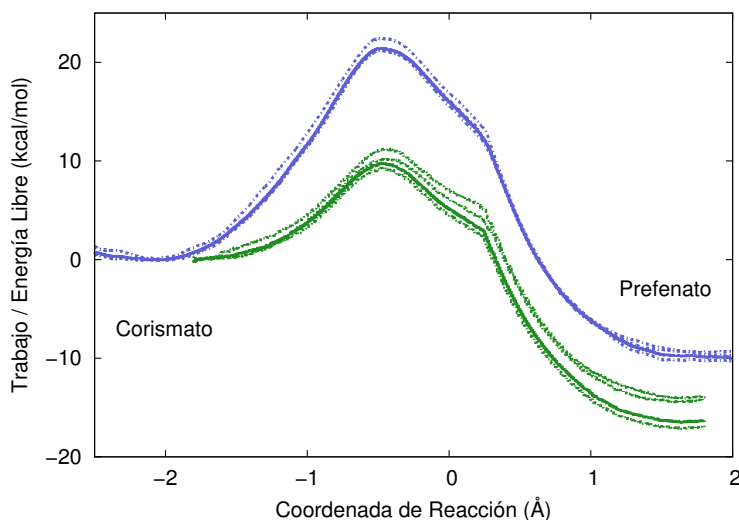


Figura 3.6: Trabajo y FEP para la conversión de Corismato a Prefenato en solución, en azul, y en la enzima BsCM, en verde. Las curvas de línea punteada corresponden al trabajo obtenido al conducir al sistema a lo largo de la coordenada de reacción, con una velocidad de $0,02\text{\AA}/ps$. Mientras que el FEP obtenido mediante la IJ se encuentra representado en línea continua.

solución), donde las curvas de línea punteada representan los perfiles de trabajo y la curva de línea sólida el estimador de Jarzynski para la energía libre correspondiente. Esta última curva será considerada en adelante como el perfil de energía libre de referencia para esta reacción, con un ΔG^0 de -10kcal/mol y una barrera de activación, $\Delta G^\ddagger$, de 21kcal/mol .

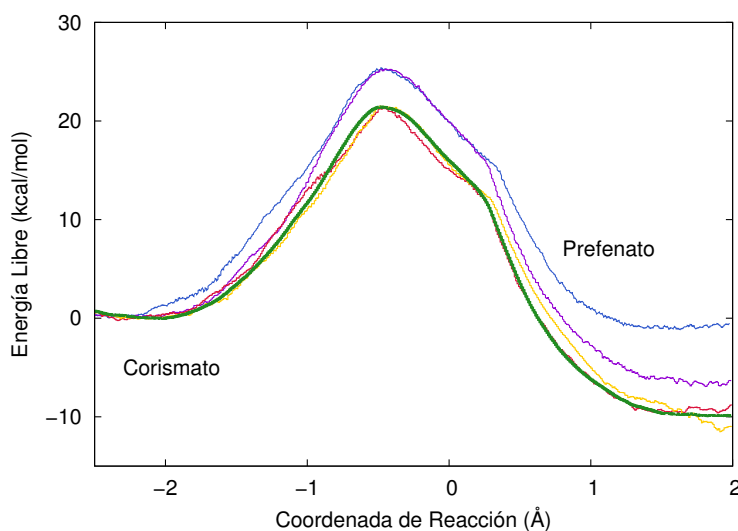
Bajo este mismo criterio se calculó un perfil de referencia para la reacción en la enzima BsCM (en línea continua verde en la Figura 3.6). Los valores obtenidos fueron $9,6\text{kcal/mol}$ para la barrera de activación y un ΔG^0 de $-16,5\text{kcal/mol}$. Estos valores, tanto para la reacción en la enzima como los obtenidos en solución, correlacionan bien con los valores experimentales en la literatura, como se discutirá en la sección 3.4.

3.2.3 Aplicación de HyDRA en Solución

Para poder estudiar la influencia de HyDRA en los perfiles de energía libre, se analizó la precisión de los mismos para la reacción de corismato a prefenato, empleando múltiples dinámicas moleculares guiadas de no equilibrio en conjunción con la ecuación de Jarzynski, utilizando diferentes valores de DRAr (1, 4, 10 y 20). Usualmente, y dependiendo del tipo de reacción, la obtención de un perfil de energía libre con esta metodología (MDMG con la ecuación de Jarzynski, sin HyDRA) involucra la realización de entre 5 y 40 trayectorias independientes, que se realizan a una velocidad entre 10 y 500 veces más rápida que si se realizara en condiciones cercanas al equilibrio. Estos parámetros, velocidad y número de trayectorias empleadas, son variables que influyen en la precisión y calidad del perfil de energía libre, y en el costo computacional asociado. A continuación se detallará el análisis realizado para estudiar esta dependencia para la reacción de corismato a prefenato, y la influencia del valor de DRAr en estos cálculos.

En un primer lugar se considerarán los perfiles de energía libre obtenidos empleando 10 trayectorias de no equilibrio a una velocidad de $1\text{\AA}/ps$ (50 veces más rápido que el FEP de referencia) y DRAr 1, 4, 10 y 20. Analicemos primero el perfil obtenido con dinámica conven-

Figura 3.7: Perfiles de energía libre para la reacción de Corismato a Prefenato en solución, obtenidos mediante 10 dinámicas de no equilibrio, con una velocidad de $1\text{\AA}/ps$ y diferentes valores de DRAr: 1, 4, 10 y 20 -en azul, lila, rojo y amarillo respectivamente-; junto con el perfil de referencia en verde.



cional, es decir $\text{DRAr}=1$, representado por la curva azul en la Figura 3.7, y comparémoslo con el perfil de referencia obtenido (en color verde en la misma figura). Se puede observar que para el caso de mayor velocidad se obtiene un perfil de energía libre menos preciso, en el que tanto $\Delta G^\ddagger$ como ΔG° son significativamente sobreestimados. Si bien se produce un incremento en la energía de la barrera de activación (de 5 kcal/mol), el error más grosero se produce en el ΔG° de la reacción que, si sólo se tuviera en cuenta este resultado, desconocería el perfil de referencia (como ocurre en la práctica), prediría que reactivos y productos tienen la misma energía libre, lo que llevaría a concluir que en esta reacción la termodinámica no tiene ninguna injerencia.

En el caso del perfil obtenido empleando un DRAr de 4 (curva de color lila en la Figura 3.7), podemos observar que también ocurre una sobreestimación en los valores de energía libre, sin embargo en este caso la termodinámica de la reacción se prediría correctamente, ya que se observa que el prefenato se encuentra en un estado de menor energía libre que el corismato. Empleando el esquema de HyDRA con valores de DRAr de 10 y 20 (perfiles en rojo y amarillo, respectivamente, en la Figura 3.7) los resultados obtenidos para esta reacción no se distinguen de aquel obtenido con una velocidad 50 veces menor, por lo que pueden considerarse correctos. Y a su vez, los valores de energía libre obtenidos en estos últimos casos con el esquema de HyDRA muestran muy poca variación entre sí. Esto se debe a que el esquema de HyDRA acelera la relajación del solvente, evitando que el sistema reactivo realice trabajo en exceso sobre el solvente, sin embargo, una vez que el mismo relajó, incrementar el número de pasos clásicos no produce ningún efecto observable sobre la energía libre. En otras

palabras, existe un límite para la relajación del subsistema clásico frente a la perturbación del sistema cuántico.

Sin embargo, el valor neto de energía libre no es la única medida a tener en cuenta. Para llegar al perfil correcto empleando la igualdad de Jarzynski, es indispensable que el ensamble se encuentre convergido. Para poder contar con una primer medida que nos permitiera estudiar la convergencia del FEP en relación al valor de DRAr, se analizaron las curvas de trabajo mediante las cuales se obtuvieron los perfiles de energía libre representados en la Figura 3.7. Estos resultados pueden observarse en la Figura 3.8, donde se muestran los perfiles

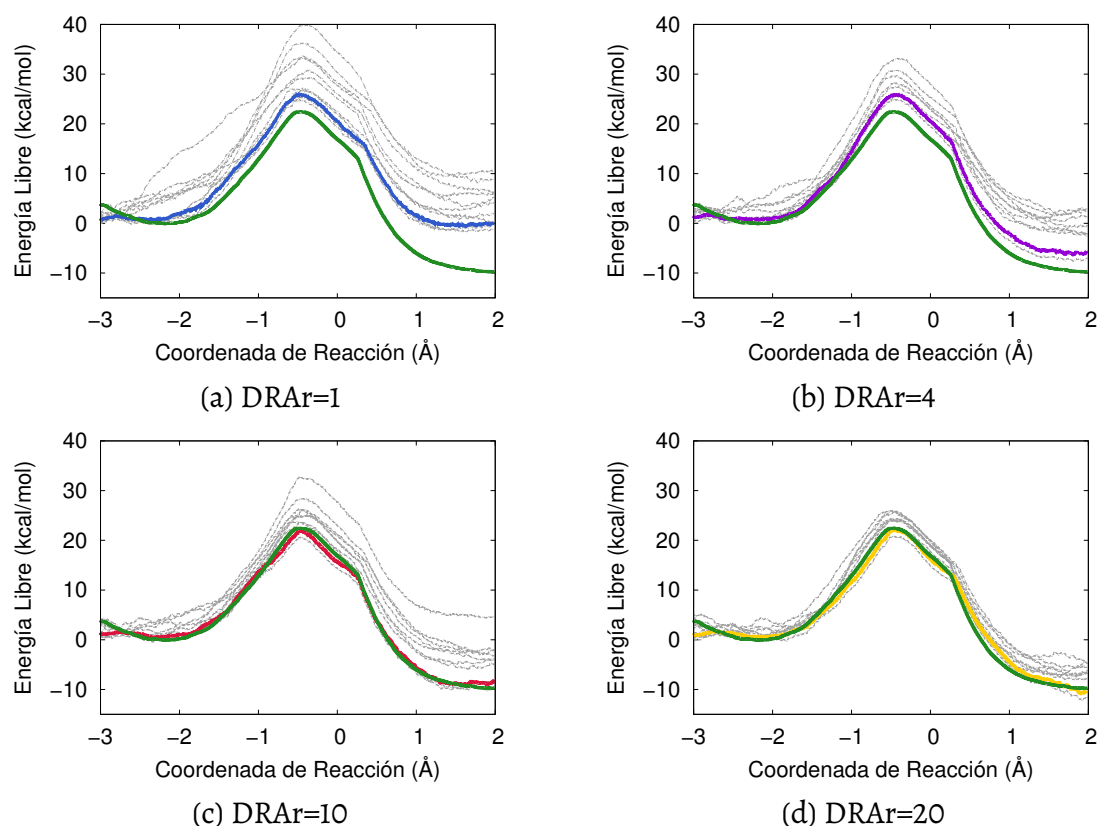


Figura 3.8: Perfiles de energía libre obtenidos para diferentes valores de DRAr mediante la igualdad de Jarzynski, empleando valores de trabajo obtenidos para 10 trayectorias (curvas color gris) y una velocidad $1\text{\AA}/ps$. Se muestra además el perfil de referencia en color verde, mientras que los FEP obtenidos con HyDRA se representan empleando los colores azul, lila, rojo y amarillo para los valores de DRAr de 1, 4, 10 y 20 respectivamente.

de energía libre obtenidos en cada caso siguiendo el código de colores de la Figura 3.7 (donde DRAr 1, 4, 10 y 20 se representan con los colores azul, lila, rojo y amarillo respectivamente), junto con las curvas de trabajo empleadas en la obtención de cada perfil (en color gris). A fin de comparar con mayor claridad los perfiles, en todos los casos se muestra el FEP de referencia en verde, y todos los gráficos se encuentran en la misma escala. Puede observarse

que, en el caso de la dinámica convencional (Figura 3.8.a), la dispersión entre las curvas de trabajo es de aproximadamente 20kcal/mol , mientras que para las DMG con valores de DRAr de 4 y 10 la dispersión disminuye tomando valores entre $10-15\text{kcal/mol}$. Sin embargo el caso realmente notorio se encuentra para DRAr=20 donde, si bien el perfil de energía libre no varía respecto a DRAr=10 (demostrando que en el segundo caso el muestreo es suficiente como para obtener la energía libre correcta), la dispersión alcanza un máximo de 7kcal/mol , alcanzando un nivel de precisión muy elevado. De esta forma puede observarse como la convergencia del perfil mejora al aumentar el valor de DRAr, disminuyendo notoriamente la dispersión entre las curvas de trabajo.

Como se mencionó en la sección 2.5.3 (Figura 2.4), la convergencia del perfil se encuentra directamente relacionada con la velocidad y el número de trayectorias utilizadas, ya que son estos los valores que determinarán la calidad del muestreo estadístico. Por esta razón se estudió en detalle la relación que existe entre estas variables y el valor de DRAr, resultados que se presentan a continuación.

3.2.3.1 Influencia de la velocidad

Para ser capaces de estudiar la influencia del esquema HyDRA en la obtención de un FEP y su relación con la velocidad, se realizaron diversos perfiles manteniendo fijo el número de trayectorias empleadas en un valor de 10, variando ahora la velocidad para que la misma fuera entre 25 y 100 veces más rápida que la requerida para considerar la reacción reversible. Los resultados obtenidos se presentan en la Tabla 3.1, donde lo primero que puede observarse es que, para cualquier velocidad, un aumento en el valor de DRAr conduce no sólo a menores valores de energía libre - tanto en la barrera como en el ΔG° de la reacción - sino que se obtienen también resultados más exactos (en comparación con los valores de referencia). Es importante además notar que la contribución del esquema de HyDRA resulta más notoria a medida que la velocidad de la reacción se incrementa. Por ejemplo, para una velocidad 25 veces mayor que la de referencia (de $0,5\text{\AA}/ps$) la diferencia entre los resultados es apenas significativa entre DRAr 1 y 10. Sin embargo, cuando la velocidad es 100 veces mayor que la de referencia ($2\text{\AA}/ps$) las MDMG arrojan un resultado no sólo cinéticamente incorrecto, prediciendo una barrera de activación casi 50% por encima del valor esperado, sino que ni siquiera logra la termodinámica correcta, prediciendo un producto más estable que el reactivo (contrario a la realidad). En cambio, empleando el esquema de HyDRA con un DRAr de 20 los resultados son equivalentes a los obtenidos con una velocidad 4 veces menor. Esto implica que es posible obtener, empleando un DRAr correcto, un perfil de energía libre exacto a una velocidad superior a la que se requeriría para alcanzar el mismo resultado con múltiples DMG QM/MM convencionales.

Velocidad	DRAr	$\Delta G^\ddagger$	ΔG^o
0.02*	1	21.0	-10.0
0.5	1	24.6	-7.0
	4	18.4	-10.3
	10	22.6	-8.1
	20	20.6	-12.8
1	1	25.9	-0.9
	4	25.5	-6.5
	10	21.6	-9.5
	20	21.6	-11.3
2	1	33.8	8.4
	4	29.0	-0.2
	10	23.6	-6.3
	20	23.4	-9.1

Cuadro 3.1: Valores de $\Delta G^\ddagger$ y ΔG^o para los perfiles de energía libre de la reacción de Corismato a Prefenato en solución obtenidos mediante la ecuación de Jarzynski para 10 trayectorias y diferentes velocidades y valores de DRAr. Resaltado en gris se muestra el perfil de referencia, obtenido con sólo 3 trayectorias y una velocidad de $0,02\text{\AA}/ps$. La energía libre se encuentra en unidades de $kcal/mol$.

También es interesante comparar los FEP obtenidos para diferentes velocidades, pero un mismo DRAr. Como es de esperarse, en la dinámica convencional (DRAr=1) existe una mejora notoria en la precisión de los valores de energía libre al disminuir la velocidad, como puede observarse al pasar de un resultado incorrecto a una velocidad de $2\text{\AA}/ps$ a un resultado razonable empleando una velocidad cuatro veces más lenta que ésta ($0,5\text{\AA}/ps$). Para un DRAr de 4, la mayor velocidad aquí presentada continúa dando resultados alejados del correcto, sin embargo ya a velocidades moderadas ($1\text{\AA}/ps$) los resultados obtenidos presentan una mejoría notoria. En los casos en los que se emplea un DRAr mayor (valores de 10 y 20), los resultados son casi indistinguibles para las velocidades presentadas, obteniéndose valores notablemente cercanos a los de referencia incluso para la velocidad más elevada. Estos resultados nos llevan a concluir que el esquema HyDRA puede ser implementado empleando velocidades más elevadas que las tradicionales obteniendo aún perfiles de energía libre exactos.

3.2.3.2 Influencia del Número de Trayectorias

Como se mencionó con anterioridad, la obtención de un perfil de energía libre depende de dos variables esenciales: el número de trayectorias y la velocidad. En la sección anterior se pudo observar cual es la influencia de la velocidad en un FEP y cómo es la relación entre estos y el valor de DRAr cuando se emplea el esquema de HyDRA. Por esta razón, continuamos el análisis de la influencia de HyDRA en la obtención del FEP, manteniendo fija ahora la velocidad y variando el número de trayectorias empleadas para realizar el promedio expo-

Cuadro 3.2: Valores de $\Delta G^\ddagger$ y ΔG^o para los perfiles de energía libre obtenidos mediante la ecuación de Jarzynski para una velocidad de $1\text{\AA}/ps$ y diferente número de trayectorias y valores de DRAr. Resaltado en gris se muestra el perfil de referencia, obtenido con una velocidad de $0,02\text{\AA}/ps$ y sólo 3 trayectorias. La energía libre se encuentra en unidades de $kcal/mol$.

Nº de Trayectorias	DRAr	$\Delta G^\ddagger$	ΔG^o
3*	1	21.0	-10.0
5	1	25.1	-0.6
	4	25.1	-7.8
	10	21.1	-9.9
	20	22.6	-9.9
10	1	25.9	-0.9
	4	25.5	-6.5
	10	21.6	-9.5
	20	21.6	-11.3
15	1	25.6	-0.9
	4	25.1	-7.5
	10	21.1	-9.9
	20	21.4	-11.5
20	1	25.2	-1.6
	4	25.1	-7.0
	10	21.1	-9.9
	20	21.1	-11.9

nencial de Jarzynski. Los resultados obtenidos se presentan en la Tabla 3.2. En la misma se puede observar que se han realizado 4 aproximaciones diferentes, variando el número de trayectorias de 5 a 20, y en cada una de ellas el valor de DRAr se ha variado también entre los valores de 1, 4, 10 y 20. En líneas generales, y en concordancia con la sección previa, los resultados obtenidos empleando el esquema HyDRA son más exactos que los obtenidos mediante MDMG tradicionales.

En particular, tomemos el caso de los perfiles obtenidos con sólo 5 trayectorias - valor cercano al empleado para realizar el perfil de referencia, aunque en éste último se empleó una velocidad 50 veces mayor a la utilizada en los FEPs de la Tabla 3.2-. Puede observarse para este caso que los valores de energía libre obtenidos aumentan su exactitud al hacerlo el valor de DRAr utilizado, especialmente para ΔG^o , donde puede observarse que para la dinámica tradicional el valor obtenido es termodinámicamente incorrecto, mientras que ya en el caso de DRAr=10 se predicen los valores correctos de energía libre a lo largo de todo el perfil. Esta tendencia entre la exactitud de los valores de energía libre y el valor de DRAR se mantiene para los diferentes valores del número de trayectorias aquí representados. Este resultado es de vital importancia para el sistema escogido, ya que permite reducir enormemente el tiempo total de simulación computacional.

3.2.4 Aplicación de HyDRA en la Enzima BsCM

Para analizar en mayor detalle las ventajas que se obtienen empleando la estrategia presentada, completamos el análisis del sistema Corismato-Prefenato estudiando esta reacción dentro de la enzima *Bacillus subtilis* Chorismate Mutase (BsCM). Esto nos permitirá comparar los valores de energía libre obtenidos computacionalmente con los experimentales, de manera de contar con una validación externa al sistema computacional, que es además el empleado cuando se busca validar un mecanismo de reacción propuesto. Los resultados obtenidos se presentan resumidos en la Tabla 3.3, donde, de forma similar a la reacción en solución se ha computado como referencia un FEP empleando 3 trayectorias a una velocidad de $0.02 \text{ \AA/ps}$. Debido a los resultados obtenidos en la sección 3.2.3.2, donde se pone de manifiesto que, para este sistema, el parámetro que más afecta la exactitud del FEP es la velocidad, en la obtención de los perfiles de energía libre en la enzima se mantuvo fijo el número de trayectorias en un valor de 10. Por esta misma razón, se analizó la relación entre la exactitud del FEP, el valor de DRAr y la velocidad de la simulación, tomándose en este caso valores entre 50 y 200 veces más grandes que la velocidad de referencia. En el caso de la enzima no se muestran los resultados para DRAr=20 ya que los mismos no muestran ninguna variación respecto a aquellos en los cuales se empleó el valor de DRAr igual a 10.

Velocidad	Nº de Trayectorias	DRAr	$\Delta G^\ddagger$	ΔG°
0.02	3	1	9.6	-16.5
1	10	1	12.2	-11.4
		4	10.1	-15.4
		10	8.3	-15.6
2	10	1	15.2	-9.5
		4	12.5	-12.6
		10	9.9	-12.4
4	10	1	15.4	-5.2
		4	11.8	-7.5
		10	9.6	-12.0

Cuadro 3.3: Valores relevantes del perfil de energía libre para la reacción de Corismato a Prefenato en la enzima BsCM, obtenidos con 10 trayectorias empleando la ecuación de Jarzynski y diferentes velocidades y valores de DRAr. La velocidad tiene unidades de $\text{\AA/ps}$ mientras que la energía libre se encuentra en kcal/mol .

Los resultados en la enzima muestran la misma tendencia que aquellos obtenidos en solución, aunque las diferencias netas entre los valores son menores. Se puede notar que, por ejemplo, ΔG° decrece al implementar el esquema de HyDRA para todos los casos considerados, haciéndose especialmente notorio para el caso de las velocidad de $4 \text{ \AA/ps}$.

Para $\Delta G^\ddagger$ el incremento en el valor de DRAr mantiene la tendencia antes observada de disminuir la barrera de activación. De hecho, es posible obtener con DRAr=10 un perfil de energía libre con barreras similares a la de la referencia para todas las velocidades aquí

presentadas, mientras que para los perfiles calculados sin utilizar el esquema de HyDRA existe una sobreestimación de la barrera de entre $3 - 5 \text{ kcal/mol}$.

3.3 Análisis del Costo Computacional

Para completar el análisis de la eficiencia computacional esquema de HyDRA es necesario realizar un estudio del tiempo de cómputo asociado a dicho esquema, en relación al tiempo requerido para realizar dinámica computacional tradicional. Este análisis se realizará teniendo en cuenta el tiempo requerido para realizar un paso de DM, para el caso de dinámica convencional y para cuando se emplea el esquema HyDRA, de manera que pueda analizarse además el tiempo total de cómputo necesario para obtener resultados que se ajusten a valores experimentales.

