

Tesis Doctoral

Estudio teórico-experimental de migración y unión de ligandos pequeños en hemoproteínas

Boubeta, Fernando Martín

2017-03-10

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:

Boubeta, Fernando Martín. (2017-03-10). Estudio teórico-experimental de migración y unión de ligandos pequeños en hemoproteínas. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires.

Cita tipo Chicago:

Boubeta, Fernando Martín. "Estudio teórico-experimental de migración y unión de ligandos pequeños en hemoproteínas". Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires. 2017-03-10.

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.

Estudio teórico-experimental de migración y unión de ligandos pequeños en hemoproteínas

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

Autor

Lic. Fernando Martín Boubeta

Director de Tesis

Dr. Darío Ariel Estrin

Director Asistente

Dr. Leonardo Boechi

Consejero de Estudios

Dr. Federico Williams

Lugar de Trabajo

Grupo de Modelado Molecular

Departamento de Química Inorgánica, Analítica y Química Física.

Instituto de Química Física de Materiales, Ambiente y Energía (INQUIMAE-CONICET).

Facultad de Ciencias Exactas y Naturales

Universidad de Buenos Aires

Ciudad Autónoma de Buenos Aires, 2017

Fecha de defensa: 10 de Marzo de 2017

Resumen

Estudio teórico-experimental de la migración y unión de ligandos pequeños en hemoproteínas

Las hemoproteínas se encuentran en todos los organismos vivos y tienen una gran variedad de funciones, desde el transporte de ligandos hasta fenómenos bioquímicos redox. Las funciones de las hemoproteínas están típicamente asociadas a la interacción con pequeños ligandos, como O_2 , NO, HNO, H_2S , O_2^- , etc. Una de las etapas fundamentales de esta interacción son la migración del ligando a través de la matriz proteica y su posterior unión al centro metálico del hemo. Estos procesos pueden limitar la velocidad del proceso global, con su consecuente impacto biológico sobre el organismo en que se encuentra la hemoproteína.

La complejidad de las hemoproteínas en función de su estructura, o condiciones del medio (temperatura, presión, etc) requieren de la utilización de diversos enfoques para abordar un estudio de los procesos mencionados.

El presente trabajo de tesis tuvo como objetivo estudiar, a partir de estudios teóricos, y métodos tanto de simulación computacional como experimentales, el proceso de migración y unión de ligandos a hemoproteínas

En primer lugar, se abordó un estudio teórico sobre los estimadores de energía libre derivados de la ecuación de Jarzynski, para distribuciones de trabajo generalmente obtenidas en simulaciones de dinámica molecular guiada, tomando como caso paradigmático la migración de ligandos en metaloproteínas.

En segundo lugar, se estudió mediante diversas técnicas de simulación computacional clásica y multi-escala cuántico-clásica (QM/MM) la especiación de especies de sulfuro (H_2S) en hemoproteínas, utilizando además las metodologías desarrolladas en el primer capítulo.

Finalmente, se estudió mediante técnicas experimentales de láser flash fotólisis el efecto de la temperatura en la migración y unión de ligandos diatómicos en hemoglobinas truncadas pertenecientes a bacterias que habitan en condiciones diferentes de temperatura: *P. Haloplanktis*, *B. Subtilis*, y *T. Fusca*, que son consideradas psicrófilas, mesófilas, y termófilas, respectivamente.

Palabras clave: hemoproteínas, pequeños ligandos, migración, dinámica molecular guiada, ecuación de Jarzynski, dinámica molecular, QM/MM, cinética de Laser Flash Fotólisis, constantes de asociación k_{on}

Abstract

Theoretical-experimental study of the migration and binding of small ligands to heme proteins

Heme proteins are found in all living organisms and have a gamut of functions, from the transport of ligands to redox chemistry. The functions of heme proteins are typically associated with their interactions with small ligands like O₂, NO, HNO, H₂S and O₂⁻. The fundamental stages of this interaction are the migration of the ligand through the protein matrix and its subsequent binding to the heme metal center. In many cases, these processes may limit the rate of the global association process, with a consequent biological impact on the organism to which the heme protein belongs.

The complexity of heme proteins as a function of their structures, or the conditions of the medium (temperature, pressure, etc.) require the use of diverse perspectives to encompass the study of the mentioned processes.

The main goal of the present dissertation is to study from theoretical studies, computational simulations and experimental methods, the process of migration and binding of small ligands to heme proteins.

First, a theoretical/mathematical study was carried out on the use of different estimators of the free energy derived from the Jarzynski equality for underlying work distributions obtained using steered molecular dynamics simulations, using the migration of small ligands in metalloproteins as a paradigmatic case.

Second, a combination of several classical and multi-scale quantum-classical (QM/MM) simulation techniques were used to study the speciation of hydrogen sulfide (H₂S) in heme proteins, employing besides the tools developed in the first chapter.