Para poder comparar los tiempos de dinámica entre el esquema tradicional y HyDRA se realizaron cálculos de dinámica molecular para diferentes sistemas químicos-biológicos empleando ambos esquemas. La diferencia entre los sistemas radica principalmente en el número de átomos, tanto en el valor total como también la relación entre el número de átomos presentes en el subsistema clásico y en el cuántico. La importancia de este análisis radica en el hecho de que es necesario conocer cuánto tiempo computacional es agregado (*overhead*) a la dinámica convencional cuando se emplea el esquema HyDRA, debido a que este es el tiempo que será necesario compensar reduciendo el número de pasos cuánticos o disminuyendo la velocidad, sin perder exactitud. Es decir que, hasta el momento hemos demostrado que es posible obtener perfiles de energía libre exactos empleando el esquema HyDRA en un número menor de trayectorias o una velocidad menor, sin embargo, no hemos demostrado hasta el momento que esto se traduzca efectivamente en una reducción del tiempo total de cómputo. Para esto realizaremos primero un análisis teórico del costo computacional total, en el cual consideraremos que $t_{MM} = \alpha \cdot t_{QM}$, donde t_{MM} es el tiempo empleado en realizar un único paso clásico, t_{QM} es el tiempo empleado en realizar un paso híbrido QM/MM (o conjunto), y α es un parámetro que depende de cada sistema particular y que permite relacionar ambos tiempos, y que asume valores $\alpha \leq 1$. Es decir que el costo temporal total de un paso de dinámica convencional híbrida será t_{QM} . Con esta notación el tiempo total consumido en una trayectoria de dinámica molecular de n_s pasos (cuánticos) será

$$(3.1) \quad T_1 = n_s [t_{QM} + (DRAr - 1) t_{MM}]$$

donde t_{QM} es el tiempo de dinámica sin HyDRA y $(DRAr - 1) t_{MM}$ es el tiempo agregado por realizar $DRAr$ pasos dentro del nuevo esquema. A su vez, el número de pasos de una DMG puede relacionarse con la velocidad de la misma según $n_s = \delta \cdot v^{-1}$ donde δ es el cociente entre la distancia recorrida por la coordenada de reacción (en Å) y el paso de

tiempo empleado, dt . Así, combinando las expresiones anteriores, el tiempo computacional total requerido para realizar un FEP completo empleando la ecuación de Jarzynski y N_{Tr} trayectorias independientes será

$$(3.2) \quad T = \delta \cdot \left(\frac{N_{Tr}}{v} \right) \cdot (t_{QM}(1 + \alpha(DRAr - 1))) = \delta \cdot SS \cdot Sy$$

donde se han agrupado las diferentes contribuciones al tiempo total, T . Tendremos así tres términos donde δ se relaciona con los parámetros técnicos de la simulación (dt y coordenada de reacción), SS agrupa los valores que dependen de la estrategia de simulación utilizada en el cálculo del FEP (número de trayectorias y velocidad), y Sy es el término que depende exclusivamente de cada sistema y de cómo se realiza la división del mismo en la región cuántica y clásica (t_{QM} , $DRAr$ empleado y α). A fin de poder comparar el costo computacional agregado por el esquema HyDRA en las simulaciones, definimos el tiempo total relativo como

$$(3.3) \quad \tau_R = \frac{T}{T^o} = \frac{SS \cdot Sy}{SS^o \cdot Sy^o}$$

donde el subíndice o hace referencia a los valores de la simulación de referencia. Es decir que τ_R representa el tiempo necesario para obtener un dado FEP, relativo a otro esquema de simulación diferente. En el mismo se tienen en cuenta todos los parámetros relevantes, como ser la velocidad, el número de trayectorias, el valor de $DRAr$ y las características particulares del sistema.

El análisis de τ_R se realizó considerando un valor constante para δ en todos los casos, pero variando los términos SS y Sy . Consideremos entonces una estrategia de simulación uniforme para todos los sistemas y analicemos la influencia de Sy en el tiempo total T . Para esto se realizó una dinámica molecular (no guiada) de 1ps para diferentes valores de $DRAr$ en seis sistemas diferentes en solución - Nitrato, metanotiolato y peróxido de hidrógeno, corismato, corismato en la enzima BsCM, mioglobina y kinasa ERK (extracelular signal-regulated kinase)- empleando el resultado obtenido para $DRAr=1$ (es decir, dinámica convencional) como simulación de referencia, Sy_o . Cada uno de los sistemas presenta un número diferente de átomos clásicos y cuánticos, donde los últimos han sido tratados a nivel DFT-PBE. Las dinámicas fueron realizadas en un procesador Intel Core i3, siendo este el único trabajo ejecutado. Los valores promedios para τ_R y Sy se presentan en la Tabla 3.4. Los resultados obtenidos demuestran que, para la mayor parte de los sistemas estudiados, emplear un $DRAr$ de 4 representa apenas un incremento temporal del 10% respecto a una dinámica convencional QM/MM (τ_R asume un valor de 1,1). Para valores de $DRAr$ de 10 y 20, el costo asociado con emplear el esquema de HyDRA representa un incremento de 100% del tiempo, aproximadamente. Se evidencia de esta forma que cualquier mejora en el tiempo de

Sistema	NO_3^- Agua	CH_3S^- + H_2O_2	Cor Agua	Cor BsCM	Mb	ERK
N° átomos QM	4	9	24	24	42	55
N° átomos MM	2070	8808	9834	40763	18758	53478
Sy° (s)	7.2	9.8	185.0	209.1	1059.4	776.8
$\tau_R(DRAr = 4)$	1.1	1.1	1.2	1.1	1.1	1.1
$\tau_R(DRAr = 10)$	1.4	1.5	1.3	1.4	1.2	1.3
$\tau_R(DRAr = 20)$	1.9	2.2	1.6	1.7	1.5	1.5

Cuadro 3.4: Costo computacional para los diferentes esquemas híbridos QM/MM a nivel DFT, empleando HyDRA. Sy° es el tiempo promedio en segundos requerido para realizar un paso conjunto (es decir un paso QM y uno MM) en una dinámica molecular convencional, o equivalentemente $DRAr=1$, mientras que $\tau_R(DRAr)$ es el cociente entre el tiempo promedio (en segundos) requerido para realizar un paso de dinámica en el esquema HyDRA (1 paso QM y $DRAr$ pasos clásicos) y Sy° .

cómputo total, requerido para calcular un FEP exacto, vendrá dada por una combinación de los valores de SS y Sy .

Para poder analizar esta relación se combinaron los resultados volcados en la Tabla 3.4 y los FEPs obtenidos en la Sección 3.2, con el fin de poder comparar los tiempos de simulación totales y los resultados obtenidos para cada esquema de simulación realizado. A continuación, se calcularon los tiempos de simulación totales relativos para los distintos casos de la reacción de Corismato a Prefenato estudiados previamente, donde el valor Sy_o se tomó como el tiempo computacional total requerido para realizar el perfil que fue empleado como referencia para esta reacción ($v = 0,02\text{\AA}/ps$, $N_{Tr} = 3$ y $DRAr = 1$). Entre todos los resultados obtenidos, se seleccionaron algunos casos ejemplificadores que fueron volcados en la Tabla 3.5, donde se especifica el tiempo de cómputo relativo (τ_R), los parámetros que influyen en el tiempo de cómputo total (N_{Tr} , v y $DRAr$) y la diferencia en energía libre entre el perfil obtenido y el de referencia.

Cuadro 3.5: Comparación de FEP para la reacción de Corismato a Prefenato en solución, para casos obtenidos en el mismo tiempo de cómputo relativo, τ_R . Velocidad en $\text{\AA}/ps$ y ΔG en $kcal/mol$.

τ_R	N_{Tr}	v	$DRAr$	$\Delta G^\ddagger$	ΔG°
1	3	0.02	1	21.0	-10.0
0.1	10	0.5	1	24.6	-7.0
	15	1	10	21.1	-9.9
0.03-0.04	5	1	1	25.1	-0.6
			10	21.1	-9.9

Tomemos como primer caso el mejor estimador para $DRAr=1$, obtenido mediante 10 DMG con una velocidad de $0,5\text{\AA}/ps$ y que difiere en aproximadamente $3kcal/mol$ de la referencia.

Para este caso $\tau_R = 0,1$, es decir que es un resultado que se alcanza en aproximadamente el 10% del tiempo necesario para reproducir el perfil de referencia. Empleando el esquema de HyDRA con un valor de DRAr igual a 10 es posible obtener, en el mismo tiempo de cómputo total que el mejor estimador para DRAr=1, un FEP que es indistinguible de la referencia (empleando 15 trayectorias a una velocidad de $1\text{\AA}/ps$). Más aún, es posible obtener un perfil indistinguible de la referencia en un tiempo de cómputo cerca de 3 veces menor, utilizando DRAr=10. Puede observarse de la Tabla 3.5 que, para este caso, la disminución en el tiempo se debe directamente a la reducción del número de trayectorias empleado para obtener el FEP, de 15 en el primer caso mencionado a 5 en el que requiere 3 veces menos tiempo. Además de esto, el incremento en el tiempo relativo de cómputo al emplear el esquema de HyDRA para la velocidad de $1\text{\AA}/ps$ y $N_{Tr} = 5$ es despreciable, haciendo que puedan obtenerse ambos perfiles para un valor de τ_R semejante. Sin embargo, aunque el tiempo de cómputo sea equivalente, el FEP obtenido con dinámica convencional carece de la exactitud alcanzada con el esquema de HyDRA. Esto muestra que, empleando HyDRA y con los parámetros escogidos apropiadamente, es posible reducir el tiempo de cómputo total sin perder exactitud.

3.3.1 Poder Catalítico

Una de las cantidades más requeridas en el campo de la enzimología es el poder catalítico de una enzima particular, por lo que, medir el tiempo que se emplearía en realizar este tipo de cálculo, nos otorga una forma adicional de comparar el esquema HyDRA con el tradicional. Por esta razón utilizamos los valores obtenidos previamente para calcular el poder catalítico de la enzima BsCM, y el tiempo de cómputo necesario para alcanzar un valor cercano al experimental ($\Delta\Delta G^\ddagger = 9,1kcal/mol$) dentro del esquema tradicional y con HyDRA. En este caso no consideraremos el perfil de referencia, ya que tendremos en cuenta aquellos valores de energía libre que permitan aproximar con mayor exactitud el valor de $\Delta\Delta G^\ddagger$ al experimental. Dentro del esquema tradicional, el mejor estimador del poder catalítico se consigue al combinar los FEPs obtenidos para una velocidad de $v = 0,5\text{\AA}/ps$ en solución y $v = 1\text{\AA}/ps$ en BsCM, empleando 10 trayectorias en ambos casos, obteniendo un valor de $\Delta\Delta G^\ddagger$ de $12,4kcal/mol$. Mientras que para DRAr=10 el poder catalítico es de $12,8kcal/mol$, combinando los perfiles obtenidos a una velocidad de $v = 1\text{\AA}/ps$ y 5 trayectorias para la reacción en solución y 10 en la enzima, resultado que es posible obtener en aproximadamente un tercio del tiempo requerido para calcular el mejor estimador con DRAr=1.

3.4 Discusión

El presente trabajo introduce el esquema HyDRA, realizado durante este trabajo de tesis, que permite la relajación selectiva del entorno clásico en simulaciones híbridas QM/MM de reacciones químicas y/o enzimáticas, resultando así en una convergencia más rápida (en términos de tiempo computacional) del perfil de energía libre obtenido mediante la ecuación de Jarzynski. Para validar este esquema se estudió la conversión de Corismato a Prefenato, donde HyDRA nos permitió obtener energías libres de reacción, y las barreras asociadas, con notoria exactitud en menos del 10 % del costo computacional tradicional (el tiempo que hubiese requerido obtener el mismo nivel de exactitud sin emplear HyDRA). En esencia, estas observaciones reflejan el comportamiento esperado del uso de HyDRA, que da lugar a trayectorias más cercanas al equilibrio, incluso para velocidades más de 100 veces mayores que las necesarias para ser consideradas cuasi-estáticas.

Desde el punto de vista de la mecánica estadística, nuestros resultados ponen también de relieve algunas propiedades interesantes del comportamiento de las dinámicas de no-equilibrio, y los perfiles de trabajo asociados, en relación con los parámetros de simulación referidos a la estrategia, como ser la velocidad y el número de trayectorias. Los resultados actuales, y la experiencia previa del grupo de trabajo en el tema [15, 16, 64, 75], ponen en evidencia la dificultad de obtener una FEP preciso cuando la velocidad es demasiado rápida, independientemente del número de trayectorias realizadas (a excepción de valores muy elevados). Esto se debe a que, cuando el sistema evoluciona lejos del equilibrio, la desviación estándar (σ) asociada a los valores de trabajo es mucho mayor que $k_b.T$, dificultando enormemente la tarea de conseguir resultados significativos y estadísticamente convergidos. Este efecto fue analizado desde un punto de vista teórico por Pohorille y colaboradores [62], los mismos estiman que si $\sigma = k_b.T$, un sexto de las trayectorias muestreadas tendrán valores cercanos a la energía libre real, asegurando así un resultado preciso (convergado) y exacto. Sin embargo, este número disminuye a 1/40 cuando $\sigma = 2k_b.T$, y más aún, cuando $\sigma = 5k_b.T$ esta fracción decae a 1/3,500,000, lo que significa que podríamos nunca muestrear una trayectoria cercana al valor correcto de energía libre. Los resultados obtenidos con HyDRA se encuentran más cercanos a las trayectorias de equilibrio, incluso a velocidades elevadas, reduciendo así significativamente la desviación estándar de los valores de trabajo y, por tanto, logrando un aumento significativo de la exactitud media y la convergencia del perfil.

Con respecto al costo computacional adicional ocasionado por el aumento de los pasos puramente clásicos, este ha sido extensamente estudiado para la reacción modelo (Corismato a Prefenato) resultando, para cálculos con un nivel teórico DFT-PBE, en apenas un leve incremento temporal. Del mismo podemos concluir que el uso de un DRAr entre 4-10 tiene el potencial de generar estimaciones de energía libre más exactas empleando un menor número

de trayectorias y velocidades más rápidas, pudiendo emplearse en estudios QM/MM de reacciones químicas o enzimáticas. Los resultados obtenidos en la sección 3.2 pueden además ser contrastados con datos experimentales y computacionales disponibles en literatura. Para la conversión de Corismato a Prefenato en solución, los estudios experimentales (Andrews et. al. [1]) muestran que existe una energía libre de activación de $24,5\text{kcal/mol}$. Mientras que para la enzima BsCM la barrera de activación es de $15,4\text{kcal/mol}$ (Kast et. al.[43]), evidenciando la eficiencia de dicha enzima, que logra disminuir la energía libre de activación en $9,1\text{kcal/mol}$. Además, se pueden encontrar datos en literatura [44] que determinan que el ΔG° de la reacción es de $-13,4\text{kcal/mol}$. Para poder comparar estos resultados con los obtenidos en este capítulo es necesario tener en cuenta el conocido hecho [15, 16] de que DFTB tiende a subestimar los valores de energía libre. Por esta razón, la mejor forma de evaluar la exactitud de los FEP obtenidos es comparando los valores de energía libre de activación para la reacción en la enzima y en solución ($\Delta\Delta G^\ddagger$).

Si consideramos los dos perfiles de referencia utilizados (tanto para la reacción en solución como la que sucede en la enzima), obtenemos una eficiencia enzimática de $11,4\text{kcal/mol}$, sobreestimando ligeramente el valor experimental. Si calculamos este mismo valor para el mejor estimador sin HyDRA, obtenido para 10 trayectorias a $v = 0,5\text{\AA/ps}$ en solución y $v = 1\text{\AA/ps}$ en BsCM, se obtiene un $\Delta\Delta G^\ddagger$ de $12,4\text{kcal/mol}$, similar al valor de la referencia. Un resultado equivalente ($\Delta\Delta G^\ddagger = 12,8$) podría obtenerse empleando $DRAr = 10$ en un tercio del tiempo requerido para obtener el mejor estimador sin HyDRA. Más aún, es posible obtener un valor en el mismo orden de magnitud (14kcal/mol) empleando los resultados obtenidos para las velocidades más rápidas ($2\text{\AA/ps}$ en solución y $4\text{\AA/ps}$ en BsCM) y $DRAr = 10$, evidenciando una vez más el nivel de precisión alcanzado con el esquema de HyDRA, incluso a velocidades elevadas.

En resumen, los resultados del estudio de la reacción de Corismato a Prefenato muestran que el uso de la estrategia HyDRA resulta en perfiles de energía libre más exactos y precisos, incluso cuando se emplean velocidades rápidas. La evidencia aquí presentada muestra la potencialidad del esquema HyDRA para obtener FEP en tiempos computacionales moderados, presentándose así como una herramienta teórica de gran relevancia en el estudio de reacciones enzimáticas o en solución.

3.5 Conclusiones

Hemos implementado (en el paquete de simulación molecular AMBER) y probado exhaustivamente el Esquema de Relajación Híbrida Diferencial (HyDRA), el cual permite obtener perfiles exactos de energía libre de reacciones químicas en solución y enzimas a un costo

computacional reducido, en el contexto de simulaciones de múltiples dinámicas moleculares guiadas y la igualdad de Jarzynski. El poder del método se basa en su capacidad para permitir una relajación más rápida del entorno clásico (proteína/solvente), frente al subsistema reactivo cuántico. La aplicación del método permitirá estudiar reacciones, mediante el método de MDMG en combinación la IJ, que eran inaccesibles hasta el momento, debido a su elevado costo computacional.

APLICACIÓN DE HyDRA A OTROS SISTEMAS DE INTERÉS

Debido al interés despertado en el grupo de trabajo por el desarrollo del esquema HyDRA, el mismo fue aplicado para el estudio de diversas reacciones de interés, tanto en enzimas como en solución. En particular, y en estrecha colaboración con otros miembros del grupo de investigación, se estudiaron las reacciones de oxidación de metanotiol por H_2O_2 , y se propuso un mecanismo consistente para las Hidrolasas de Zn de *Mycobacterium tuberculosis*. En todos los casos se realizaron pruebas de velocidad, número de trayectorias y DRAr, para encontrar los valores óptimos de estos parámetros.

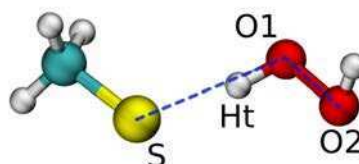
4.1 Oxidación de Metanotiol por Peróxido de Hidrógeno

La oxidación de compuestos celulares que contienen tiol desempeña un rol fundamental en muchos procesos biológicos. Entre los posibles oxidantes, el peróxido de hidrógeno (H_2O_2) es producido en muchos tipos de células como una respuesta a una gran variedad de estímulos extracelulares y podría funcionar como un mensajero intracelular. Por esta razón, la reacción de oxidación de metanotiol en solución ($\text{CH}_3\text{S}^- + \text{H}_2\text{O}_2 \longrightarrow \text{CH}_3\text{SO}^- + \text{H}_2\text{O}$) representa un paso clave en la comprensión de estos procesos. La misma ha sido previamente estudiada empleando la estrategia de *Umbrella Sampling*[46, 85], donde se ha concluido que el solvente juega un papel clave en el posicionamiento de los reactivos. Esto se debe a que existe una redistribución de carga significativa en las primeras etapas de la reacción, y a que existe una transferencia de hidrógeno entre los átomos de oxígeno de H_2O_2 que se produce luego de alcanzar el estado de transición. Debido al papel clave que desempeña la dinámica del solvente, se buscó estudiar esta reacción empleando el esquema de HyDRA.

4.1.1 Condiciones de Simulación

Para poder llevar a cabo la reacción se solvató al soluto ($\text{CH}_3\text{S}^- + \text{H}_2\text{O}_2$) en una caja octaédrica con 4290 moléculas de agua de tipo TIP3P. Inicialmente se realizó una equilibración del sistema para alcanzar la temperatura y densidad correctas, equivalente a la realizada en la sección 3.2, seguida de una dinámica QM/MM de $50ns$, donde el soluto fue tratado a nivel cuántico DFTB. De esta dinámica se seleccionaron 20 estructuras para realizar las MDMG, las cuales fueron sometidas a una DM híbrida de $1ps$ con un nivel cuántico de teoría DFT-PBE. La coordenada de reacción propuesta se encuentra esquematizado en la figura

Figura 4.1: Coordenada de reacción para la oxidación de metanotiol por peróxido de hidrógeno.



4.1. En la misma, sólo se consideran las distancias entre los átomos de oxígeno y el azufre, mientras que la transferencia de protón, necesaria para formar el producto, se espera ocurra espontáneamente durante la dinámica. Debido a esto no es posible realizar la reacción en sentido inverso, razón por la cual el esquema de Jarzynski fue descartado con anterioridad [85] para obtener el perfil de energía libre. En este contexto, nos propusimos realizar el FEP para esta reacción empleando el esquema de HyDRA, efectuando sólo el proceso *forward*. Además de la mencionada transferencia de protón, esta reacción involucra cambios significativos en la solvatación del sistema, por lo cual esperamos que los beneficios de emplear el esquema de HyDRA sean notorios.

La reacción fue estudiada empleando el nivel de teoría DFT, haciendo uso del programa LIO, que posee una interfaz con AMBER para realizar cálculos de dinámica híbrida QM/MM y que implementa el funcional de intercambio y correlación propuesto por Perdew, Burke y Ernznheof (PBE). El programa LIO fue testeado en AMBER en numerosos sistemas [31, 32, 55, 56] demostrando un poder de predicción alto. Sin embargo requiere un costo computacional mucho más elevado que el requerido para realizar DFTB (el método empleado en la sección anterior), por lo que la utilización de HyDRA en este esquema híbrido podría representar una ganancia significativamente superior en el costo computacional final.

En todos los casos, los perfiles de energía libre de la reacción fueron realizados empleando 10 trayectorias de DMG con la ecuación de Jarzynski. Como en los casos previos, se realizaron FEPs con diferentes valores de velocidad, tomando como referencia en este caso el tiempo de simulación total empleado en la obtención del perfil con *Umbrella Sampling* y disminuyéndolo

en hasta la mitad el tiempo. A su vez, para cada velocidad se realizaron diferentes perfiles para los valores de DRAr de 1, 4 y 10.

4.1.2 Resultados

Velocidad	DRAr	$\Delta G^\ddagger$	$\sigma^\ddagger$	ΔG^o	σ^o
0.5	1	12.7	6.1	-35.9	11.4
	4	12.0	2.5	-34.8	8.6
	10	10.1	1.1	-35.9	5.6
1	1	14.3	6.6	-27.0	9.5
	4	11.7	5.6	-31.0	9.4
	10	12.2	4.5	-31.9	9.4

Cuadro 4.1: Valores de energía libre representativos del FEP para la reacción de oxidación de metanotiol por peróxido de hidrógeno en solución y la desviación estándar asociada a cada uno de estos valores. Los resultados fueron obtenidos utilizando 10 DMG con la ecuación de Jarzynski, para diferentes velocidades y valores de DRAr. La velocidad se encuentra en unidades de $\text{\AA}/ps$ y la energía libre en $kcal/mol$.

La tabla 4.1 muestra los valores de energía libre obtenidos para esta reacción, junto con los valores de desviación estándar asociados. Para la velocidad más rápida, $1\text{\AA}/ps$, los resultados obtenidos muestran la misma tendencia que para la reacción de Corismato a Prefenato, donde la exactitud del FEP aumenta cuando se emplea el esquema HyDRA (recordemos que una de las ventajas de emplear la igualdad de Jarzynski es que el estimador de energía libre será siempre mayor o igual que el FEP real), mientras que los valores de desviación estándar son similares para los diferentes DRAr empleados, indicando que si bien la exactitud de las medidas mejora, no necesariamente su convergencia.

Para la velocidad más lenta ($0,5\text{\AA}/ps$), en cambio, podemos observar que los valores de energía libre no muestran variaciones significativas, alcanzando una diferencia máxima para el valor de la barrera de activación de $3kcal/mol$ entre las simulaciones con DRAr 1 y 10. Sin embargo la desviación estándar asociada a dichas mediciones disminuye notoriamente al incrementar el valor de DRAr, lo cual indica una mejora clara en la convergencia de los perfiles obtenidos empleando el esquema de HyDRA frente a aquellos de dinámica convencional. Estos resultados reflejan, nuevamente, la ganancia en precisión que se obtiene al emplear el esquema de HyDRA, contribuyendo así a obtener FEP con mayor credibilidad.

4.2 Hidrolasas de Zinc

Las infecciones por *Mycobacterium tuberculosis* (Mtb) continúan siendo un problema a escala global, a tal punto que más de un millón de personas mueren anualmente por es-

ta enfermedad, por lo que la caracterización y validación de nuevos blancos biológicos¹ drogables es un intenso campo de investigación. Entre los posibles blancos drogables se han propuesto dos hidrolasas de zinc, MshB (Rv1170) y MA-amidasa (Rv3717), que han sido caracterizadas estructuralmente. Las hidrolasas de Zn son capaces de catalizar la hidrólisis de enlaces químicos empleando agua u otros nucleófilos en su sitio activo, y conforman uno de los grupos más grandes de enzimas que contienen zinc por lo que han sido objeto de extensos estudios, especialmente en humanos. Sin embargo, el mecanismo de reacción de las hidrolasas de Zn en bacterias no había sido clarificado hasta el momento. Algunos intentos de clarificar dicho mecanismo han sido propuestos en analogía a los correspondientes en hidrolasas de Zn humanas, a pesar de que estos fallaban en resolver cuestiones claves, como ser la coordinación del Zn durante la reacción y su rol en la catálisis.

Debido a la relevancia de la temática y a la dificultad que presentaba realizar esta reacción con otras metodologías, se realizó el estudio de mecanismos de reacción para estas enzimas utilizando el esquema de HyDRA.

4.2.1 Condiciones de Simulación

Todos los parámetros clásicos para ambos sistemas fueron tomados del campo de fuerzas de AMBER ff99SB [38] para los 20 aminoácidos, mientras que las moléculas de aguas fueron representadas con el modelo TIP3P [42]. Los parámetros clásicos para el ión Zn y su esfera de coordinación (empleados durante las dinámicas de equilibración) fueron construidos empleando el procedimiento estándar recomendado por el paquete AMBER. Las simulaciones híbridas QM/MM se llevaron a cabo empleando el nivel de teoría SCC-DFTB [19], implementado en el programa SANDER [22]. Ambos sistemas fueron sujetos a un protocolo de relajación empleando dinámica clásica, luego de esto el sistema fue lentamente calentado hasta alcanzar los 300K y 50ns de dinámica molecular de equilibrio (en el ensamble NPT) fueron realizados para permitir que el sistema relaje. Durante las simulaciones, la temperatura y la presión se mantuvieron constantes empleando el termostado de Langevin y el barostato de Berendsen [7], se emplearon además condiciones periódicas de contorno. Partiendo de las estructuras equilibradas se realizaron 300ps de DM-QM/MM, con un paso de integración de 1fs, para cada estructura, y de esas se obtuvieron las estructuras necesarias para realizar las MDMG. Para ambas proteínas, el subsistema cuántico consistió en el ión de Zn, las cadenas laterales de los residuos coordinados y de la base del sitio activo, las moléculas de agua coordinadas y/o reactivas, y el enlace amida del sustrato, cerrado con un grupo metilo (el sustrato se corresponde con una N-metil-acetamida).

¹Un blanco biológico (o *target*) es cualquier elemento dentro de un organismo vivo al cual otra entidad es dirigida, o se une, de forma tal que su comportamiento o función se vea modificado, siendo las proteínas los blancos más usuales.

4.2.2 Resultados

Para simular la reacción completa, se dividió el proceso en tres pasos: generación del ión hidroxido (activación), ataque nucleofílico (primer paso de reacción) y ruptura del enlace C-N (segundo paso de reacción). Para llevar a cabo la reacción y calcular el perfil de energía

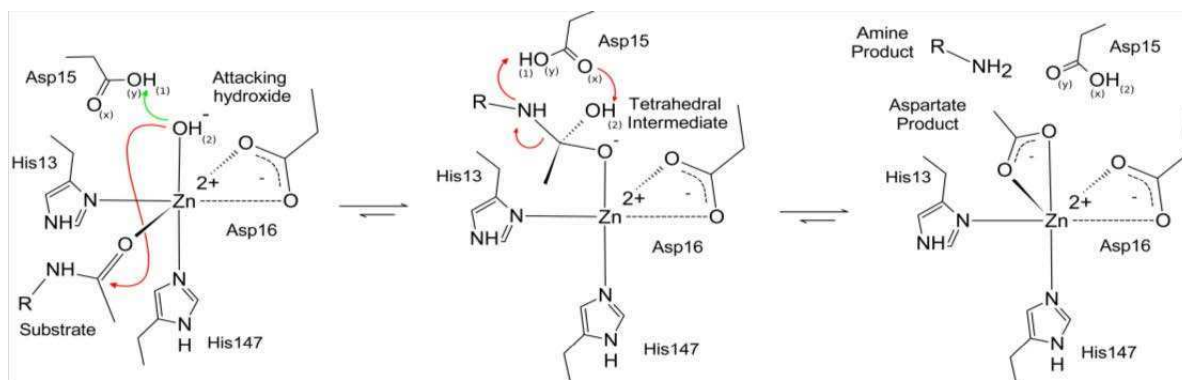


Figura 4.2: Mecanismo de reacción propuesto para las Hidrolasas de Zn.

libre correspondiente, se emplearon las siguientes coordenadas de reacción:

- Para el paso de activación se empleó una coordenada doble, que resulta de la diferencia entre las distancias entre el oxígeno del agua y el protón que se transfiere (dO_wHt) y la distancia entre el protón que se transfiere y el carboxilato de la base activa que acepta al oxígeno (dH_tAcO):

$$RC_0 = dO_wHt - dH_tAcO$$

- Para el primer paso de reacción se utilizó la distancia entre el oxígeno del hidróxido (O_{OH}) y el carbonilo del enlace amida (C):

$$RC_1 = dO_{OH}C$$

- En el segundo paso de la reacción la coordenada es notoriamente más compleja, combinando la distancia del enlace amida (dCN) con una diferencia de distancias entre la base del sitio activo (dH_tAcO) y el nitrógeno de la amida al cual se transfiere el protón (dNH_t):

$$RC_2 = dCN - (dNH_t - dH_tAcO)$$

En la Figura 4.2 se encuentra esquematizado el mecanismo de reacción propuesto.

Inicialmente, se realizó un estudio estructural de los residuos que conforman el sitio activo, donde se observó que la generación del ión hidróxido no tiene energía de activación ² (los detalles pueden encontrarse en el trabajo publicado[66]).

4.2.2.1 Primer Paso de Reacción

Una vez formado el ión hidróxido, la primera etapa de reacción corresponde al ataque del ión formado, al carbonilo del enlace amida que será hidrolizado. Para obtener el perfil de energía libre asociado a esta reacción, se realizaron 20 simulaciones independientes de DMG con una velocidad de $0,1\text{\AA}/ps$ (es decir, 12.000 pasos QM/MM en cada trayectoria), y empleando valores de DRAr de 1, 4 y 10. Para MshB, el perfil se obtuvo combinando, de manera óptima, las reacciones en sentido directo e inverso (10 curvas de trabajo para cada sentido de la reacción). Mientras que, para MA-Amidasa sólo se consideraron las reacciones en sentido directo, debido a que las de sentido inverso realizaban un camino diferente, mostrando una alta histéresis. Los perfiles de energía libre correspondientes a este paso se muestran en la Figura 4.3, MshB en (a) y MA-Amidasa en (b). En el mismo se muestran los valores obtenidos para diferentes valor de DRAr (1, 4 y 10 en negro, rojo y azul, respectivamente).

El efecto de DRAr, se evidencia claramente en ambas enzimas y muestra que, para la velocidad y el número de trayectorias empleado, es necesario utilizar un valor de DRAr igual a 10 para obtener los mejores resultados. Como se demostró anteriormene, el efecto de emplear el esquema de HyDRA consiste en reducir tanto la barrera como la energía libre de reacción. En comparación con las DMG estándar, el efecto de HyDRA es dramático, reduciendo tanto las alturas de la barrera como las energías libres de reacción hasta 10kcal/mol . Recordemos que, al utilizar la igualdad de Jazrynski, los valores de barrera y de energía libre de reacción, contienen sólo errores por sobreestimación, por lo que la disminución de su valor neto se asocia directamente con una mayor precisión.

Su efecto también resalta la relevancia del ajuste dinámico del entorno proteico durante la reacción. Curiosamente, el efecto es diferente en ambas enzimas, ya que mientras para la MA-Amidasa se observa una reducción más significativa en la barrera, para MshB se observa un cambio mayor en la energía libre de reacción y, por tanto, la estabilidad prevista del intermediario tetraédrico. El origen de esta diferencia no está claro, pero posiblemente se derive del papel desempeñado por la coordinación Zn al hidróxido atacante.

Para analizar más a fondo el papel de la cordinación del Zn con el carbonilo del sustrato, se calculó también el FEP del ataque nucleofílico del hidróxido, comenzando esta vez desde una

²Para obtener el perfil de energía libre asociado a esta reacción, se realizaron 20 simulaciones independientes de DMG con una velocidad de $0,4\text{\AA}/ps$ (es decir, 3000 pasos QM/MM en cada trayectoria), empleando un DRAr de 10. Las simulaciones se realizaron tanto en sentido directo como inverso y la energía libre informada corresponde a una combinación óptima de ambos.

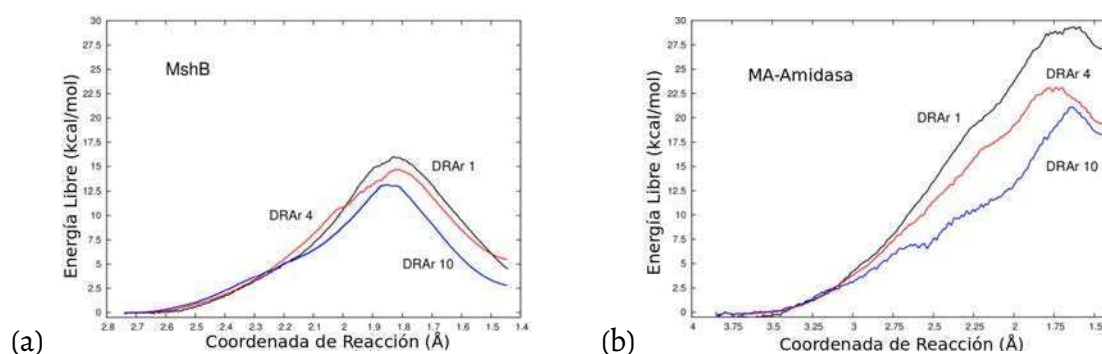


Figura 4.3: Perfiles de energía libre correspondientes al primer paso de la reacción (ataque nucleofílico del ión hidróxido), obtenidos empleando 20 trayectorias independientes de SMD con DRAr igual a 1,4 y 10 (negro, rojo y azul, respectivamente), en combinación con la ecuación de Jarzynski, utilizando la distancia entre el OH que ataca y el carbono del sustrato. (a) FEP para MshB. (b) FEP para MA-Amidasa.

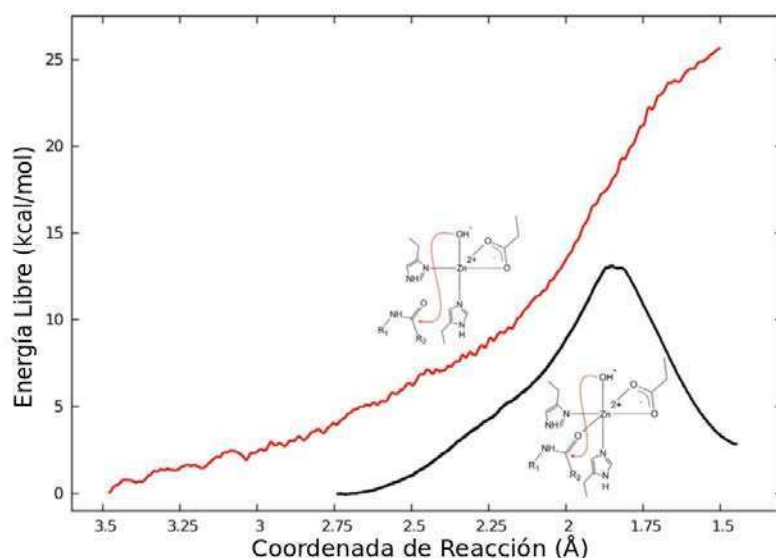


Figura 4.4: Perfiles de energía libre correspondientes al primer paso de la reacción (ataque nucleofílico del ión hidróxido), para la enzima MshB, partiendo desde una configuración inicial con el sustrato coordinado (en negro) y no coordinado (en rojo). Cada perfil se obtuvo con la IJ y 20 trayectorias de SMD con DRAr=10.

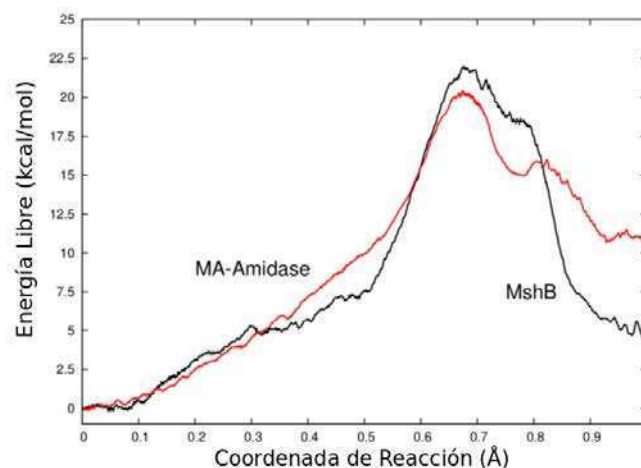
conformación donde el sustrato no se encuentra coordinado al Zn catalítico. Los resultados obtenidos se muestran en la Figura 4.4, donde se observa que el cofactor es crítico para estabilizar el intermediario tetraédrico “meta”-estable. Cuando el sustrato no se encuentra coordinado al Zn, no se alcanza un intermediario estable (mínimo en el FEP) para ninguna de las dos enzimas. Estos resultados apuntan de nuevo al papel primordial del ión Zn en la estabilización de la carga negativa, acumulada en el oxígeno del intermediario tetraédrico (ex-carbonilo) luego del ataque nucleofílico.

4.2.2.2 Segundo Paso de Reacción

Para ir del intermediario tetraédrico al producto, es necesario que el enlace C-N se rompa. Clave para que esto suceda, es necesario que la red de enlaces hidrógeno se reorganice

de forma tal de promover la transferencia de un protón a la amida NH (caso contrario la barrera energética para la ruptura es muy elevada). Para analizar la etapa de ruptura del enlace amida C-N, se calcularon los FEP correspondientes empleando 20 trayectorias de DMG a una velocidad de $1,4\text{\AA}/ps$ empleando un valor de DRAr de 10, y utilizando como coordenada de reacción una combinación de distancias que representa la ruptura del enlace C-N y la transferencia de protón al nitrógeno de la amida. Debido a que la reacción no es reversible molecularmente, dado el número de procesos involucrados (principalmente las transferencias de protón involucradas), las DMG se realizaron sólo en sentido directo.

Figura 4.5: Perfil de energía libre correspondiente al segundo paso de la reacción (ruptura del enlace C-N), para la enzima MshB (en negro) y MA-Amidasa (en rojo). Cada uno de los perfiles fue obtenido empleando 20 trayectorias independientes de SMD con DRAr=10, en combinación con la ecuación de Jarzynski.



Los resultados obtenidos se encuentran representados en la Figura 4.5, donde puede observarse que los FEP para ambas enzimas son similares. Esto es de esperarse, debido a que, para los residuos involucrados en este paso, las enzimas no muestran diferencias estructurales significativas.

El análisis de los perfiles, en el contexto de los parámetros estructurales relevantes (presentados en la Figura 4.6, para MshB) muestra que la reacción posee cuatro fases distintas. En la primera (zona 1), la energía libre aumenta lentamente, aproximadamente 7kcal/mol . Este segmento corresponde a la rotación de la base protonada (Asp15 en MshB y Glu200 en MA-Amidasa) y a la reducción de las distancias entre cada protón y su aceptor, generando la configuración apropiada para el doble intercambio de protones. Posteriormente, la energía se eleva abruptamente hasta alcanzar el estado de transición, cuando el N de la amida toma el protón de la base y el enlace C-N se rompe (zona 2). Poco después, se produce una pequeña disminución en la energía libre debido a la relajación del recién formado NH_2 , y a la transferencia del protón del intermediario OH al Asp15 (zona 3). Finalmente, la energía disminuye como resultado de la relajación de todo el sistema (zona 4).

Puede destacarse en esta reacción el papel clave que juega la correcta relajación del entorno para llevar a cabo la misma. El hecho de que el esquema HyDRA sea capaz de predecir el comportamiento de un sistema con una coordenada de reacción tan compleja, donde

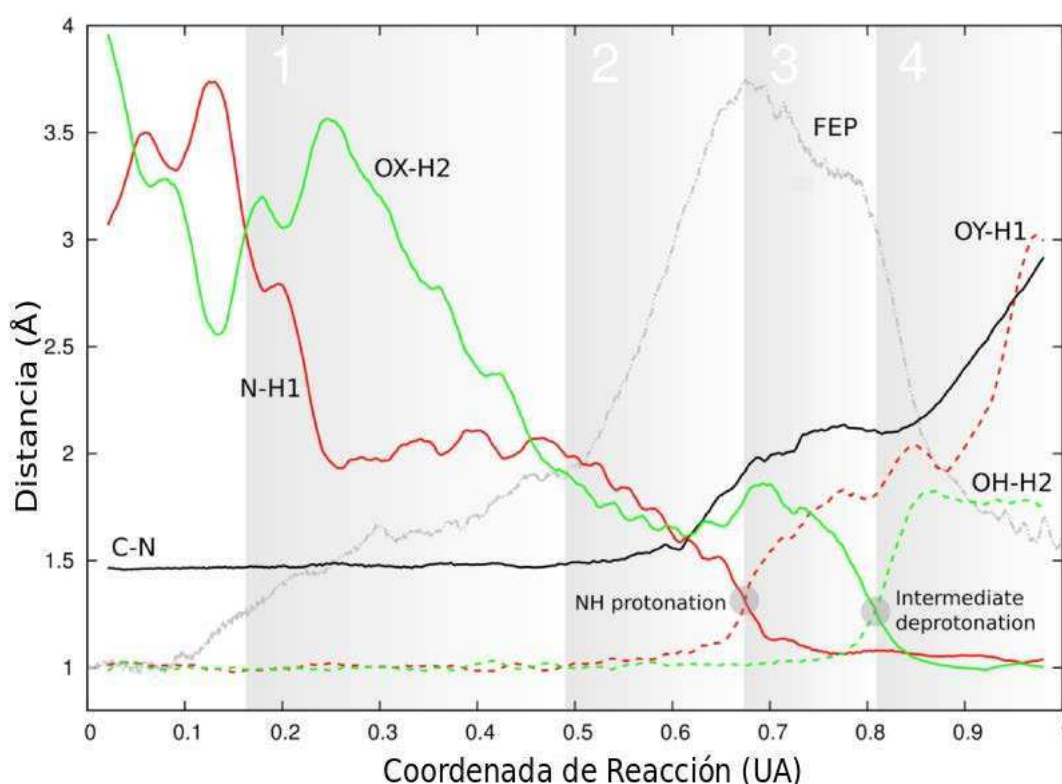


Figura 4.6: Evolución de los parámetros estructurales claves, a lo largo de la coordenada de reacción para MshB: Enlace C-N (curva negra), N-H1 (línea roja continua), OY-H1 (línea roja punteada), OX-H2 (línea verde continua), y OH-H2 (línea verde punteada). Se muestra además el perfil de energía libre obtenido en este paso (en gris).

ocurren varios eventos de manera sincronizada, indica que el mismo podría emplearse en una gran variedad de problemas, logrando una ventaja clara frente al esquema tradicional.

4.3 Discusión

Posteriormente al desarrollo de HyDRA, se buscó extender su aplicabilidad a sistemas diferentes al de la Cosismato Mutasa, los resultados obtenidos en este proceso fueron resumidos en este capítulo. Si bien el esquema HyDRA podría aplicarse a una infinidad de sistemas, escogimos aquellos que, a nuestro entender, se beneficiarían más de la ventaja presentada por la relajación temporal diferencial. Bajo este enfoque, el primer sistema que se estudió fue la reacción de oxidación de metanotiol por H_2O_2 en solución, con una carga neta significativa (-2) que debe ser reordenada a lo largo de la reacción, permitiendo además una transferencia de protón. Debido a esto, durante la reacción, las propiedades de solvatación del sistema reactivo se modifican significativamente, con lo que resulta de vital importancia el reordenamiento apropiado del entorno. Los resultados obtenidos (Tabla

4.1) confirmaron que permitir la relajación diferencial del subsistema clásico, resulta en una mejora significativa de la precisión del FEP obtenido, donde la desviación estándar asociada a las mediciones de los trabajos realizados, se ve reducida en incluso más de la mitad de su valor neto (al comparar los resultados obtenidos para DRAr 1 y 10). Este aumento notorio en la precisión del FEP obtenido para la reacción de oxidación de CH_3S^- por peróxido de hidrógeno, se condice con las observaciones realizadas para el sistema modelo de la reacción Corismato-Prefenato en solución, donde la diferencia en la energía libre obtenida con y sin HyDRA es notoria. Esto es de esperarse, debido a que, cuando la reacción de interés se encuentra en solución, el efecto del solvente es clave para permitir que la misma ocurra, por lo que la aplicación de HyDRA resulta siempre ventajosa. Sin embargo, las proteínas ofrecen, en muchos casos, un entorno más “rígido”, puesto a punto por la evolución, para llevar a cabo esa reacción particular, por lo que la ventaja del esquema HyDRA podría no siempre resultar en un beneficio neto.

Para estudiar más detalladamente el efecto de HyDRA sobre proteínas, se estudiaron dos enzimas de la familia de las Hidrolasas de Zn de *Mycobacterium tuberculosis*, MshB y MA-Amidasa. Los resultados obtenidos demuestran que, para ambas hidrolasas micobacterianas, existe un mecanismo de reacción general basado en un ataque nucleofílico de agua/hidróxido de dos pasos. Las características clave de este mecanismo son: la coordinación del Zn-hidróxido, gracias a una base conservada presente en el sitio activo; el ataque nucleofílico al carbonilo escindible para producir el intermediario tetraédrico metaestable (paso 1), cuya estabilidad es promovida por el ión Zn que estabiliza (y acepta) la carga negativa desarrollada en el oxígeno del ex-carbonilo; y la ruptura del enlace C-N promovido por la transferencia de protón de la base del sitio activo al nitrógeno de la amida. Este mecanismo, involucra un reacomodamiento constante de las cadenas laterales de los residuos involucrados, por lo que la ventaja del esquema de HyDRA es notoria. Para el caso de la MA-Amidasa la ventaja del método es especialmente llamativa, en particular en el primer paso de la reacción puede notarse una disminución de 10kcal/mol en la barrera de activación, al emplear DRAr=10. Semejante disminución en el valor de energía libre resulta particularmente prometedor para este caso debido a que la velocidad empleada es lenta, y disminuir más su valor resultaría en un incremento sustancial del tiempo de cómputo total. Además, debido a la histéresis observada, no es posible para este caso utilizar una combinación de las reacciones directa e inversa, lo que, como se ha comentado con anterioridad, produce una mejora en la precisión del resultado final. Es por esto que, para esta enzima particular, la aplicación del esquema HyDRA es de vital importancia para obtener un resultado preciso, especialmente para el ΔG de reacción, en un tiempo de cómputo accesible.

Habiendo determinado el mecanismo y obtenido los FEP correspondientes, es interesante también comparar nuestros resultados con los obtenidos por otras hidrolasas de Zn.

Aunque un estudio detallado de la reducción del costo computacional se presentó en el capítulo anterior, es interesante analizar el rendimiento del método HyDRA en la contexto de trabajos similares sobre enzimas de la misma familia. Posiblemente, el caso de la carboxipeptidasa-A (CPA) sea el más estudiado, donde en un estudio reciente [84], empleando un enfoque QM/MM basado en DFTB, los autores reportaron que para la hidrólisis de péptidos, el mecanismo de ataque nucleofílico agua/hidróxido coordinado con Zn es el más probable, llevando también a un intermediario tetraédrico, con un barrera de $17,9\text{kcal/mol}$, similar a la obtenida para las hidrolasas aquí estudiadas. Más interesante es que el mismo mecanismo también se informó para la termolisina [9], la enzima convertidora de angiotensina (ACE) [86] y la peptidasa del factor letal del ántrax [72], donde todas las barreras para el ataque nucleofílico muestran valores en el rango de $13 - 18\text{kcal/mol}$ y para la ruptura del enlace C-N en el orden de $5 - 14,7\text{kcal/mol}$.

Si bien es difícil realizar una comparación detallada cuando se emplearon diferentes códigos, niveles de teoría y sistemas, el número total de pasos QM/MM requeridos para obtener el perfil podría emplearse como un parámetro independiente del sistema y del nivel teórico/código. El número de pasos híbridos requeridos para obtener la FEP para la carboxipeptidasa A, termolisina y ACE fue de aproximadamente 2, 1,5 y 8 millones, respectivamente. Empleando el esquema HyDRA, con un valor de DRAr de 10, y utilizando sólo las reacciones en sentido directo, en este trabajo pudieron obtenerse los FEP en sólo 400.000 pasos QM/MM, que es aproximadamente cinco veces más rápido que el utilizado para los casos de CPA y termolisina, y 20 veces más rápido más rápido en comparación con ACE. Estos resultados ponen de relieve el tremendo poder y la ventaja obtenida con la utilización del esquema HyDRA.

4.4 Conclusiones

La aplicación del esquema HyDRA en sistemas más complejos, como lo son las hidrolasas de Zn y la reacción de oxidación de metanotiol por peróxido de hidrógeno, nos permitió estudiar con más detalle las ventajas obtenidas en la obtención de un perfil de energía libre de reacción, al emplear el método desarrollado. Los resultados obtenidos nos permiten concluir que el esquema HyDRA puede ser empleado en una gran variedad de sistemas, y que las ventajas, tanto en tiempo de cómputo como en precisión, pueden observarse en cualquier sistema que requiera de un reacomodamiento significativo del entorno para llevar a cabo una reacción. Estos resultados ponen de manifiesto el enorme potencial del esquema HyDRA para elucidar las propiedades físico-químicas subyacentes en reacciones en fases condensadas.

COORDENADAS DE REACCIÓN PARA CAMBIOS CONFORMACIONALES

Los cambios conformacionales se emplean dentro de los sistemas biológicos para modular la función de diversas macromoléculas. Una de las formas de regulación de enzimas más usuales es el alosterismo, donde la unión de una molécula en el sitio alostérico, modifica (a través de un cambio conformacional) las condiciones de unión de otra molécula en el sitio catalítico de la enzima, distante del primero. Sin embargo, la caracterización cuantitativa de la energética asociada a dichas interconversiones estructurales continua en la actualidad siendo un desafío, carente de un método unificado de estudio. Desarrollar herramientas que permitan comprender mejor este tipo de procesos, por métodos de simulación molecular, es la temática que abordaremos a lo largo de este capítulo.

5.1 La Problemática

Como se mencionó en los capítulos anteriores, la simulación molecular ha permitido el estudio de numerosos y diversos fenómenos, que van desde la reactividad química y la catálisis enzimática a los cambios conformacionales extendidos que involucran el movimiento de grandes regiones de la proteína. Sin embargo, las simulaciones clásicas atómicas, se encuentran todavía limitadas a simular los procesos que ocurren en escalas temporales de nano a microsegundos y para sistemas de tamaño moderado. Por esta razón, es usual que los sistemas proteicos de tamaño mediano o grande (aproximadamente, más de 500 residuos) se encuentren atrapados en mínimos locales, incapaces de explorar otras conformaciones, debido a las barreras energéticas que las separan de las mismas. En este contexto, las simu-

laciones de grano grueso (CG) surgen como una herramienta que permitiría aumentar el muestreo del espacio conformacional para este tipo de sistemas, así como también el tamaño de los mismos, debido a su costo computacional reducido.

En los últimos años, se ha realizado un esfuerzo extraordinario para desarrollar modelos CG que proporcionen una representación adecuada de la dinámica de proteínas en su estado nativo [21, 36, 54] y que permitan retener información sobre las posiciones de las cadenas laterales y de las interacciones entre los aminoácidos. Con el fin de realizar simulaciones CG de proteínas, se han desarrollado diversos campos de fuerza, como ser la extensión del campo de fuerza Martini para proteínas [54], el campo de fuerzas Sirah[21] o el modelo CG de proteínas desarrollado por Hills y colaboradores[36]. Este último será el utilizado en este capítulo y ha sido descrito en la Sección 2.3.2.

Si bien el desarrollo de los campos de fuerza de grano grueso permitió el estudio de un gran número de fenómenos, inaccesibles hasta el momento, es aún restrictivo frente a otros, principalmente debido a que este modelo teórico aún no cuenta con un desarrollo metodológico equiparable al de las metodologías tradicionales clásicas o cuánticas. Por esta razón, estudiar cambios conformacionales desde un punto de vista energético, continua siendo una tarea complicada en el mejor de los casos. Esto se debe, principalmente, a que este tipo de transiciones son energéticamente costosas (requiriendo el pasaje de una barrera) y por esta razón su frecuencia de ocurrencia, incluso en dinámicas de grano grueso, es muy baja, lo que no permite realizar un muestreo estadístico apropiado. Contar con una coordenada de reacción que permita realizar DMG para atravesar dichas barreras permitiría muestrear estos fenómenos de forma eficaz, con el fin de comprender con mayor claridad procesos como el alosterismo, que son de vital importancia en el camino a comprender la regulación proteica. En este capítulo se presentará una coordenada de reacción capaz de dirigir un movimiento conformacional, con la cual se realizará un análisis sobre una proteína representativa de un sistema alostérico, como lo es la hemoglobina humana.

5.2 La Desviación Cuadrática Media como coordenada de reacción

Para poder guiar un cambio conformacional mediante una fuerza externa -que actúa sobre la coordenada de reacción- es necesario que la misma pueda identificar los diferentes estados conformacionales de una misma proteína. Si bien esta condición es fácil de cumplir en una coordenada de pocos átomos, encontrar variables colectivas que presenten valores directamente relacionados con los estados conformacionales de una proteína no es una tarea sencilla. Diversos modelos han sido propuestos para lidiar con los movimientos con-

formacionales de las proteínas, la gran mayoría de ellos se encuentra basado en propiedades particulares de cada sistema, con lo cual su reproducción sobre otras proteínas de interés se vuelve dificultoso, cuando no inaccesible. La desviación cuadrática media, o RMSD (por sus siglas en inglés *Root Mean Square Deviation*), ha sido empleada en algunos sistemas de diferente naturaleza, en la mayoría de los casos en combinación con el esquema de *Umbrella Sampling*. Si bien en los casos mencionados ha sido posible un estudio del fenómeno de interés, en el mismo no pueden llegar a distinguirse posibles casos de histéresis. Para poder observar esto es necesario realizar una DMG en el sentido directo e inverso del cambio conformacional que se desea estudiar. Teniendo entonces como objetivo poder realizar dinámicas moleculares guiadas, que permitan a una proteína adoptar distintos estados conformacionales, se escogió el RMSD como coordenada de reacción para tal fin.

El potencial que se utilizará para guiar la transición tiene, como es habitual, forma armónica:

$$(5.1) \quad V_{CC}(\vec{X}) = \frac{1}{2}k[RMSD(\vec{X}, \vec{X}^{Ref}) - RMSD^0]^2$$

donde k es la constante de fuerza y $\vec{X}$ es el vector de posición para el sistema de interés en un determinado momento. En cambio, $\vec{X}^{Ref}$ son las coordenadas fijas correspondientes a la estructura de referencia. La misma es necesaria para poder emplear el RMSD como una coordenada de reacción, ya que esta permitirá realizar el cálculo de la variable en todos los pasos. Esta misma estructura de referencia se empleará también para calcular $RMSD^0$, que es un parámetro (Ecuación 5.1) que se corresponde con el mínimo del potencial, y es usualmente el valor de RMSD final que se desea alcanzar. De esta forma, la desviación cuadrática media en el potencial queda definida como:

$$(5.2) \quad RMSD(\vec{X}, \vec{X}^{Ref}) = \sqrt{\frac{1}{N} \sum_{i=1}^N [(x_i - x_i^R)^2 + (y_i - y_i^R)^2 + (z_i - z_i^R)^2]}$$

donde la suma se realiza sobre los átomos que se desea comparar entre ambas estructuras.

En otras palabras, el potencial o fuerza externa guía la conformación proteica acercándola a una estructura de referencia, utilizando como coordenada de reacción el valor de RMSD. Si bien es sabido que, en la dinámica de proteínas, los valores de RMSD fluctúan en gran medida y que existe una combinación infinita de coordenadas con las cuales se obtiene el mismo valor de RMSD, es posible reducir esta variabilidad con el potencial escogido. Esto se logra escogiendo apropiadamente la estructura de referencia en conjunción con el valor de $RMSD^0$. Una forma de garantizar la efectividad de esta coordenada de reacción, consiste en escoger como estructura de referencia la conformación final que se desea alcanzar. Si bien de esta forma se logra reducir la variabilidad de estructuras, el RMSD como coordenada

de reacción sigue siendo lo suficientemente laxo como para permitir al sistema explorar los diferentes caminos conformacionales, de forma de escoger aquel que lo conduzca, por el camino de mínima energía, a la estructura final.

Con el fin de poder realizar las pruebas pertinentes para poner a punto la coordenada de reacción, y teniendo en cuenta los tiempos de simulación atomística, el trabajo se realizó empleando modelos de grano grueso. Este modelo se presenta como una alternativa precisa para descubrir el camino por el cual transcurre un cambio conformacional, debido a que el mismo logra disminuir las barreras de activación presentes en la SEP, sin eliminar las de mayor significancia. Para poder realizar esto, la coordenada del RMSD fue insertada en el programa *pmemd* (del paquete AMBER [11]), en donde se encontraba funcionando una versión para realizar DM con el modelo CG de Voth.

5.2.1 Implementación en AMBER

Con el fin de poder probar la coordenada de reacción se implementó el potencial descrito en la Ecuación 5.1 dentro del programa *pmemd* del paquete AMBER. Primeramente describiremos el aspecto práctico de la implementación, es decir, los archivos y modificaciones necesarias para realizar DM con el RMSD como coordenada de reacción desde el punto de vista del usuario. Para esto es necesario generar 2 archivos nuevos, y modificar el archivo de entrada “*mdin*” a fin de que el mismo contenga las variables necesarias para aplicar del potencial. Dicha modificaciones se pueden observar en la Figura 5.1, marcadas en negrita, donde se representa un archivo de entrada simplificado para realizar DMG con el modelo de CG de Voth (“*cg_voth=1*”). La primera modificación que se puede observar es la

Figura 5.1: Ejemplo simplificado de un archivo de entrada para el programa PMEMD, dentro del paquete de simulación AMBER, para realizar una DMG (<i>cm_option=2</i>) de <i>Ins</i> (<i>dt = 4fs</i>) con el RMSD como coordenada de reacción, con el modelo de Voth de CG (<i>cg_voth=1</i>).	<pre> 1ns de DMG-CG &cntrl nstlim=25000, dt=0.004, [...] cg_voth=1, cmrst=1 / &cmrst cm_option=2, rmsdk= 150., xtarget=0.5 &end </pre>
--	---

inclusión de la variable “*cmrst*”, la cual es usada para activar el potencial de la Ecuación 5.1, lo que ocurre (como es usual en el paquete AMBER) cuando la misma asume el valor 1. Una vez activada esta opción, se leen y guardan las variables especificadas dentro del *namelist* “*cmrst*”. El mismo contiene información acerca de que tipo de DM se realizará, a través de la

variable “*cm_option*”, que asumirá el valor de 1 cuando sólo se agregue una restricción fija durante toda la dinámica, mientras que cuando valga 2 se realizará una DMG. Además de esta, se encuentran las variables relacionadas con el potencial armónico, donde “*rmsdk*” es siempre la constante de fuerza, mientras que “*xtarget*” es el mínimo del potencial durante toda la DM (cuando “*cm_option*=1”) o el valor final del mínimo del potencial aplicado en la DMG (“*cm_option*=2”).

Como se mencionó, será necesario incluir dos archivos nuevos para que la dinámica guiada por RMSD pueda realizarse. Uno de ellos debe contener las coordenadas de la estructura de referencia (en el formato de archivos de coordenadas de AMBER, *.crd* o *.rst7*), mientras que el otro contendrá la lista de átomos sobre los cuales el potencial será aplicado. Esta lista permite especificar la región de la proteína sobre la cual actúa el potencial guía. Al aplicar una lista de átomos sobre los cuales calcular el RMSD y aplicar la fuerza, se logra independizar el cambio conformacional de los movimientos restantes en la proteína.

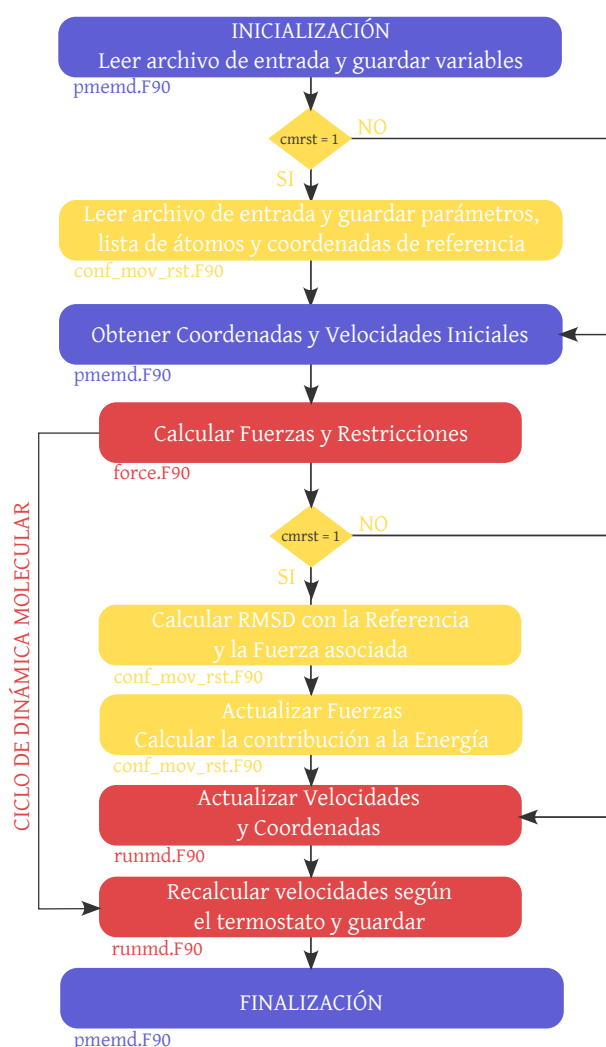


Figura 5.2: Diagrama de flujo que representa el esquema general de realización de una dinámica molecular serial dentro del paquete AMBER y el módulo PMEMD, donde los pasos han sido sobresimplificados para mayor claridad. En azul se representan los pasos de inicialización y finalización (que sólo se realizan una vez en la dinámica), mientras que en rojo se esquematizan los pasos de DM (que se realizan “*nstlim*” veces). En amarillo se incluyen las modificaciones implementadas para poder realizar la coordenada de reacción generalizada de RMSD, activada con la variable “*cmrst*”. Debajo de cada recuadro se nota además la subrutina en la que se encuentra cada uno de los pasos descritos, incluyendo la subrutina originada en este trabajo de tesis para utilizar al RMSD como coordenada para mover al sistema, “*conf_mov_rst.F90*”. Esta nueva subrutina ha sido programada como un módulo externo dentro de PMEMD, donde la única interacción con el código raíz es a través de los llamados a la misma, de manera de aumentar la portabilidad del nuevo código.

Con respecto a las modificaciones realizadas dentro del código *pmemd*, las mismas fueron

minimizadas para aumentar la portabilidad del código de RMSD. Para lograr esto se generó una subrutina independiente “conf_mov_rest.F90” que contiene toda la información necesaria para realizar la DM con el RMSD como coordenada de reacción. En el código raíz de *pmemd* sólo se incluyeron los llamados a dicha subrutina, los cuales ocurren cuando “cmrst=1”. La descripción general de la implementación se encuentra esquematizada en la Figura 5.2, donde los cambios realizados se encuentran en recuadros color amarillo. Las modificaciones pertinentes se realizaron de manera tal que el RMSD pudiera computarse como un *restraint* normal, para lo que fue necesario calcular la fuerza sobre cada átomo debido al nuevo potencial. Esto se traduce en una expresión del tipo:

$$(5.3) \quad F(x)_i = -\frac{k}{N} \frac{RMSD(X, X^{Ref}) - RMSD^0}{RMSD(X, X^{Ref})} [x_i - x_i^R]$$

donde sólo la última expresión entre corchetes es particular de cada coordenada (x, y, z) de cada átomo (i), mientras que el resto de la expresión es común para todos los átomos sobre los cuales actúa el potencial. El código desarrollado fue empleado para estudiar la transición R-T de la hemoglobina humana, con el modelo de grano grueso de Voth.

5.3 El Caso de la Hemoglobina

La hemoglobina humana (Hb) es una hemo proteína tetramérica, conformada por dos heterodímeros, cada uno formado por dos subunidades conocidas como α y β (véase Figura 5.3). Cada subunidad contiene a su vez un grupo hemo, que le otorga a esta proteína la capacidad de unir reversiblemente moléculas de oxígeno (una en cada grupo prostético). Cuando esto sucede, la Hb sufre una transición alostérica del estado T (“Tenso”) al R (“Relajado”). La hemoglobina constituye así un ejemplo importante de proteína alostérica, aunque no sea una enzima, sino una molécula transportadora de oxígeno. Como se mencionó en el capítulo 1, las proteínas alostéricas pueden cambiar su especificidad luego de la unión de un ligando. En el caso de la Hb, la unión de la primera molécula de oxígeno aumenta la afinidad de unión al segundo grupo hemo, la unión de la segunda aumenta la afinidad para la tercera, y así sucesivamente.

Por ser un caso emblemático dentro del alosterismo, la hemoglobina ha sido extensamente estudiada [10, 29, 39, 41, 61]. Sin embargo, debido a la importancia del grupo hemo en la misma y a la falta de parámetros para el mismo, no había sido estudiada previamente con modelos de grano grueso. La parametrización del grupo hemo fue realizada y probada durante el transcurso de esta tesis, por parte de otros miembros del grupo de investigación del cual la autora de esta tesis participa. A continuación se presentan los resultados asociados a este desarrollo, pertinentes a esta tesis.

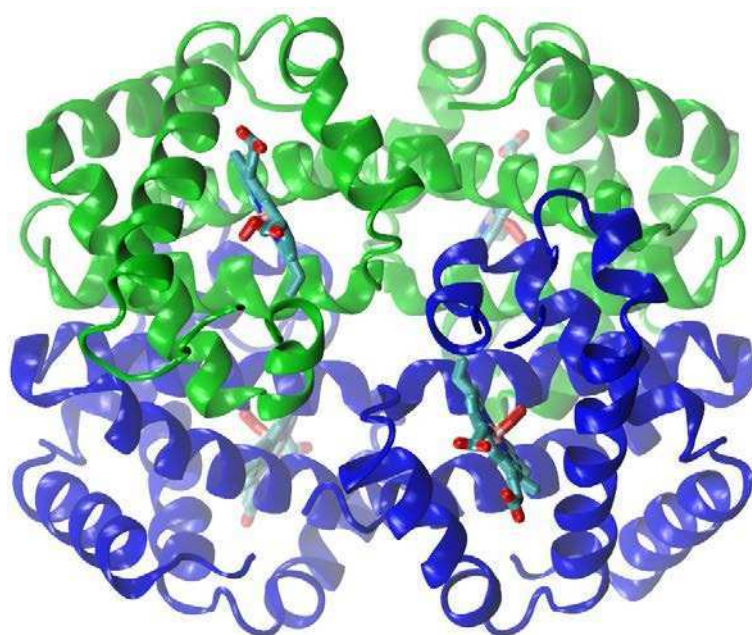


Figura 5.3: Representación de cintas de la proteína Hemoglobina en el estado R (unida a una molécula de oxígeno). Las subunidades α y β se representan en color verde y azul, respectivamente, mientras que los grupos hemo se encuentran representados con un código de colores que varía según el tipo de átomo (celeste para los carbonos, rojo para el oxígeno, azul para el nitrógeno y rosa para el hierro).

5.3.0.1 Condiciones de Simulación

Las simulaciones de grano grueso se realizaron en una versión localmente modificada del programa *pmemd* (para CPU) dentro del paquete de simulación AMBER, donde el campo de fuerzas CG desarrollado por Voth había sido implementado. Un paso de tiempo de $4fs$ fue empleado para las DM-CG, según lo recomendado en el trabajo original de Hills y colaboradores [36]. Los cristales que se emplearon para realizar las simulaciones corresponden a 2HHB [28] para el estado T y 1HHO para R [71]. Estas estructuras se convirtieron a la representación de grano grueso, y luego fueron minimizadas y lentamente termalizadas hasta alcanzar la temperatura deseada. Debido a que, en las simulaciones CG, se ha visto que la temperatura juega un papel crucial y que, para reproducir el comportamiento atomístico es necesario simular el sistema a $0,6T$, se realizaron simulaciones de prueba a las temperaturas de 100K, 125K, 150K y 200K. Para el sistema de la Hb, la temperatura se fijó en un valor de 125K, para evitar distorsiones en la estructura de la proteína, usuales debido a la capacidad de los modelos CG de disminuir las barreras de energía. Además de esto, y para evitar el posible movimiento del grupo hemo, se colocó una pequeña restricción de $10kcal/mol.\text{\AA}$ sobre las cuatro partículas polares de la Hb (posicionadas en la vecindad inmediata de los átomos de Fe, en la representación atomística).

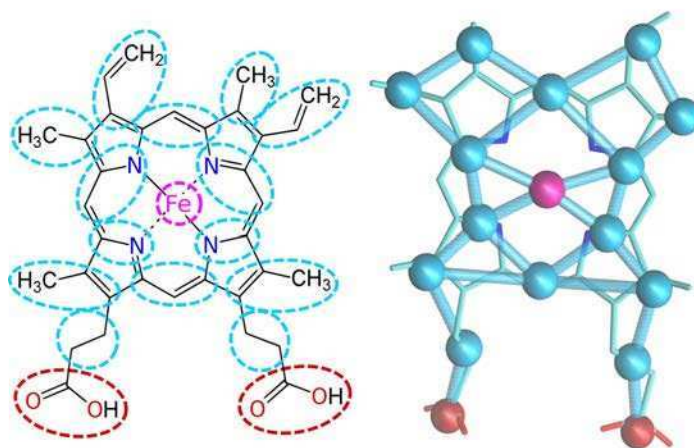
Para los propósitos comparativos, se realizaron $200ns$ de DM atomística en solvente explícito y en el ensamble NPT, empleando el campo de fuerzas Amberff99SB [51] y el programa *pmemd* del paquete AMBER. Las DMG-CG se realizaron en condiciones equivalentes a las empleadas durante las dinámicas de equilibrio, a excepción de que la transición fue realizada en $70ns$, lo que equivale a una velocidad teórica de $60\text{\AA}/ps$ (la coordenada de reacción varía

desde 5 a 0,5Å). Para realizar la transición, sólo se tuvieron en cuenta los carbonos- α del tetrámero (que representan el *backbone* en el modelo CG), por lo que el potencial externo fue aplicado exclusivamente sobre estos átomos.

5.3.1 Validación del Modelo CG del Hemo

Debido al poder de las metodologías de grano grueso, y a la relevancia biológica de las hemoproteínas, se llevó a cabo la parametrización del grupo Hemo, empleando un protocolo similar al empleado por Hills y colaboradores [36]. La elección por este modelo, frente a los disponibles en literatura, se basó en la experiencia previa del grupo en simulación molecular de hemo proteínas, lo que proporcionó los datos necesarios para realizar apropiadamente la parametrización. La geometría final del grupo hemo puede observarse en la Figura 5.4,

Figura 5.4: Modelo CG del grupo Hemo. A la izquierda se muestran los átomos que se agruparon para formar una partícula CG (fucsia, turquesa y rojo para partículas polares, no-polares y negativas, respectivamente), donde la posición está determinada por el centro de masa de cada grupo de átomos dentro del círculo de líneas punteadas. A la derecha se muestra la estructura final del grupo hemo, incluyendo enlaces.



donde se representan los modelos atomísticos y de grano grueso del mismo. Un total de 17 partículas (o granos) fue utilizado para representar al grupo hemo en el modelo CG de Voth, 1 partícula polar (fucsia), 14 no-polares (turquesa) y dos cargadas negativamente (rojo). La posición de cada una de las partículas se fijó en relación con la posición del centro de masas del grupo de átomos que esa partícula representaría (marcadas con un círculo de línea punteada en la Figura 5.4). Los enlaces se fijaron de forma tal que la geometría y rigidez del grupo hemo no se vieran afectados, y sus vibraciones fueran similares a las observadas en las DM atomísticas.

Debido a la importancia del grupo Hemo dentro de la Hb fue necesario validar la parametrización realizada para el mismo, dentro del modelo de Voth¹. Para poder analizar si el modelo CG es capaz de reproducir los diferentes estados conformacionales de la Hb, se

¹En el trabajo final publicado la validación se llevó a cabo, no sólo en Hb sino también en Mioglobina y Neuroglobina. Por una cuestión de relevancia con los resultados se muestran sólo los cálculos realizados en Hemoglobina.

realizó una DM-CG de equilibrio de $0,2\mu s$, comenzando desde las estructuras cristalinas (transformadas a CG) correspondientes de los estados R y T. En ambos casos las estructuras permanecieron estables, con respecto a la estructura inicial (Figura 5.6), manteniéndose cercanas a la conformación de partida y distantes de la conformación opuesta.

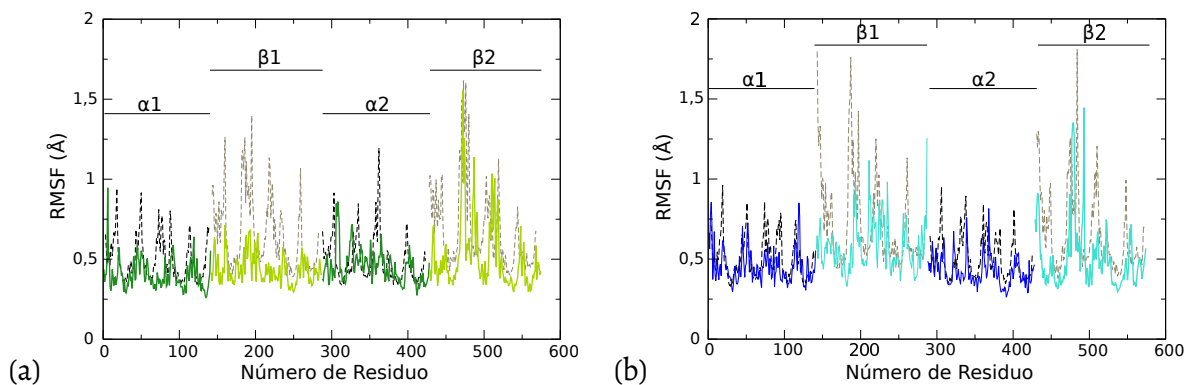


Figura 5.5: RMSF por residuo para las simulaciones CG (líneas sólidas) y AA (líneas punteadas). Los paneles (a) y (b) muestran los resultados obtenidos para R y T, respectivamente, donde se muestran además las diferentes subunidades α y β , diferenciadas por el subíndice acorde al dímero al que pertenecen.

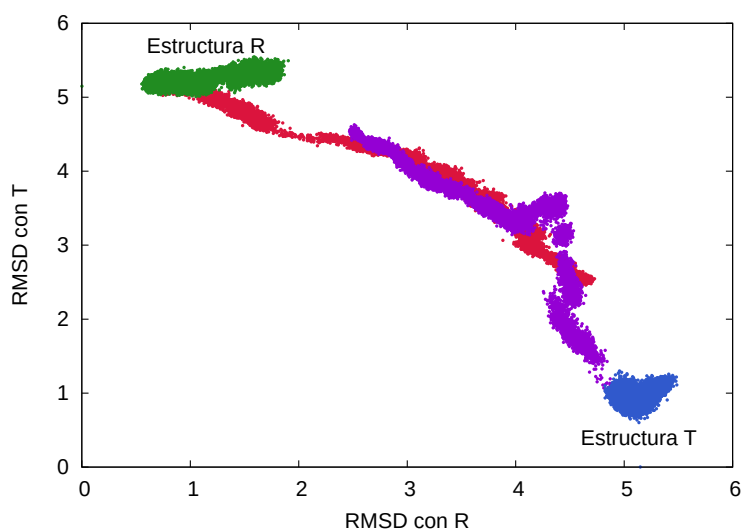
Adicionalmente, las trayectorias CG fueron comparadas con las dinámicas moleculares a nivel atómico, para ambos estados, R y T, de la hemoglobina humana. Para este fin se calculó el valor de RMSF para las 4 DM, y se superpusieron los resultados obtenidos para las dinámicas atómicas (líneas punteadas) y las de grano grueso (líneas continuas), como puede observarse en la Figura 5.5. Los valores de RMSF obtenidos para las DM-CG se encuentran en concordancia con las fluctuaciones observadas para las simulaciones AA, lo que es usual para estos casos. Interesantemente, las mayores fluctuaciones obtenidas en las subunidades atómicas se pueden observar también para en las DM-CG, con lo que se puede concluir que el modelo CG escogido puede describir la diferencia de flexibilidad entre las subunidades.

5.4 Resultados: Transición R-T en Hb

Habiendo validado el modelo CG para el caso de la Hb, en ambos estados, se procedió a estudiar la capacidad de la coordenada de reacción escogida (RMSD) para guiar el cambio conformacional R-T. Para esto, partiendo de las dinámicas convergidas empleadas en la sección anterior, se realizaron las DMG con el valor de RMSD como coordenada de reacción, empleando como referencia la estructura T cuando la DMG partía del estado R, y viceversa. Diversos parámetros de la transición fueron analizados a lo largo de la trayectoria de DMG,

de forma tal de caracterizar el camino escogido por el sistema en la dinámica CG, y compararlo con los resultados en literatura, obtenidos para dicha transición con metodologías atómicas[29, 39].

Figura 5.6: Valores de RMSD calculados sobre todos los átomos $C\alpha$ de la Hb. En el eje de abscisas se muestran los valores obtenidos cuando la alineación y el cálculo de RMSD se realizan sobre los $C\alpha$ de la estructura R, y en el eje de ordenadas, el equivalente para la estructura T. En verde se muestra la DM de Hb-R, y en azul la de Hb-T; mientras que en rojo se muestra la transición de R a T, y en lila la de T a R.



Como primer medida, es necesario evaluar la coordenada de reacción empleada. Con este fin se calculó, por un lado, el valor de RMSD tomando como referencia los C- α de la estructura R, y por otro, los C- α de la estructura T. Esto se realizó para las DM-CG de equilibrio -de manera de considerar las desviaciones normales de RMSD- y para las DMG-CG, en las cuales el sistema fue forzado a adquirir valores diferentes de RMSD. El potencial se aplica desde el valor de RMSD entre ambas estructuras iniciales ($5\text{\AA}$), y se pretende que alcance un valor mínimo de $0,5\text{\AA}$. Si bien este límite inferior es un valor extremo, se escogió este debido a que el RMSD es muy laxo como CR, y es necesario que la fuerza sea considerable para poder realizar el cambio conformacional. En la Figura 5.6 puede observarse el resultado obtenido para las diferentes simulaciones de grano grueso.

Lo primero que puede observarse en la Figura 5.6 es que, si bien existe una variación del valor de RMSD en las DM-CG de equilibrio, los valores que asume durante la DMG-CG se encuentran en un grado de mayor variabilidad y con una orientación diferente. Tomemos por ejemplo el caso de la DM para R (en color verde), que muestra una variabilidad de más de $1\text{\AA}$ con respecto a ella misma (eje de abscisas). Sin embargo, para la misma dinámica, cuando se calcula el RMSD tomando como referencia la estructura T (eje de ordenadas), la variabilidad es mínima. Si ahora intentamos conducir el estado R al T (DMG en rojo), podemos notar que la variación en el valor de RMSD es mucho más notoria, distinguiéndose

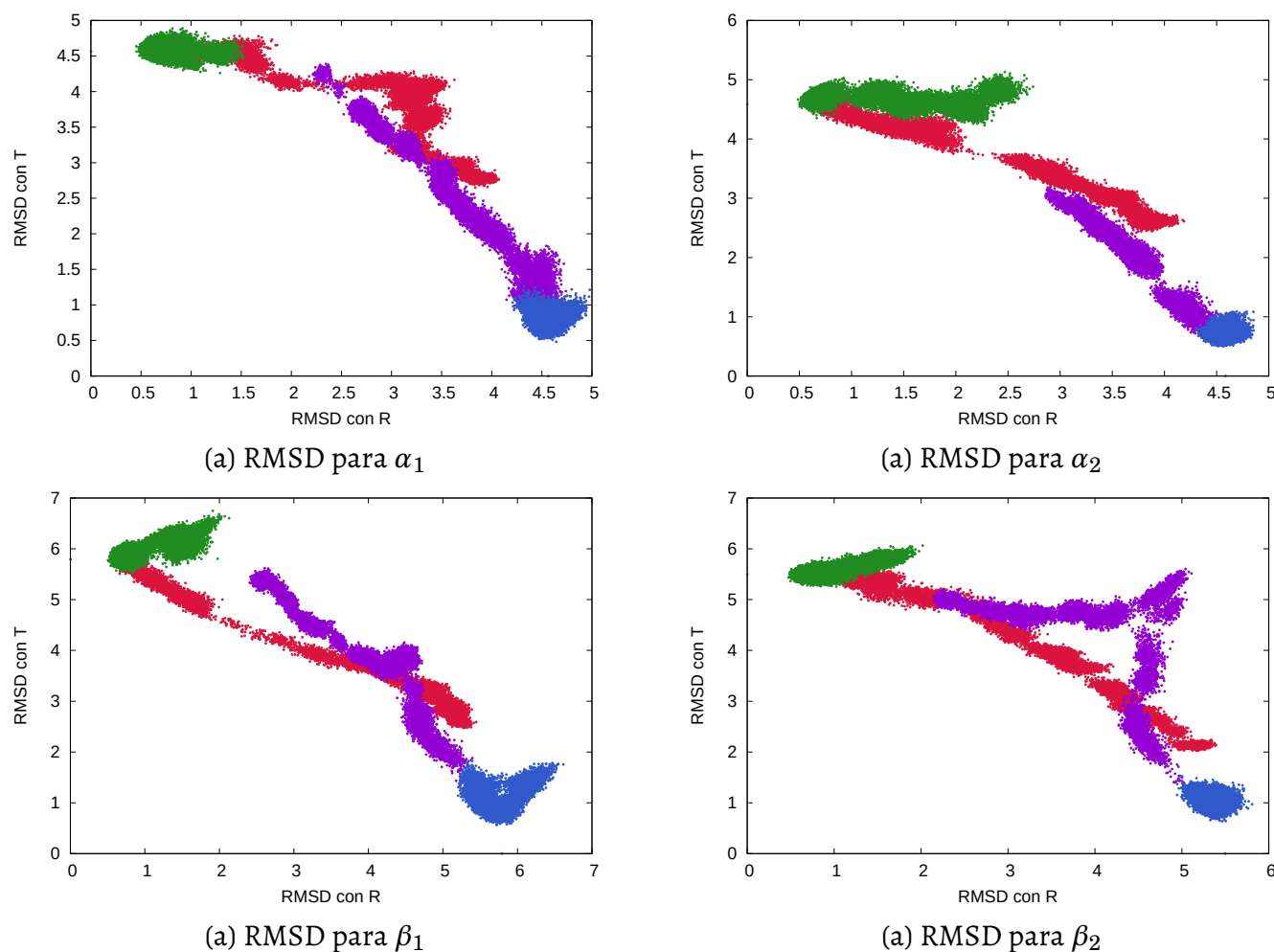


Figura 5.7: Valores de RMSD calculados sobre todos átomos $C\alpha$ de las subunidades individuales de la Hb, habiendo realizado el alineado sobre todos los $C\alpha$ de la proteína. En el eje de absicas se muestran los valores obtenidos cuando el cálculo de RMSD se realiza tomando como referencia la R, y en el eje de ordenadas, el equivalente para la estructura T. En verde se muestra la DM de Hb-R, y en azul la de Hb-T; mientras que en rojo se muestra la transición de R a T, y en lila la de T a R.

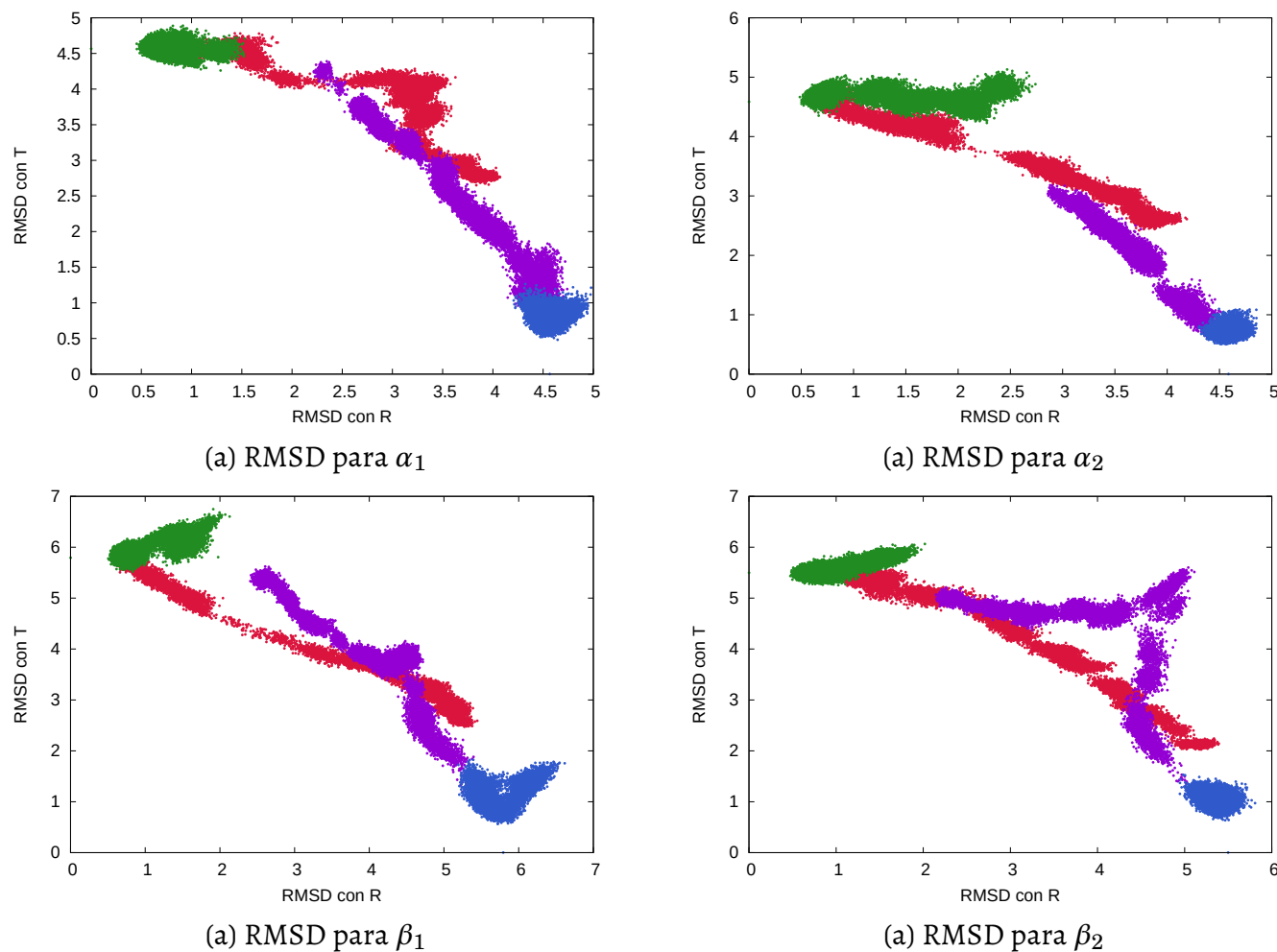


Figura 5.8: Valores de RMSD calculados sobre todos átomos $C\alpha$ de las subunidades individuales de la Hb, habiendo realizado el alineado sobre los mismos átomos. En el eje de abscisas se muestran los valores obtenidos cuando el cálculo de RMSD se realiza tomando como referencia la R, y en el eje de ordenadas, el equivalente para la estructura T. En verde se muestra la DM de Hb-R, y en azul la de Hb-T; mientras que en rojo se muestra la transición de R a T, y en lila la de T a R.

claramente del movimiento que la proteína realiza en equilibrio. Para la estructura T el caso es más notorio inclusive, debido a que la variabilidad presentada a lo largo de la DM de equilibrio es mucho menor, con lo que se vuelve incluso más evidente que la DMG (en lila para la transición T a R) logra conducir al sistema fuera de su estado de equilibrio y hacia el destino seleccionado.

Si bien la transición ocurre en el sentido correcto, es decir, alejándose de su estructura inicial y acercándose a la de referencia, ninguna de las DMG alcanza el valor de RMSD final propuesto en la CR. Esto se debe a que, debido a la forma que tiene la fuerza aplicada (Ecuación 5.3), la misma tiende rápidamente a cero a medida que la estructura se aproxima al valor de referencia, debido a que existen dos términos que disminuyen el valor de la fuerza total (la diferencia en el valor total de RMSD, y la diferencia entre las coordenadas específicas). A pesar de esto, combinando ambas trayectorias, puede observarse el camino de reacción completo. En el mismo ambas trayectorias se superponen, lo que da lugar a un camino único de reacción, salvo por una ligera discrepancia alrededor de los 4Å en el eje de abscisas. Para poder entender esta discrepancia se estudio el movimiento de cada subunidad por separado. Para esto se realizaron dos cálculos diferentes de RMSD, teniendo en cuenta sólo los C- α de cada subunidad, pero variando la alineación inicial. Primeramente se realizó un alineamiento sobre todos los C- α del tetrámero, seguido de un cálculo de RMSD para los C- α de cada subunidad (el resultado obtenido puede observarse en la Figura 5.7). En el mismo puede observarse que es el movimiento de las subunidades β , en las trayectorias de T a R, el que ocasiona la discrepancia entre ambas trayectorias. Sin embargo, puede observarse también una similitud en el mecanismo al observar las subunidades α , durante las trayectorias de R a T. A continuación, se realizó un nuevo cálculo de RMSD, realizando el alineamiento y el cálculo de RMSD únicamente sobre los C- α de cada subunidad, de manera de poder observar si esta discrepancia en las trayectorias se relaciona con el camino real de la transición o es un artefacto del cálculo del RMSD (Figura 5.8).

Para no limitarnos únicamente al estudio del valor de la CR escogida, consideramos también otras propiedades que se encuentran descriptas en literatura, para caracterizar la transición. Las mismas se detallan a continuación.

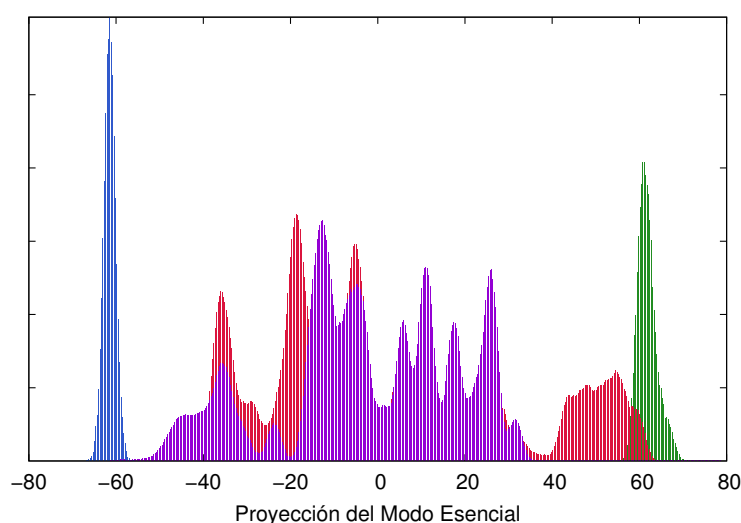
5.4.1 Modos Esenciales

Para poder comparar con mayor detalle el espacio conformacional explorado en las dinámicas de grano grueso, se realizó un análisis de los modos esenciales de las mismas. Los modos esenciales involucran la diagonalización de la matriz de covarianza de las posiciones atómicas, a lo largo de las trayectorias, dando como resultado autovectores que definen la dirección del movimiento esencial (ó dirección de mayor movimiento) de la proteína

durante la dinámica utilizada para la determinación de la matriz. Este análisis se realizó sobre los C- α únicamente (*backbone* en el modelo CG), excluyendo a los residuos terminales, debido a que la alta flexibilidad de éstos podría enmascarar el movimiento esencial del *core* de la proteína. Para esto se calculó inicialmente el modo de la transición, combinando las trayectorias de DM de equilibrio de R y T, y seleccionando el primer autovector. El mismo se asocia al concepto químico del modo vibracional responsable de convertir el complejo activado en producto, y por esto se considera que el primer modo esencial contiene esta información (cuando se combinan los estados inicial y final), ya que da como resultado la dirección de mayor variación entre dos estructuras. Considerando que dicho vector contiene la información de la transición, es posible emplearlo para medir el grado de avance del movimiento conformacional en las DMG. Por esta razón, se proyectó el modo de la transición sobre las DMG obtenidas, a fin de poder analizar el espacio configuracional explorado en las mismas y su similitud con el correspondiente a la transición. Luego se calculó la proyección de dicho modo sobre todas las DM-CG de equilibrio realizadas para la Hb, de forma tal de ubicar en el espacio conformacional los estados R y T. Los resultados obtenidos se encuentran representados en histogramas en la Figura 5.9.

Nuevamente puede observarse que es posible distinguir claramente ambos estados de la Hb (R en verde y T en azul), y que las DMG que representan la transición (rojo para R $\rightarrow$ T y lila para T $\rightarrow$ R) asumen valores intermedios semejantes entre sí (y que se superponen), pero diferenciados de las dinámicas de equilibrio. De forma similar a los resultados obtenidos para la Figura 5.6, será necesario combinar ambos resultados (R $\rightarrow$ T y T $\rightarrow$ R) para poder obtener la transición completa, y así describir todos los movimientos conformacionales intermedios.

Figura 5.9: Histogramas de la proyección del primer modo esencial de la transición R-T, sobre las trayectorias de DM-CG para R y T (en verde y azul, respectivamente) y sobre las transiciones realizadas con DMG de R a T (rojo) y de T a R (lila). Los valores de la proyección se realizaron en unidades arbitrarias.



5.4.2 Ángulo de Rotación

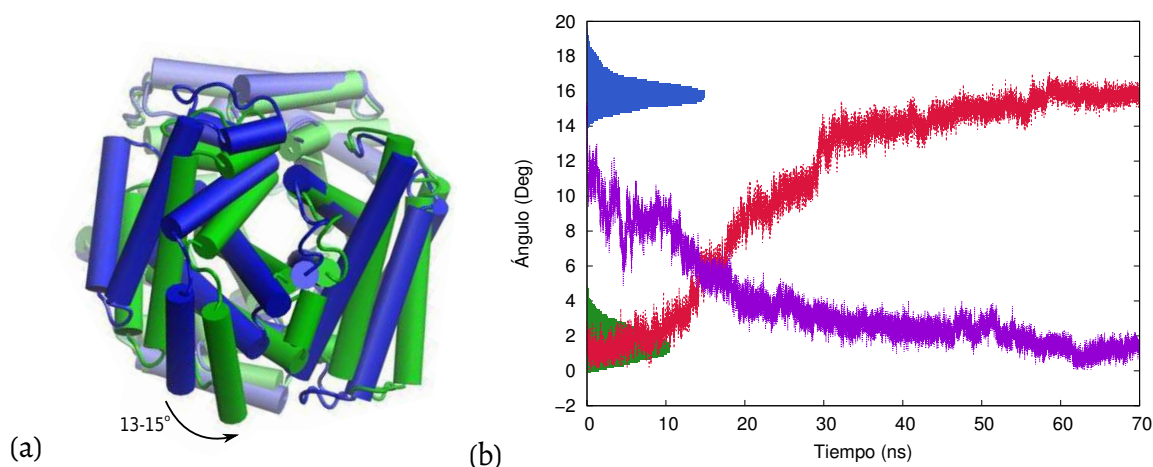


Figura 5.10: Rotación del dímero α_1/β_1 con respecto al dímero α_2/β_2 . (a) Estructuras superimpuestas de los estados R (verde) y T (azul) evidenciando el ángulo de rotación. (b) Evolución temporal del ángulo que representa la rotación, a lo largo de la transición R→T (rojo) y T→R (lila), donde también se representan los valores obtenidos para las DM de R y T.

Otro de los parámetros más empleados para caracterizar la transición cuaternaria R-T, es la rotación del dímero α_1/β_1 con respecto al dímero α_2/β_2 (Figura 5.10.a), que varía en un ángulo de 12-15° entre los dos estados conformacionales [29, 39]. Para poder realizar un análisis del comportamiento de dicho ángulo, se realizaron primeramente medidas del mismo en las dinámicas CG de equilibrio realizadas para las estructuras R y T, de manera de poder comprobar que existieran estados diferenciados y que el modelo de grano grueso pudiera reproducirlos. En la Figura 5.10.b se muestran estos datos en forma de histogramas, tomando como referencia el estado R², en verde los valores para R y en azul para el estado T. Se puede observar que, si bien existe una variación del ángulo dentro de cada dinámica, los valores se encuentran concentrados en una región del espacio. Esto permite diferenciar claramente los estados R y T, y evidenciar que, para el modelo propuesto, no ocurre una interconversión espontánea entre ambos estados en el tiempo simulado, lo que asume relevancia cuando se tiene en cuenta que la DM plana posee mayor duración que la DMG en la cual se observa la transición. Las transiciones R→T y T→R se encuentran también representadas en la Figura 5.10.b, en color rojo y lila respectivamente, donde queda en evidencia que durante ambas transiciones la DMG es capaz de guiar al sistema en la dirección correcta. Puede observarse además que, en el caso de la transición R→T, la variación del ángulo ocurre suavemente y que, a la mitad del tiempo de simulación, los dímeros ya se han establecido en

²Debido a que el ángulo de rotación, y los parámetros que analizaremos a continuación, son características relativas, es decir que su magnitud dependerá de la estructura con la cual lo estoy comparando, se toma el estado R como referencia de aquí en adelante, para simplificar el análisis.

lo que será su conformación final. Cuando se realiza el camino inverso (transición $T \rightarrow R$), el cambio es muy abrupto inicialmente, lo que indica que esta rotación inicial (de aproximadamente 4°) no es energéticamente costosa. Ambas observaciones, en conjunto con los resultados obtenidos para el valor de RMSD y los modos esenciales, parecerían indicar que la rotación de los dímeros es uno de los primeros movimientos conformacionales requeridos para la transición, mientras que el reacomodamiento interno de los residuos ocurre al final de la misma. En este caso, la transición entre ambos estados se observa en su totalidad, sin necesidad de combinar las trayectorias de DMG.

5.4.3 Switch de la Histidina

Los resultados hasta aquí presentados nos permiten afirmar que, tanto el modelo CG como la coordenada de reacción, son capaces de reproducir los parámetros estructurales globales de la transición entre los estados R y T. Sin embargo, es también importante conocer la capacidad predictiva del modelo propuesto sobre los residuos individuales y sus interacciones, teniendo en cuenta que la coordenada de reacción escogida podría, en principio, enmascarar movimientos sutiles de los aminoácidos dentro de la proteína (como se discutió en la sección 5.2), y que los modelos de grano grueso simplifican las interacciones entre residuos. Una característica muy representativa de la interacción entre residuos, en la interfaz de las subunidades, es conocida como *switch* de la His97 [29]. En la misma se considera la posición de la cadena lateral de la β His97, situada en el loop entre las hélices F y G, la cual varía en las estructuras R y T. Para la proteína en el estado relajado, la β His97 se sitúa entre las Thr38 y Thr41 de la hélice C de la subunidad alfa, mientras que en el estado tenso lo hace entre la Thr41 y la Pro44 de la misma hélice (α -C). Este movimiento se encuentra representado en la Figura 5.4.3.a, donde, para mayor claridad, se superimponen las interfaz de interés para las estructuras R y T (verde y azul respectivamente) a nivel atómico a la izquierda (estructuras cristalinas) y de grano grueso a la derecha.

Para poder diferenciar cuantitativamente ambos estados durante las dinámicas, se considera la distancia entre los carbonos alfa de los residuos Thr38 e His97, $dT38C\alpha-H97C\alpha$, la cual se encuentra indicada con líneas punteadas en la Figura 5.4.3.a. Esta distancia fue medida en las DM de equilibrio realizadas para los estados R y T, y para las DMG de la transición entre ambos estados, los resultados se volcaron en la Figura 5.4.3.b. Los mismos indican, en concordancia con las demás propiedades analizadas anteriormente, que, mientras en cada estado se conserva la posición del *switch* durante toda la DM, en la transición entre los estados R y T, la simulación realizada con DMG es capaz de inducir el cambio en la posición de la His97.

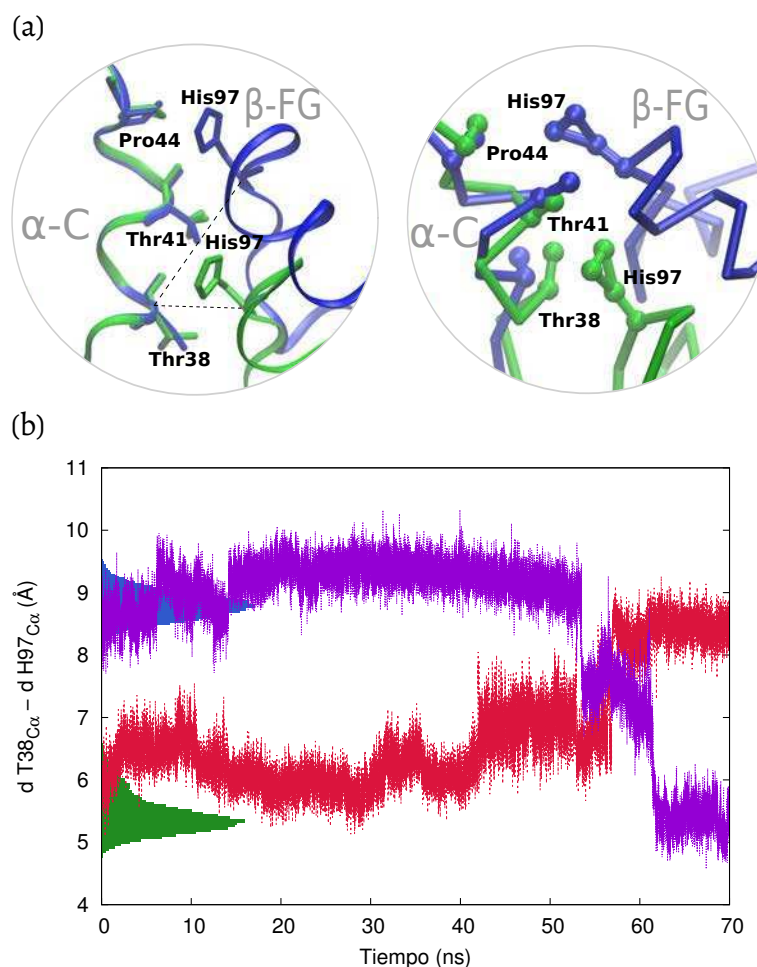


Figura 5.11: Interfaz del *switch* de la Hb. (a) Representación gráfica atomística (izquierda) y de grano grueso (derecha) para el estado R (verde) -con el residuo β_1 His97 ubicándose entre α_2 Thr38 y α_2 Thr41- y para el estado T (azul) -con la β_1 His97 ubicándose entre los residuos α_2 Thr38 y α_2 Thr41- en la interfaz del *switch*. (b) Evolución de la distancia α_2 Thr38C α - β_1 His97C α a lo largo de la transición R→T (rojo) y T→R (lila), donde también se representan los valores de dicha distancia para las DM de R y T.

5.5 Discusión

A lo largo del capítulo se presentó la implementación de una coordenada de reacción que pudiera guiar cambios conformacionales en proteínas. La validación se realizó empleando el recientemente desarrollado modelo de grano grueso para el grupo Hemo, para estudiar la transición R-T en la Hemoglobina humana. Existen trabajos previos en bibliografía [29, 39] que logran caracterizar dicha transición utilizando modelos atómicos, en los cuales nos basaremos para realizar nuestro análisis comparativo. Si bien es necesario mencionar que en ambos trabajos sólo se estudió la transición T→R, por lo que el abordaje de la posible histéresis en el sistema será abordado basándonos exclusivamente en nuestros resultados.

En primer lugar analizaremos la variación de la estructura cuaternaria, tomando como referencia el trabajo realizado por Fischer y colaboradores [29]. Puede observarse en la Figura 5.6 que el movimiento cuaternario pareciera transitar por un único camino de reacción, independientemente del sentido en el cual se simule el cambio conformacional. A pesar de esto, el movimiento cuaternario por subunidad muestra una clara diferencia respecto al

comportamiento de las subunidades α y β . Esta variación puede observarse con mayor claridad en la Figura 5.7, donde el camino T $\rightarrow$ R (el reportado en bibliografía) seguido por las subunidades β se basa en diferenciarse inicialmente de la estructura inicial, sin por esto presentar similitud con R, para luego realizar un “salto” en el cual la similitud con T se mantiene y se aumenta la cercanía con la estructura R. En cambio, las subunidades α muestran un movimiento sin grandes variaciones. Ambos resultados se encuentran en concordancia con lo descrito por Fischer en 2011. Los valores de RMSD obtenidos para la transición R $\rightarrow$ T muestran un patrón similar pero invertido, donde las subunidades β transicionan lentamente al estado final, mientras que las α lo hacen más abruptamente. Un análisis de la variación temporal del RMSD en las DMG pareciera indicar que existe una tendencia de las subunidades α a permanecer en el estado T, mientras que las β parecieran tender al estado R, independientemente del sentido de la transición. Esto puede observarse por los valores de RMSD obtenidos por cada subunidad en las diferentes transiciones, que se condicen también con lo observado por Hub en 2010 [39].

Combinar ésta con la información obtenida para las demás propiedades medidas puede ayudarnos a elucidar el comportamiento general del sistema durante las transiciones R-T (y viceversa). Como se mencionó anteriormente, la variación del ángulo entre subunidades difiere también dependiendo del sentido de la reacción, siendo inicialmente abrupta para el caso T $\rightarrow$ R, mientras que, para el movimiento en la dirección contraria, el sistema permanece en la misma conformación los primeros 10 ns de simulación. Esto se condice con los resultados publicados por Fischer para el movimiento T $\rightarrow$ R, donde postulan dos movimientos cuaternarios secuenciales diferenciados. El segundo movimiento cuaternario viene acompañado por el reacomodamiento de la His97 en la interfaz del *switch*, lo cual puede ocurrir hacia el final de ambas DMG. El mismo se encuentra representado en la Figura 5.4.3, donde pueden distinguirse dos saltos abruptos, en el camino T $\rightarrow$ R, que podrían relacionarse con la secuencia movimiento cuaternario-reacomodamiento terciario, descrito por Fischer *et al.* Para el proceso R $\rightarrow$ T se observa un comportamiento similar, aunque sólo se distingue una variación abrupta, alrededor de los 55 ns. La combinación de estos resultados pareciera indicar que existe evidencia para sostener que el movimiento cuaternario no transcurre por un único camino, haciéndose más evidente este fenómeno a nivel de la estructura terciaria, donde se observan claramente fenómenos de histéresis durante las transiciones. Una generalización del mecanismo parecería indicar que existen dos movimientos cuaternarios, el primero de los cuales se produce variando el ángulo de rotación entre los dímeros, seguido de un acomodamiento marcado de las subunidades β (α) cuando el sentido de la transición es T $\rightarrow$ R (R $\rightarrow$ T), produciendo a continuación el movimiento de las subunidades contrarias, lo que permite un segundo movimiento cuaternario que se condice con el reacomodamiento interno de las subunidades en la interfaz (revelado por el *switch* de la His97).

El trabajo presentado demuestra que es posible emplear este modelo para, no sólo distinguir claramente los dos estados, sino que también describir apropiadamente el movimiento conformacional de dicha proteína. La comparación realizada con las dinámicas atomísticas, indican que las DM-CG fueron capaces de explorar el espacio conformacional, incluso cuando el sistema fue reducido en un tercio de su tamaño original (esto sin tener en cuenta las moléculas de agua) al transformarlo a un modelo de grano grueso.

A pesar de la pérdida del detalle atomístico, los procesos que implican reordenamientos locales o globales pueden describirse adecuadamente desde un punto de vista estructural y dinámico, con una reducción significativa del costo computacional, cuando se emplea el modelo propuesto. Los resultados presentados para la transición R-T de la Hb muestran que es posible explorar el cambio conformacional asociado con un esquema de muestreo sesgado. Por lo tanto, en principio, sería posible obtener los correspondientes perfiles de energía libre, y por lo tanto un enlace directo con los datos experimentales, como ser constantes cinéticas y de equilibrio. Sin embargo, es importante notar que la exactitud, y por lo tanto la relevancia de la comparación, dependerá de la capacidad del campo de fuerza CG para describir correctamente la superficie de energía libre del sistema de interés.

Gracias a la portabilidad del código desarrollado, es posible también aplicar el potencial externo asociado al RMSD en sistemas atomísticos a su vez. Sin embargo, en estos casos, si bien la energética será correcta, será necesario considerar el incremento sustancial en el tiempo de cálculo resultante de considerar todos los átomos del sistema.

5.6 Conclusiones

En este trabajo hemos desarrollado y probado una coordenada de reacción capaz de dirigir cambios conformacionales en proteínas, en el contexto de modelos de grano grueso. En conjunto, los resultados obtenidos en la Sección 5.4, demuestran que, empleando el valor de RMSD como coordenada de reacción, es posible estudiar la transición R-T de la hemoglobina humana, obteniendo resultados similares a los que se obtendrían por métodos atomísticos tradicionales. Pudiendo predecir correctamente, no sólo los movimientos cuaternarios globales, sino también los rearrreglos de la estructura terciaria, incluso al nivel de las posiciones relativas de las cadenas laterales.

Demostramos que este enfoque puede aplicarse al estudio de diferentes cambios conformacionales en proteínas, conservando la información estructural y dinámica clave que rige los procesos. El caso aquí estudiado es uno de los más estudiados para proteínas alostéricas, su correcta descripción abre la posibilidad de estudiar nuevos procesos complejos mediante este enfoque.

CONCLUSIONES Y PERSPECTIVAS FUTURAS

A lo largo de esta tesis he recorrido un largo camino de aprendizaje, adentrándome en el mundo de la química y la física. Mediante esto, hemos logrado abordar temáticas relacionadas con la energética de fenómenos biológicos, y hemos presentado metodologías computacionales que buscan reducir el tiempo de cómputo asociado a dichos fenómenos complejos. Esto nos permitió poder simularlos de manera eficiente, y así comprender con mayor profundidad la físico-química subyacente a los mismos.

6.1 Conclusiones Generales

Desde la introducción de la igualdad de Jarzynski en 1997 en el mundo académico, muchos estudios teóricos [17, 18, 58] han aprovechado su simplicidad y fácil paralelización para obtener perfiles de energía libre de diferentes procesos a nivel molecular. Sin embargo, la presunta reducción en el costo computacional, en comparación con otros esquemas, como el de *Umbrella Sampling*, y la precisión del FEP obtenido, son más una cuestión de fe que un hecho. Esto es especialmente notorio en simulaciones de reacciones enzimáticas (o en solución), donde que la complejidad de los sistemas por lo general impide la aplicación de diferentes esquemas con fines comparativos, y donde, con frecuencia, no existen valores experimentales de precisión con los cuales efectuar una comparación directa de los resultados obtenidos.

6.1.1 Esquema HyDRA

El esquema HyDRA presentado en esta tesis proporciona un método que tiene la potencialidad de generar perfiles de energía libres más exactos y precisos a un costo computacional significativamente menor. Además de esto, el análisis presentando resalta algunos datos interesantes sobre la relación entre la velocidad, el número de trayectorias y la precisión de los FEP resultantes, en el contexto de MDMG e IJ. Los datos actuales, se suman a la experiencia del grupo en el tema, y muestran claramente que es muy difícil obtener un FEP preciso cuando la velocidad empleada es demasiado rápida. Más importante aún, la velocidad excesivamente rápida no puede ser corregida realizando más trayectorias (a menos que se consideren números enormes) y, generalmente, un conjunto pequeño -en el orden de diez- es suficiente para obtener una buena estimación de la distribución del trabajo. Por lo tanto, al realizar estimaciones de un perfil de energía libre empleando la estrategia MDMG en combinación con la ecuación de Jarzynski, la distribución del valor de trabajo debe mostrar desviaciones en el orden de $k_B T$ para obtener resultados significativos. HyDRA permite, mediante la relajación diferencial del entorno clásico, que el sistema completo permanezca más cercano al equilibrio, incluso cuando el sistema QM es perturbado mediante el potencial externo -en los pasos conjuntos QM/MM- obteniéndose así distribuciones de trabajo más estrechas y, de esta forma, perfiles de energía libre más precisos.

Más allá de las ventajas del método HyDRA, también es interesante discutir otros esquemas de muestreo basados en la ecuación de Jarzynski, que permiten una convergencia más rápida del FEP (sin afectar por esto su exactitud), al reducir la desviación estándar entre las trayectorias de trabajo, que podrían aplicarse a esquemas híbridos QM/MM. Un ejemplo de este tipo de esquemas fue introducido por el grupo de Amzel [24], donde la reacción es dividida en segmentos, donde en cada uno de los mismos se realiza una DMG. La particularidad radica en que, luego de la producción en un segmento, se lleva a cabo un DM de equilibrio para permitir que el sistema se relaje y, a continuación, con el nuevo conjunto de coordenadas y velocidades, se realiza la DMG correspondiente al siguiente segmento. Cada segmento del perfil de trabajo es luego unido en múltiples combinaciones, generando así muchos más perfiles de trabajo completos para la reacción, que luego serán utilizados para calcular un FEP. De esta forma se logra conseguir un muestreo mayor de la SEP, con un mismo número de trayectorias realizadas. Otro enfoque de este tipo es conocido como *Adaptive Steered Molecular Dynamics* (ASMD), y fue desarrollado por Ozer y colaboradores [58]. En el mismo, también se divide la coordenada de reacción en segmentos, pero en lugar de realizar dinámicas de equilibrio para todas las trayectorias, que incrementan el costo computacional total, se emplean las coordenadas correspondientes al perfil de menor trabajo para comenzar el segmento siguiente. De esta forma sólo se realiza una dinámica de equilibrio por segmento

(la que se realizará a partir de las coordenadas de la configuración que haya alcanzado el final del segmento con el menor trabajo). El FEP final se obtiene uniendo los perfiles de energía libre parciales obtenidos para cada segmento usando la IJ. Hasta el momento, y hasta donde sabemos, ambas estrategias se han utilizado en simulaciones clásicas para estudiar procesos biológicos, tales como plegamiento de péptidos pequeños o unión a ligandos, y ninguno de ellos fue probado en un esquema QM/MM para estudiar o reacciones enzimáticas. En este contexto, podría ser interesante abordar la pregunta de qué estrategia es la más rentable para calcular reacciones en entornos complejos, cuando para esto se emplean esquemas híbridos QM/MM. Adicionalmente, podrían combinarse los esquemas de HyDRA y ASMD, ya que los mismos toman ventaja de diferentes condiciones de la igualdad de Jarzynski, lo que podría resultar en incluso una disminución mayor del costo computacional total.

Luego del desarrollo y validación del método, en un sistema muy estudiado -como lo es la reacción de Corismato Mutasa- el mismo fue aplicado para estudiar la reacción de dos hidrolasas de Zn esenciales de *Mycobacterium tuberculosis*, a saber, MshB (Rv1170) y MA-amidasa (Rv3717). Ambas enzimas hidrolizan el difícil enlace C-N de una amida. Los resultados de ese trabajo permitieron la obtención de perfiles de energía libre precisos para ambas etapas de la reacción, la primera involucrando el ataque nucleofílico coordinado de hidróxido de Zn al carbonilo escindible de la amida, que conduce al intermediario tetraédrico meta-estable. La segunda etapa, que implica la ruptura del enlace C-N y que es promovida por la transferencia de un protón, desde la base del sitio activo, hacia el N amídico. Las barreras y el mecanismo obtenidos fueron similares a los reportados para otras hidrolasas de Zn, como la carboxipeptidasa-A [84], termolisina [9], la enzima convertidora de angiotensina (ACE) [86] y la peptidasa del factor letal del ántrax [72], apuntando hacia un mecanismo conservado (o convergente) para este tipo de enzimas, pertenecientes a diferentes familias de proteínas. Más notorio aún, es que se realizó un análisis de los pasos híbridos QM/MM requeridos para obtener el FEP en cada caso, el cual demostró que la aplicación del esquema HyDRA resultó en una reducción del tiempo computacional de 5 a 20 veces, resaltando el poder de la estrategia HyDRA.

Los resultados obtenidos, presentados en este trabajo de tesis, nos permiten concluir que el esquema HyDRA puede ser empleado en una gran variedad de sistemas. Las ventajas observadas, en sistemas que requieren de un reacomodamiento significativo del entorno para llevar a cabo una reacción, son notorias, tanto en el tiempo de cómputo total, como en la precisión del perfil de energía libre logrado. Por esta razón, el esquema HyDRA se presenta como una herramienta poderosa dentro del campo, permitiendo el acceso al estudio de propiedades físico-químicas relacionadas con reacciones en fases condensadas complejas.

6.1.2 Coordinada de Reacción para Cambios Conformacionales

El segundo desarrollo presentado en esta tesis se encuentra relacionado con la dificultad de simular cambios conformacionales proteicos. Para esto, se introdujo una coordenada de reacción capaz de guiar movimientos conformacionales en una proteína, lo que implica mover coordinadamente un gran número de átomos. La coordenada de reacción escogida, la Desviación Cuadrática Media (RMSD), fue exitosamente implementada en el programa *pmemd* del paquete de simulación AMBER, y fue probada mediante el estudio de la transición R-T en la Hemoglobina humana, en un modelo simplificado de grano grueso. La transición se estudió realizando DMG en ambos sentidos de la reacción, de forma tal de poder estudiar posibles efectos de histéresis, los cuales no podría observarse empleando esquemas de muestreo sesgado como *Umbrella Sampling*. Los resultados obtenidos para la transición T→R pudieron ser comparados con simulaciones a nivel atómico, realizadas por otros autores [29, 39], demostrando el poder de predicción de la técnica empleada al ser capaz de reproducir, no sólo los movimientos cuaternarios de la transición, sino también las variaciones que ocurren al nivel de las subunidades. Este resultado es especialmente notorio, debido a que la transición entre ambos estados es sutil (sólo 5Å de diferencia en RMSD) y a que las interacciones se encuentran simplificadas dentro de los modelos de grano grueso, por lo que la capacidad del modelo de predecir cambios estructurales a nivel de cadenas laterales, como es el caso del *switch* de la His97 (Figura 5.4.3), es especialmente notorio. Los resultados obtenidos para la transición R→T, nos permitieron realizar inferencias acerca de los diferentes caminos escogidos por la proteína para realizar el cambio conformacional. De esta forma puede observarse que, si bien existe una conservación del camino a nivel de la estructura cuaternaria (Figura 5.6), el comportamiento de las subunidades (Figuras 5.7, 5.8, 5.4.3 y 5.10), en cambio, presenta claras diferencias en las dinámicas guiadas de grano grueso, pudiendo indicar histéresis al nivel de estructura terciaria. En conjunto, los resultados obtenidos permiten confirmar que, empleando el esquema utilizado, es posible predecir el comportamiento cuaternario y terciario de la Hemoglobina humana.

Estas conclusiones nos permiten suponer que el enfoque DMG-CG con RMSD como coordenada de reacción, podría aplicarse en el estudio de cambios conformacionales para distintos sistemas proteicos, describiendo apropiadamente la información estructural y dinámica que rige los procesos, en un tiempo computacional reducido, permitiendo incluso identificar fenómenos de histéresis. A pesar del prometedor resultado obtenido, es necesario tener en cuenta que el valor de RMSD puede no ser apropiado para describir el movimiento conformacional de algunas proteínas, por lo que la coordenada de reacción presentada podría no ser capaz de guiar el cambio conformacional. Algunas alternativas a esto, así como también la posibilidad de calcular perfiles de energía libre empleando este tipo de esquemas,

se discutirán en las perspectivas futuras.

En conjunto, el esquema de HyDRA y el RMSD como coordenada de reacción, pueden ser utilizadas para estudiar procesos proteicos de gran relevancia, en un tiempo computacional menor al que ofrecen otras alternativas en el campo. Debido a que ninguno de los dos esquemas se encuentra condicionado a una representación particular, sería posible utilizarlos incluso también en otros marcos teóricos. Por ejemplo, el esquema de HyDRA podría aplicarse en otras metodologías híbridas (como ser AA-CG), debido a que el mismo explota la potencialidad de la diferencia en el costo computacional de dos niveles teóricos diferentes. Inclusive, las mismas podrían combinarse para tratar diferentes sistemas, donde el cambio conformacional precede a una reacción, o donde el reacomodamiento del entorno es crucial para la correcta transición entre dos estados conformacionales, haciendo, de esta forma, posible su aplicación en una gran variedad de sistemas y esquemas de simulación.

6.2 Perspectivas Futuras

A pesar del esfuerzo realizado, han quedado proyectos pendientes que, debido al tiempo acotado de trabajo, no han podido finalizarse al momento de redactar este trabajo. Gracias al mayor entendimiento de la termodinámica asociada a los fenómenos de interés, y de las implementaciones prácticas de la teoría sobre el código fuente del programa escogido, fueron surgiendo, de manera paralela a los proyectos aquí presentados, nuevas preguntas e hipótesis, que dieron lugar a nuevos desarrollos, los cuales se encuentran actualmente en proceso. Los mismos se presentan, discriminados por temática, a continuación.

Respecto al esquema HyDRA, actualmente nos encontramos realizando los manuales y tutoriales necesarios su implementación en el *release* de AMBER17. Esto permitiría el acceso del esquema a una infinidad de usuarios, que de otra forma necesitarían realizar las modificaciones pertinentes en su propio código. Es por esta razón que este es uno de los proyectos principales en los que me encuentro trabajando actualmente, a fin de que el esquema de HyDRA pueda ser implementado por todos los usuarios de AMBER que así lo deseen.

En relación al trabajo realizado para ser capaces de guiar cambios conformacionales nos quedan aún, hipótesis, desarrollos y demostraciones a futuro. En este sentido, se presentan a continuación alguno de los temas que están siendo abordados en la actualidad por la autora de esta Tesis y el grupo de trabajo:

- **Paralelizar el código asociado al cálculo de RMSD como coordenada de reacción, dentro del paquete AMBER.** Los sistemas sobre los cuales se podría aplicar esta CR son de un tamaño tal que, realizar los cálculos en un único procesador, incrementaría el

costo computacional de una forma abismal. Para poder disminuir dicho costo, y por ende volver esta CR de reacción accesible al usuario, es imperioso que se realice una paralelización del código.

- **Calcular la energía libre de la transición R→T.** El mismo nos permitiría analizar la influencia de histéresis observada en los movimientos terciarios, sobre la energía libre total durante la transición, lo que sería un paso de suma importancia para llegar a comprender la transición de la Hb humana. Para poder calcular dicho perfil se plantea realizar un muestreo estadístico significativo de la transición R-T con el modelo CG de la Hb humana, empleando el modelo CG de Voth y la CR de RMSD, de forma tal de contar con la transición completa en el sentido de interés. Si bien la SEP del sistema CG representa, en esencia, el comportamiento real del sistema, la pérdida de grados de libertad, asociada a la transformación AA→CG, podría enmascarar algunos fenómenos significativos para alcanzar el valor correcto de energía libre. Por esta razón nos proponemos obtener el FEP de la transición R→T a nivel atomístico, empleando, sin embargo, la información obtenida en la DMG CG, de manera de poder reducir el tiempo computacional total necesario para obtener la energía libre del movimiento conformacional. Planificamos, con este fin, tomar estructuras representativas de la transición R→T, obtenidas con el modelo de grano grueso una vez muestreada apropiadamente la transición, y convertirlas nuevamente al modelo atomístico. Este paso se conoce como *backmapping*, y su funcionamiento fue recientemente generalizado y aplicado al modelo de Voth, dentro del grupo de investigación [?]. Esto nos permitiría acceder a estructuras que, en el modelo atomístico, pueden requerir un costo energético muy elevado, pero que, sin embargo, son necesarias para realizar el cambio conformacional completo. Una vez que las estructuras escogidas se encuentran en la representación atomística, es necesario realizar simulaciones de equilibración, que permitan que las cadenas laterales (que ahora contienen una mayor cantidad de átomos) se ubiquen en la conformación geométrica de menor energía. Una vez que las estructuras se encuentren equilibradas en el nuevo modelo atomístico, se planea realizar un esquema de muestro sesgado de *Umbrella Sampling*, nuevamente con el RMSD como coordenada de reacción para obtener la FEP-AA final.
- **Probar una coordenada de reacción asociada al primer vector del modo esencial.** Motivados por los resultados obtenidos por los valores de modos esenciales obtenidos para el caso de la Hb (véase Figura 5.9), nos propusimos estudiar el modo de la transición como coordenada de reacción. Esto surge también como una posibilidad para los sistemas, tal como la Hb humana, que poseen un modo de la transición sumamente definido, que permitiría, en principio, guiar dicha transformación. Para esto se

implementó, dentro de la misma subrutina en la que se ejecuta el RMSD como CR (`conf_mov_rst.F90`, dentro del programa *pmemd* en el paquete de simulación molecular AMBER), una coordenada de reacción que utiliza el vector del modo de la transición (archivo de entrada, que debe ser otorgado por el usuario) para aplicar una fuerza sobre el sistema. La misma se aplica sobre los átomos sobre los cuales se realizó el cálculo de modos esenciales (por lo que una lista de átomos involucrados también es necesaria) en forma equivalente a una DMG, de manera que va variando acorde la coordenada transcurre. El parámetro que permite dicha variación es la diferencia entre los valores de la proyección del modo, sobre las DM de equilibrio de las estructuras inicial y final. Por ejemplo, para el caso de la Hb, este valor sería de aproximadamente 120 (véase Figura 5.9), valor que surge de la diferencia entre los valores de la proyección del modo de la transición sobre la DM para R (verde) y sobre la DM de T (azul). En este momento nos encontramos probando dicha coordenada de reacción en la ya estudiada Hb, para la cual contamos con una caracterización profunda de la transición que nos permitiría validar los resultados obtenidos.

En resumen, y mas allá de las propuestas futuras particulares, un punto importante del desarrollo futuro reside en la aplicabilidad y disponibilidad de la estrategia desarrollada para su uso por parte de terceros. Pensamos que esto es de vital importancia para ampliar la aplicación de los métodos desarrollados, y con esto el impacto que los mismos producirían en la comunidad científica.

BIBLIOGRAFÍA

- [1] P. R. ANDREWS, G. D. SMITH, AND I. G. YOUNG, *Transition-state stabilization and enzymic catalysis. Kinetic and molecular orbital studies of the rearrangement of chorismate to prephenate*, Biochemistry, 18 (1973), pp. 3492–3498.
- [2] P. A. BASH, M. J. FIELD, AND M. KARPLUS, *Free energy perturbation method for chemical reactions in the condensed phase: a dynamic approach based on a combined quantum and molecular mechanics potential*, J. Am. Chem. Soc., 109 (1987), pp. 8092–8094.
- [3] A. D. BECKE, *Density-functional thermochemistry. III. The role of exact exchange*, The Journal of Chemical Physics, 98 (1993), p. 5648.
- [4] C. H. BENNETT, *Efficient estimation of free energy differences from Monte Carlo data*, J. Comput. Phys., 22 (1976), pp. 245–268.
- [5] H. J. C. BERENDSEN, J. R. GRIGERA, AND T. P. STRAATSMA, *The missing term in effective pair potentials*, The Journal of Physical Chemistry, 91 (1987), pp. 6269–6271.
- [6] H. J. C. BERENDSEN, J. P. M. POSTMA, W. F. VAN GUNSTEREN, A. DI NOLA, AND J. R. HAAK, *Molecular dynamics with coupling to an external bath*, The Journal of chemical physics, 81 (1984), pp. 3684–3690.
- [7] H. J. C. BERENDSEN, J. P. M. POSTMA, W. F. VAN GUNSTEREN, A. DI NOLA, AND J. R. HAAK, *Molecular dynamics with coupling to an external bath*, The Journal of Chemical Physics, 81 (1984), p. 3684.
- [8] H. J. C. BERENDSEN, J. P. M. POSTMA, W. F. VAN GUNSTEREN, AND J. HERMANS, *Interaction Models for Water in Relation to Protein Hydration*, Springer Netherlands, 1981, pp. 331–342.
- [9] J. BLUMBERGER, G. LAMOUREUX, AND M. L. KLEIN, *Peptide Hydrolysis in Thermolysin: Ab Initio QM/MM Investigation of the Glu143-Assisted Water Addition Mechanism*, Journal of Chemical Theory and Computation, 3 (2007), pp. 1837–1850.

- [10] M. CAMMARATA, M. LEVANTINO, M. WULFF, AND A. CUPANE, *Unveiling the Timescale of the R-T Transition in Human Hemoglobin*, *Journal of Molecular Biology*, 400 (2010), pp. 951–962.
- [11] D. A. CASE, T. A. DARDEN, T. E. I. I. I. CHEATHAM, C. L. SIMMERLING, J. WANG, R. E. DUKE, R. LUO, R. C. WALKER, W. ZHANG, K. M. MERZ, B. ROBERTS, S. HAYIK, A. ROITBERG, G. SEABRA, J. SWAILS, A. W. GOETZ, I. KOLOSSVÁRY, K. F. WONG, F. PAESANI, J. VANICEK, R. M. WOLF, J. LIU, X. WU, S. R. BROZELL, T. STEINBRECHER, H. GOHLKE, Q. CAI, X. YE, J. WANG, M. J. HSIEH, G. CUI, D. R. ROE, D. H. MATHEWS, M. G. SEETIN, R. SALOMON-FERRER, C. SAGUI, V. BABIN, T. LUCHKO, S. GUSAROV, A. KOVALENKO, AND K. P. A., *AMBER 12*, University of California, San Francisco, 2012.
- [12] Y. M. CHOOK, J. V. GRAY, H. KE, AND W. N. LIPSCOMB, *The monofunctional chorismate mutase from Bacillus subtilis. Structure determination of chorismate mutase and its complexes with a transition state analog and prephenate, and implications for the mechanism of the enzymatic reaction*, *J.Mol.Biol.*, 240 (1994), pp. 476–500.
- [13] F. CLAEYSSSENS, K. E. RANAGHAN, N. LAWAN, S. J. MACRAE, F. R. MANBY, J. N. HARVEY, AND A. J. MULHOLLAND, *Analysis of chorismate mutase catalysis by QM/MM modelling of enzyme-catalysed and uncatalysed reactions*, *Org. Biomol. Chem.*, 9 (2011), pp. 1578–1590.
- [14] D. COLLIN, F. RITORT, C. JARZYNSKI, S. B. SMITH, I. TINOCO, AND C. BUSTAMANTE, *Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies.*, *Nature*, 437 (2005), pp. 231–4.
- [15] A. CRESPO, M. A. MARTÍ, D. A. ESTRIN, AND A. E. ROITBERG, *Multiple-steering QM-MM calculation of the free energy profile in chorismate mutase.*, *Journal of the American Chemical Society*, 127 (2005), pp. 6940–1.
- [16] A. CRESPO, D. A. SCHERLIS, M. A. MARTÍ, P. ORDEJÓN, A. E. ROITBERG, AND D. A. ESTRIN, *A DFT-Based QM-MM Approach Designed for the Treatment of Large Molecular Systems: Application to Chorismate Mutase*, *J. Phys. Chem. B*, 107 (2003), pp. 13728–13736.
- [17] G. E. CROOKS, *Nonequilibrium Measurements of Free Energy Differences for Microscopically Reversible Markovian Systems*, *Journal of Statistical Physics*, 90 (1998), pp. 1481–1487.
- [18] G. E. CROOKS, *Path-ensemble averages in systems driven far from equilibrium*, *Physical Review E*, 61 (2000), pp. 2361–2366.

- [19] Q. CUI, M. ELSTNER, E. KAXIRAS, T. FRAUENHEIM, AND M. KARPLUS, *A QM/MM Implementation of the Self-Consistent Charge Density Functional Tight Binding (SCC-DFTB) Method*, The Journal of Physical Chemistry B, 105 (2001), pp. 569–585.
- [20] T. DARDEN, D. YORK, AND L. PEDERSEN, *Particle mesh Ewald: An $N \log(N)$ method for Ewald sums in large systems*, The Journal of chemical physics, 98 (1993), pp. 10089–10092.
- [21] L. DARRÉ, M. R. MACHADO, A. F. BRANDNER, H. C. GONZÁLEZ, S. FERREIRA, AND S. PANTANO, *SIRAH: A Structurally Unbiased Coarse-Grained Force Field for Proteins with Aqueous Solvation and Long-Range Electrostatics*, Journal of Chemical Theory and Computation, 11 (2015), pp. 723–739.
- [22] G. DE M SEABRA, R. C. WALKER, M. ELSTNER, D. A. CASE, AND A. E. ROITBERG, *Implementation of the SCC-DFTB method for hybrid QM/MM simulations within the amber molecular dynamics package.*, The journal of physical chemistry. A, 111 (2007), pp. 5655–64.
- [23] S. L. DIXON AND K. M. MERZ JR., *Semiempirical molecular orbital calculations with linear system size scaling*, J. Chem. Phys., 104 (17) (1996), pp. 6643–6649.
- [24] I. ECHEVERRIA AND A. L. M., *Helix propensities calculations for amino acids in alanine based peptides using Jarzynski's equality*, Proteins, 78 (2010), pp. 1302–1310.
- [25] M. ELSTNER, *The SCC-DFTB method and its application to biological systems*, Theoretical Chemistry Accounts, 116 (2006), pp. 316–325.
- [26] M. ELSTNER, D. POREZAG, G. JUNGnickel, J. ELsNER, M. HAUGK, T. FRAUENHEIM, S. SUHAI, AND G. SEIFERT, *Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties*, Physical Review B, 58 (1998), pp. 7260–7268.
- [27] H. ESCHRIG AND I. BERGERT, *An optimized LCAO version for band structure calculations application to copper*, Physica Status Solidi (b), 90 (1978), pp. 621–628.
- [28] G. FERMI, M. F. PERUTZ, B. SHAANAN, AND R. FOURME, *The crystal structure of human deoxyhaemoglobin at 1.74 Å resolution.*, Journal of molecular biology, 175 (1984), pp. 159–74.
- [29] S. FISCHER, K. W. OLSEN, K. NAM, AND M. KARPLUS, *Unsuspected pathway of the allosteric transition in hemoglobin.*, Proceedings of the National Academy of Sciences of the United States of America, 108 (2011), pp. 5608–5613.

- [30] N. GO AND H. ABE, *Noninteracting local-structure model of folding and unfolding transition in globular proteins. I. Formulation*, Biopolymers, 20 (1981), pp. 991–1011.
- [31] M. C. GONZÁLEZ LEBRERO, D. E. BIKIEL, M. D. ELOLA, D. A. ESTRIN, AND A. E. ROITBERG, *Solvent-induced symmetry breaking of nitrate ion in aqueous clusters: A quantum-classical simulation study*, J. Chem. Phys., 117 (2002), pp. 2718–2725.
- [32] M. C. GONZÁLEZ LEBRERO AND D. A. ESTRIN, *QM-MM investigation of the reaction of peroxynitrite with carbon dioxide in water*, J. Chem. Theory Comput., 3 (2007), pp. 1405–1411.
- [33] H. GRUBMÜLLER, H. HELLER, A. WINDEMUTH, AND K. SCHULTEN, *Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions*, Molecular Simulation, 6 (1991), pp. 121–142.
- [34] X. HE AND K. M. J. MERZ, *Divide and Conquer Hartree-Fock Calculations on Proteins*, J. Chem. Theory Comput., 6 (2) (2010), pp. 405–411.
- [35] J. HÉNIN AND C. CHIPOT, *Overcoming free energy barriers using unconstrained molecular dynamics simulations.*, The Journal of chemical physics, 121 (2004), pp. 2904–14.
- [36] R. D. HILLS, L. LU, AND G. A. VOTH, *Multiscale Coarse-Graining of the Protein Energy Landscape*, PLoS Computational Biology, 6 (2010), p. e1000827.
- [37] P. HOHENBERG AND W. KOHN, *Inhomogeneous Electron Gas*, Physical Review, 136 (1964), pp. B864–B871.
- [38] V. HORNAK, R. ABEL, A. OKUR, B. STROCKBINE, A. ROITBERG, AND C. SIMMERLING, *Comparison of multiple Amber force fields and development of improved protein backbone parameters*, Proteins: Structure, Function, and Bioinformatics, 65 (2006), pp. 712–725.
- [39] J. S. HUB, M. B. KUBITZKI, AND B. L. DE GROOT, *Spontaneous quaternary and tertiary T-R transitions of human hemoglobin in molecular dynamics simulation.*, PLoS computational biology, 6 (2010), p. e1000774.
- [40] C. JARZYNSKI, *Nonequilibrium Equality for Free Energy Differences*, Physical Review Letters, 78 (1997), pp. 2690–2693.
- [41] E. M. JONES, G. BALAKRISHNAN, AND T. G. SPIRO, *Heme Reactivity is Uncoupled from Quaternary Structure in Gel-Encapsulated Hemoglobin: A Resonance Raman Spectroscopic Study*, Journal of the American Chemical Society, 134 (2012), pp. 3461–3471.

- [42] W. L. JORGENSEN, J. CHANDRASEKHAR, J. D. MADURA, R. W. IMPEY, AND M. L. KLEIN, *Comparison of simple potential functions for simulating liquid water*, The Journal of Chemical Physics, 79 (1983), p. 926.
- [43] P. KAST, M. ASIF-ULLAH, AND D. HILVERT, *Is chorismate mutase a prototypic entropy trap? - Activation parameters for the Bacillus subtilis enzyme*, Tetrahedron Letters, 37 (1996), pp. 2691–2694.
- [44] P. KAST, Y. B. TEWARI, O. WIEST, D. HILVERT, K. N. HOUK, AND R. N. GOLDBERG, *Thermodynamics of the Conversion of Chorismate to Prephenate: Experimental Results and Theoretical Predictions*, The Journal of Physical Chemistry B, 101 (1997), pp. 10976–10982.
- [45] D. E. KOSHLAND, G. NÉMETHY, AND D. FILMER, *Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits*, Biochemistry, 5 (1966), pp. 365–385.
- [46] S. KUMAR, J. M. ROSENBERG, D. BOUZIDA, R. H. SWENDSEN, AND P. A. KOLLMAN, *THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method*, Journal of Computational Chemistry, 13 (1992), pp. 1011–1021.
- [47] A. LAIO AND M. PARRINELLO, *Escaping free-energy minima.*, Proceedings of the National Academy of Sciences of the United States of America, 99 (2002), pp. 12562–6.
- [48] A. R. LEACH, *Molecular modelling : principles and applications*, Longman, 1996.
- [49] H. LEE WOODCOCK, M. HODOŠČEK, P. SHERWOOD, Y. S. LEE, H. F. SCHAEFER III, AND B. R. BROOKS, *Exploring the quantum mechanical/molecular mechanical replica path method: a pathway optimization of the chorismate to prephenate Claisen rearrangement catalyzed by chorismate mutase*, Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta), 109 (2003), pp. 140–148.
- [50] I. N. LEVINE, *Quantum chemistry*, Prentice Hall, 2000.
- [51] K. LINDORFF-LARSEN, S. PIANA, K. PALMO, P. MARAGAKIS, J. L. KLEPEIS, R. O. DROR, AND D. E. SHAW, *Improved side-chain torsion potentials for the Amber ff99SB protein force field.*, Proteins, 78 (2010), pp. 1950–8.
- [52] J. LIPHARDT, S. DUMONT, S. B. SMITH, I. J. TINOCO, AND C. BUSTAMANTE, *Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality*, Science, 296 (2002), pp. 1832–1835.

- [53] J. MONOD, J. WYMAN, AND J.-P. CHANGEUX, *On the Nature of Allosteric Transitions: A Plausible Model*, Journal of Molecular Biology, 12 (1965), pp. 88–118.
- [54] L. MONTICELLI, S. K. KANDASAMY, X. PERIOLE, R. G. LARSON, D. P. TIELEMAN, AND S.-J. MARRINK, *The MARTINI Coarse-Grained Force Field: Extension to Proteins*, Journal of Chemical Theory and Computation, 4 (2008), pp. 819–834.
- [55] U. N. MORZAN, F. F. RAMÍREZ, M. B. OVIEDO, C. G. SÁNCHEZ, D. A. SCHERLIS, AND M. C. G. LEBRERO, *Electron dynamics in complex environments with real-time time dependent density functional theory in a QM-MM framework*, J. Phys. Chem., 140 (2014), pp. 164105–164114.
- [56] M. A. NITSCHKE, M. FERRERIA, E. E. MOCSKOS, AND M. C. GONZÁLEZ LEBRERO, *GPU Accelerated Implementation of Density Functional Theory for Hybrid QM/MM Simulations*, J. Chem. Theory Comput., 10 (2014), pp. 959–967.
- [57] G. OZER, S. QUIRK, AND R. HERNANDEZ, *Adaptive steered molecular dynamics: validation of the selection criterion and benchmarking energetics in vacuum.*, The Journal of chemical physics, 136 (2012), p. 215104.
- [58] G. OZER, E. F. VALEEV, S. QUIRK, AND R. HERNANDEZ, *Adaptive Steered Molecular Dynamics of the Long-Distance Unfolding of Neuropeptide Y.*, Journal of chemical theory and computation, 6 (2010), pp. 3026–38.
- [59] S. PARK, F. KHALILI-ARAGHI, E. TAJKHORSHID, AND K. SCHULTEN, *Free energy calculation from steered molecular dynamics simulations using Jarzynski’s equality*, The Journal of Chemical Physics, 119 (2003), p. 3559.
- [60] J. P. PERDEW, K. BURKE, AND M. ERNZERHOF, *Generalized Gradient Approximation Made Simple*, Physical Review Letters, 77 (1996), pp. 3865–3868.
- [61] M. F. PERUTZ, A. J. WILKINSON, M. PAOLI, AND G. G. DODSON, *The stereochemical mechanism of the cooperative effects in hemoglobin revisited.*, Annual review of biophysics and biomolecular structure, 27 (1998), pp. 1–34.
- [62] A. POHORILLE, C. JARZYNSKI, AND C. CHIPOT, *Good practices in free-energy calculations.*, The journal of physical chemistry. B, 114 (2010), pp. 10235–53.
- [63] D. POREZAG, T. FRAUENHEIM, T. KÖHLER, G. SEIFERT, AND R. KASCHNER, *Construction of tight-binding-like potentials on the basis of density-functional theory: Application to carbon*, Physical Review B, 51 (1995), p. 12947.

- [64] C. RAMÍREZ, M. MARTÍ, AND A. ROITBERG, *Chapter Six – Steered Molecular Dynamics Methods Applied to Enzyme Mechanism and Energetics*, in *Methods in Enzymology*, vol. 578, 2016, pp. 123–143.
- [65] C. L. RAMÍREZ, A. PETRUK, M. BRINGAS, D. A. ESTRIN, A. E. ROITBERG, M. A. MARTI, AND L. CAPECE, *Coarse-Grained Simulations of Heme Proteins: Validation and Study of Large Conformational Transitions*, *Journal of Chemical Theory and Computation*, 12 (2016), pp. 3390–3397.
- [66] J. M. ROMERO, M. MARTIN, C. L. RAMIREZ, V. G. DUMAS, AND M. A. MARTI, *Efficient Calculation of Enzyme Reaction Free Energy Profiles Using a Hybrid Differential Relaxation Algorithm: Application to Mycobacterial Zinc Hydrolases.*, *Advances in protein chemistry and structural biology*, 100 (2015), pp. 33–65.
- [67] J.-P. RYCKAERT, G. CICCOTTI, AND H. J. BERENDSEN, *Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes*, *Journal of Computational Physics*, 23 (1977), pp. 327–341.
- [68] O.-P. SAIRA, Y. YOON, T. TANTTU, M. MÖTTÖNEN, D. V. AVERIN, AND J. P. PEKOLA, *Test of the Jarzynski and Crooks fluctuation relations in an electronic system.*, *Physical review letters*, 109 (2012), p. 180601.
- [69] R. SALOMON-FERRER, A. W. GOETZ, D. POOLE, S. LE GRAND, AND R. C. WALKER, *Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald*, *Journal of Chemical Theory and Computation*, 9 (2013), pp. 3878–3888.
- [70] M. S. K. P. A. SETTLE, *An analytical version of the SHAKE and RATTLE algorithm for rigid water molecules*, *J. Comput. Chem*, 13 (1992), pp. 952–962.
- [71] B. SHAANAN, *Structure of human oxyhaemoglobin at 2.1 Å resolution.*, *J.Mol.Biol.*, 171 (1983), pp. 31–59.
- [72] C. R. SMITH, G. K. SMITH, Z. YANG, D. XU, AND H. GUO, *Quantum mechanical/molecular mechanical study of anthrax lethal factor catalysis*, *Theoretical Chemistry Accounts*, 128 (2010), pp. 83–90.
- [73] J. J. P. STEWART, *Application of the PM6 method to modeling proteins*, *Journal of Molecular Modeling*, 15 (7) (2009), pp. 765–805.
- [74] J. J. P. STEWART, *MOPAC2016*, Stewart Computational Chemistry, Colorado S (2016).

- [75] J. TORRAS, G. M. DE SEABRA, E. DEUMENS, S. B. TRICKEY, AND A. E. ROITBERG, *A versatile AMBER-Gaussian QM/MM interface through PUPIL*, J. Comput. Chem., 29 (2008), pp. 1564–1573.
- [76] M. W. VAN DER KAMP AND A. J. MULHOLLAND, *Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology.*, Biochemistry, 52 (2013), pp. 2708–28.
- [77] J. WANG, R. M. WOLF, J. W. CALDWELL, P. A. KOLLMAN, AND D. A. CASE, *Development and testing of a general amber force field.*, Journal of computational chemistry, 25 (2004), pp. 1157–74.
- [78] A. B. WARD, O. GUVENCH, AND R. D. HILLS, *Coarse grain lipid-protein molecular interactions and diffusion with MsbA flippase*, Proteins: Structure, Function, and Bioinformatics, 80 (2012), pp. 2178–2190.
- [79] A. WARSHHEL, *Computer Simulations of Enzyme Catalysis: Methods, Progress, and Insights*, Annu. Rev. Biophys. Biomol. Struc., 32 (2003), pp. 425–443.
- [80] A. WARSHHEL AND M. LEVITT, *Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme*, J. Mol. Biol., 103 (1976), pp. 227–249.
- [81] T. K. WOO, P. MARGL, P. E. BLÖCHL, AND T. ZIEGLER, *Sampling Phase Space by a Combined QM/MM ab Initio Car-Parrinello Molecular Dynamics Method with Different (Multiple) Time Steps in the Quantum Mechanical (QM) and Molecular Mechanical (MM) Domains*, The Journal of Physical Chemistry A, 106 (2002), pp. 1173–1182.
- [82] H. L. WOODCOCK, M. HODOSCEK, AND B. R. BROOKS, *Exploring SCC-DFTB paths for mapping QM/MM reaction mechanisms.*, The journal of physical chemistry. A, 111 (2007), pp. 5720–8.
- [83] H. XIONG, A. CRESPO, M. MARTI, D. ESTRIN, AND A. E. ROITBERG, *Free Energy Calculations with Non-Equilibrium Methods: Applications of the Jarzynski Relationship*, Theoretical Chemistry Accounts, 116 (2006), pp. 338–346.
- [84] D. XU AND H. GUO, *Quantum mechanical/molecular mechanical and density functional theory studies of a prototypical zinc peptidase (carboxypeptidase A) suggest a general acid-general base mechanism.*, Journal of the American Chemical Society, 131 (2009), pp. 9780–8.

- [85] A. ZEIDA, R. BABBUSH, M. C. GONZÁLEZ LEBRERO, M. TRUJILLO, R. RADI, AND D. A. ESTRIN, *Molecular basis of the mechanism of thiol oxidation by hydrogen peroxide in aqueous solution: challenging the SN2 paradigm*, Chem. Res. Toxicol., 25 (2012), pp. 741–746.
- [86] C. ZHANG, S. WU, AND D. XU, *Catalytic mechanism of angiotensin-converting enzyme and effects of the chloride ion.*, The journal of physical chemistry. B, 117 (2013), pp. 6635–45.
- [87] L. ZHENG, M. CHEN, AND W. YANG, *Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems.*, Proceedings of the National Academy of Sciences of the United States of America, 105 (2008), pp. 20227–32.
- [88] R. W. ZWANZIG, *High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases*, The Journal of Chemical Physics, 22 (1954), p. 1420.



PUBLICACIONES

Todo el trabajo presentado en esta tesis fue exitosamente publicado a lo largo de los años de doctorado, en revistas internacionales con referato. Dichas publicaciones se detallan a continuación.

Improving Efficiency in SMD Simulations Trough a Hybrid Differential Relaxation Algorithm.

C. L. Ramírez, A. Zeida, G. E. Jara, A. E. Roitberg and M. A. Martí; *Journal of Chemical Theory and Computation*, 2014, 10(10):4609-4617.

Chapter Two – Efficient Calculation of Enzyme Reaction Free Energy Profiles Using a Hybrid Differential Relaxation Algorithm: Application to Mycobacterial Zinc Hydrolases. J. M. Romero, M. Martin, **C. L. Ramírez**, V. G. Dumas, M. A. and Martí ; *Advances in Protein Chemistry and Structural Biology*, 2015, 100:33-65.

Chapter 23: Steered molecular dynamics methods applied to enzyme mechanism and energetics.

C. L. Ramírez, M. A. Martí and A. E. Roitberg; *Methods in Enzymology*, Vol. 578, *Computational Approaches for Studying Enzyme Mechanism Part B*, Chapter 6, Elsevier, 2016,

Coarse-grained simulations of heme proteins: validation and study of large conformational transitions.**C. L. Ramírez**, A. Petruk, D. A. Estrín, A. E. Roitberg, M. A. Martí and L. Capece; *Journal of Chemical Theory and Computation*, 2016



FORMACIÓN ACADÉMICA DURANTE EL DOCTORADO

Durante el período comprendido en mi doctorado realicé una gran variedad de cursos y materias que me ayudaron a comprender más las diversas temáticas relacionadas con mi Tesis, participé de diversos congresos, realicé pasantías en otros laboratorios, organicé actividades de extensión y di clases. Siento que mi formación y mi crecimiento como científica, y persona, durante el doctorado se encuentran íntegramente ligadas no sólo al trabajo realizado y presentado en esta Tesis, sino también a la formación que recibí de las demás actividades.

Durante este tiempo realicé las siguientes materias y cursos:

- **Química Cuántica.** 2º cuatrimestre de 2012.
- **Química General e Inorgánica II.** 1º cuatrimestre de 2013.
- **Curso de Simulación Computacional.** 07/2013.
- **Química Orgánica para Paleontólogos.** 1º cuatrimestre de 2014.
- **Cinética Química.** 2º cuatrimestre de 2014.
- **Escuela ECAR 2014 de High Performance Computing.** Universidad Técnica Federico Santa María, Valparaíso, Chile. 10/2014.
- **Química Biológica 2B.** 2º cuatrimestre de 2015.
- **Sistemas Electroquímicos de Almacenamiento y Conversión de Energía.** 1º cuatrimestre de 2016.

- **Curso Internacional “Introductory School on Parallel Programming and Parallel Architecture for High-Performance Computing”.** International Centre for Theoretical Physics, Trieste, Italia. 10/2016.

Y tuve el placer de poder presentar mi trabajo en diversos congresos:

- **XL Reunión Anual Sociedad Argentina de Biofísica.** Instituto Leloir, Buenos Aires, Argentina. Diciembre 2011. Presentación de Poster.
- **53rd Sanibel Symposium.** St. Simons Island, Georgia, Estados Unidos. Febrero de 2013. Presentación de Poster.
- **98° Reunión Nacional de Física de la Asociación Física Argentina.** Instituto Balseiro, Río Negro, Argentina. Octubre de 2013. Presentación de Poster.
- **Primer Congreso Latinoamericano de HPC.** Universidad Técnica Federico Santa María, Valparaíso, Chile. Marzo de 2014. Asistente.
- **249th ACS National Meeting & Exposition.** Colorado Convention Center, Denver, Estados Unidos. Marzo 2015. Presentación Oral.

Durante este tiempo también realicé pasantías en laboratorios en el exterior, siendo éstas experiencias que me enriquecieron profundamente, permitiéndome no sólo establecer contactos de gran importancia, sino también conocer algunas de las diversas formas de hacer ciencia en el mundo. Las dos pasantías que realicé en el laboratorio del Dr. Adrián Roitberg (de enero a marzo de 2013 y de diciembre de 2014 a abril de 2015) en la Universidad de Florida -Gainesville, Estados Unidos- fueron experiencias muy valiosas para la labor aquí presentada y, supongo también, para el resto de mi carrera científica. Además de estas dos, tuve la maravillosa experiencia de pasar 5 semanas del año 2015 trabajando en el laboratorio de la Dra. Lynn Kammerlin -Structural Biology Program of the Department of Cell and Molecular Biology- en la Universidad de Uppsala, aprendiendo de la metodología EVB. No sólo a nivel científico esta fue una increíble oportunidad para mí, sino que también los contactos realizados con su grupo enormemente productivo, fueron sin duda una de las mayores ventajas de este intercambio.

No me gustaría dejar atrás el trabajo de extensión realizado los 3 primeros años de mi doctorado, organizando actividades para los alumnos de colegio secundario que participarían en la *Semana de la Química*. Esta fue sin duda una preparación exitosa para lo que sería mi primer (y único) año como docente de un laboratorio de química en el 2015 (Química Inorgánica). Sin duda la experiencia de dar una materia en una carrera que no es la mía, y que nunca había cursado, fue uno de los momentos donde más aprendí, y más disfruté

de mi decisión de haber realizado un Doctorado en el área de Química. Estas maravillosas experiencias en la Universidad de Buenos Aires me llevaron a la que es sin duda mi actividad “extracurricular” más gratificante hasta el momento: Enseñar Físico-Química aplicada al Medio Ambiente, en un Bachillerato Popular en el barrio de Las Tunas (Gral. Pacheco). Aprender a “bajar” todo el contenido teórico acumulado durante tantos años, aplicarlo a problemas cotidianos-ambientales y discutirlo con alumnos de un secundario para adultos, es algo que me llena de alegría y que sin duda mantiene despierto en mí el deseo de seguir aprendiendo.

Espero este deseo continúe movilizándome por muchos años más.



AGRADECIMIENTOS

Finalizando esta etapa de crecimiento, tanto individual como laboral, se me desborda de alegría el corazón al recordar estos 5 años en estas tierras nuevas para mí y en este trabajo igual de nuevo. Haré lo imposible por recordar a todes lo que hicieron posible que hoy esté acá.

Primero que nada a Dios. O a la Naturaleza. O la evolución. O a como sea que cada uno prefiera llamarlo. A ese *algo* que llena el Universo de seres y *cosas* tan hermosas e interesantes, sin esa existencia este trabajo nunca se hubiese realizado. Gracias también, no sólo por darme la inteligencia para comprender algunos de estos fenómenos, sino también por haberme generado la curiosidad, el asombro y la diversión que me produce estudiar lo que nos rodea.

A las feministas que lucharon, y con las que lucho, para que este mundo sea un lugar más justo para todes. Insisto, nada de esto sería posible sin las luchas que supieron dar y ganar.

A mis *directores*, que aunque figure uno, todos sabemos que fue un trabajo de los dos. A **Marce**, que fue el primero en aceptarme, que no se espantó cuando lloré todos los días por una semana entera, al lado de él, cuando tenía que terminar de escribir la tesina, ni hablar de las mil y un correcciones durante esta tesis, por estar “al pie del cañón” (aunque fuera sábado o domingo). Gracias, infinitas, por las mil y un discusiones a los gritos, gracias porque no sólo estuviste para guiarme científicamente, sino que también sos una gran persona, siempre preocupandote de cómo estaba e intentando hacer todo lo que estuviera en tu poder para que yo estuviera mejor. Y en ese intento de que todo “estuviera mejor” y yo no perdiera mi

parte de “física”, se te ocurrió decirle a Adrián que te diera una mano.

Gracias **Adri** por aceptar, por las infinitas discusiones en tu oficina, que duraban toda la tarde, con unos mates de por medio claro está. Gracias también por las otras charlas interminables, a las que nunca dijiste que no, aconsejándome, alentándome y cuidando que siempre diera lo mejor de mí.

Los dos son personas increíbles, que se complementaron perfectamente en estos 5 años, nunca hubiese podido soñar un mejor dúo para que me conduzca en este camino, *gracias* por haber estado para mí, siempre.

A los **jurados**, que aceptaron gustosamente acompañarme en esta última etapa del doctorado, y que se tomaron el trabajo de leer y cuestionar todo lo que intenté explicar en este trabajo, enriqueciendo mi formación a partir de sus perspectivas.

A **Fabio** (C.), que fue más que un consejero de estudios, que se preocupó de mi formación académica (a veces incluso más que yo), que me supo enseñar puntos de vista diferentes y que supo también indignarse conmigo y por mí. Gracias Fabio por tratarme como una persona y no como un número más.

A la UNNE, que me formó para que pudiera obtener mi título de Licenciada (sin el cual no estaría acá), a la UBA por haberme formado otros 5 años y a CONICET por otorgarme dado la beca que me permitió realizar este doctorado. Por sobre todo, estoy inmensamente agradecida a la apuesta política de la **Educación Pública**, que nos ayuda a todos a crecer. Gracias a todos los que con su trabajo hicieron posible que yo llegara hasta acá.

A las secretarias!

A **Ale**, genia, crack, sin vos no hubiese podido hacer ningún papel bien, gracias por la paciencia infinita de explicarme todo las 20 veces que te preguntaba.

Eli, sin palabras, el GMM no funcionaría sin vos, yo estaría debiéndole plata a todo el mundo, o no me habría ido a ningún lado si no fuera por vos y tu paciencia infinita (incluso cuando perdí el boardingpass!).

Gracias por no sólo hacer su labor bien, sino que por hacerlo con alegría y contagiándonos sonrisas, y haciendo que tengamos ganas de ir a verlas.

Al grupo, todo.

En especial a los que tuvieron el coraje de trabajar conmigo y no huyeron! Ari, Franco, Eli, Jara, Juan y Osvald.

Al maravilloso **EI** en toda su gloria y esplendor, que por más cliché que suene, hicieron que ir

a trabajar sea super divertido, tan tan tan divertido, que tuve que irme el último año para poder concentrarme y terminar de escribir!

Gracias (sin repetir y sin soplar, en orden “geográfico”) Veci, Estefi, Ithu, Gauto, Facto, Laia, Marianito, Fede peque, Romerito, Joni, Marianito, Fran, Yami, Dieguito, Pablito, Nick, Mauro. Y a los “extras”: Nano, Darío, Damián, Lula, Gabriela. Conversaciones inolvidables que siempre que las recuerdo me sacan una sonrisa.

A Charlie, que se me escapó hace unos años para el norte, pero que me hizo sentir menos sapo de otro pozo cuando llegue, te quiero muchísimo nene!!

A las y los del **cuarto** que me adoptaron tan cariñosamente en mi mudanza, dándome la oportunidad de hacer amigos que habían quedado a medias (cuando ellos se mudaron).

A Lean, Mode, Eli, Lucas, Sol, Franco, Juan, Osvald, y a las “agregadas” de Lula y Pau, gracias por hacer que mi mudanza fuera tan fácil!!

A **Uri**, por ser mi compañero, mi apoyo, mi bufón y mi mejor amigo durante el primer tiempo acá, gracias por aguantar mis cambios de humor, mis conversaciones sin parar, y gracias por quererme tanto como lo hiciste, fue hermoso para mí. Te quiero, muchísimo.

A **las simuladoras**, ya se los dije mil veces, esta ciudad no tiene ni la mitad de sentido cuando uds no están, y no puedo explicar cuanto las quiero y cuán feliz me hace que estén en mi vida.

Lu, gracias por integrarme desde el primer momento y ser siempre tan alegre, sos de oro y no sé dónde estaría sin mi amiga *feminista*.

Vicky, gracias por el aguante increíble siempre, y especialmente en el último tiempo de la tesis, porque me entendías perfectamente y supiste acompañarme, incluso cuando vos tenías mil cosas que hacer.

Noosh, gracias por ser vos, por la calidez, por la cercanía, por los consejos, las charlas, por mostrarte siempre como sos y dejarme a mi ser quien soy, gracias por siempre incentivarme a ser *más yo*.

A las tres, las quiero, infinito, a cada una en sus particularidades, porque cada una me da algo bien distinto, pero igual de necesario.

Gracias también a los *agregados* que trajeron con ustedes: **Emi, León, Juan y Leo**.

A sus *bebés* hermosos que me alegran muchísimo y que me hacen maravillarme siempre que los veo, son peques adorables, gracias por traerlos a este mundo para hacer un mejor lugar.

A sus compañeros de vida, que se fueron haciendo mis amigos también con el paso del tiempo, de las cenas, de las salidas, de la vida en general. Gracias por las charlas, los consejos y el aguante.

Al **Bachi**, que fue lo mejor que me pasó este año. A todes les que conforman ese espacio

hermoso, profes y estudiantes, gracias por enseñarme que otra forma de educación es posible, por abrirme los ojos, por aceptarme... y por re-educarme.

A las *chaqueñas*, que son de mi ciudad natal hermosa, y que me acompañaron en la distancia, con las que la amistad sigue intacta, a pesar de la distancia.

Lu, gracias por acompañarme en la carrera, por quererme aunque yo sea un cuelgue, por los años de convivencia, y por tu bondad sin límites.

Anita, gracias por tu sinceridad sin límites, por aguantarme cuando ni yo me aguanto, por conocerme mejor que yo, y ayudarme a conocerme más, por regalarme tu amistad hermosa hace tantos años.

Numi, gracias por ser vos..por estar conectada siempre, por entenderme, por ayudarme, por las charlas de vida, por los consejos y los mimos al alma, gracias porque a pesar de todo lo que pasamos (juntas y separadas), siempre pude contar con vos.

Las quiero muchísimo, pero muchísimo de verdad.

A mi **familia**, que me crió y me acompañó durante toda mi vida, que siempre me enseñó que lo importante es hacer lo que a uno le gusta, y que me alentó siempre a seguir aquello que me apasionaba, incluso cuando eso significó irme lejos.

A mis nonos y abuelos. Ninguno está acá para acompañarme hoy, pero el cariño y el ejemplo de vida que me dieron está siempre presente en mi vida.

A mis viejos, que siempre me hicieron el aguante, que me apoyaron en mis decisiones, que me acompañan siempre, que me quieren infinito, y que siempre están buscando ayudarme. Muchas gracias por ser lo suficientemente fuertes y cariñosos como para acompañarme en la distancia, y para siempre darme alas para volar adonde quiera.

A Mime, por ir “rompiendo moldes” adelante nuestro, para que para el resto sea más fácil. Gracias también por darme a Mateo y Milo, que son mi mayor alegría cuando estoy en Resistencia.

Al flaco, con el que compartimos infinidad de charlas y consejos, gracias por ser un hermano compañero.

A la peque/la favorita/Ayelén/Belu, gracias hermana por existir, por ser tan distinta a mí, por quererme como me querés, por acompañarme, por las charlas, las risas y hasta los enojos. La vida es mucho más linda desde que existís.

Y por último gracias a mí, por haber llegado hasta acá, incluso cuando a veces pensé que no iba a llegar. A mi, también, me quiero mucho y agradezco.