Finally, using experimental techniques, namely laser flash photolysis, the effect of temperature was studied on the migration and binding of diatomic ligands in truncated hemoglobins belonging to bacteria adapted to extreme temperatures, namely *P. Haloplanktis*, *B. Subtilis*, and *T. Fusca*, which are considered psicrófiles, mesophiles, and thermophiles, respectively.

Key words: heme proteins, small ligands, migration, steered molecular dynamics, Jarzynski equality, molecular dynamics, QM/MM, laser flash photolysis, association rate constant k_{on} .

a *Tami*

a mis Viejos y

la tribu de mi calle

*“al fin! sean bienvenidos todos
al show de la linda fe sonriente!
Nos merecemos bellos milagros
y ocurrirán...”*

Carlos Alberto Solari, “Amok! Amok!”, 2013

ÍNDICE GENERAL

RESUMEN

ABSTRACT

I.	INTRODUCCIÓN GENERAL	1
I.1	PROTEÍNAS	1
I.1.1	<i>Hemoproteínas</i>	1
I.2	UNIÓN Y DISOCIACIÓN DE PEQUEÑOS LIGANDOS.	6
I.2.1	<i>Caracterización cinética y termodinámica de los procesos de unión y disociación de pequeños ligandos en hemoproteínas.</i>	8
I.2.2	<i>Mediciones experimentales de las constantes de asociación (k_{on}) y disociación (k_{off})</i>	11
I.3	IMPORTANCIA DE LA TEMPERATURA EN LA ACTIVIDAD DE PROTEÍNAS.	13
I.4	UN NUEVO GASOTRANSMISOR: ÁCIDO SULFHÍDRICO (H_2S).	16
I.5	SIMULACIONES COMPUTACIONALES PARA LA OBTENCIÓN DE INFORMACIÓN QUÍMICA.	18
I.6	OBJETIVOS Y LINEAMIENTOS DE LA PRESENTE TESIS	21
I.7	REFERENCIAS	22
II.	METODOLOGÍA GENERAL	30
II.1	PORQUÉ USAR SIMULACIÓN COMPUTACIONAL EN QUÍMICA	30
II.2	MÉTODOS DE SIMULACIÓN COMPUTACIONAL	30
II.2.1	<i>Métodos de estructura electrónica</i>	31
II.2.1.1	Breve reseña histórica	31
II.2.1.2	Principios básicos	33
II.2.1.3	El Teorema Variacional y el método variacional lineal	37
II.2.1.4	El método de Hartree-Fock (HF)	40
II.2.1.5	Teoría del Funcional de la Densidad (DFT)	44
II.2.1.5.1	Funcionales de Intercambio y correlación	47
II.2.2	<i>El método SIESTA</i>	49
II.2.3	<i>Métodos basados en la Mecánica Clásica</i>	52
II.2.3.1	Campos de fuerza clásicos – Simulaciones atómicas	53
II.2.3.1.1	Modelos de agua explícita	56
II.2.4	<i>Métodos multi-escala cuántico-clásicos (QM/MM)</i>	58
II.2.5	<i>Métodos de exploración de la superficie de energía potencial</i>	61
II.2.5.1	Optimizaciones de geometría – Búsqueda del mínimo de energía	62
II.2.5.2	Búsqueda del camino de transformación	63
II.2.5.3	Simulaciones de Dinámica Molecular	64

II.2.5.3.1 Simulación de condiciones experimentales: temperatura y presión	66
II.2.6 <i>Análisis avanzados de las simulaciones</i>	68
II.2.6.1 RMSD	68
II.2.6.2 Identificación de los Sitios de hidratación (WS)	69
II.2.7 <i>Cálculos de Energía Libre</i>	70
II.2.7.1 Dinámica Molecular Guiada (sMD)	71
II.2.7.2 Muestro de Ligando Implícito (ILS)	72
II.3 MÉTODOS EXPERIMENTALES	75
II.3.1 <i>Técnica de láser flash fotólisis (LFP)</i>	75
II.3.1.1 Principios básicos de los experimentos de LFP	76
II.3.1.1.1 Esquema experimental del equipo de Laser Flash Fotólisis	77
II.3.1.2 Determinación de parámetros termodinámicos de activación a partir de medidas cinéticas: modelos de Arrhenius y Eyring	81
II.4 REFERENCIAS	83
III. ESTUDIO TEÓRICO DE ESTIMADORES DE ENERGÍA LIBRE ASOCIADOS A LA IGUALDAD DE JARZYNSKI	87
III.1 ENERGÍA LIBRE: MOTIVACIONES	87
III.2 EL “PROBLEMA DEL MUESTREO” PARA EL CÁLCULO DE ENERGÍA LIBRE	88
III.3 OBTENCIÓN DE ENERGÍA LIBRE A PARTIR DE MEDIDAS DE TRABAJO.	90
III.3.1 <i>¿Qué significa medir un conjunto de trabajos en un experimento de simulación de Dinámica Molecular?</i>	92
III.4 DEMOSTRACIÓN DE LA IGUALDAD DE JARZYNSKI	93
III.5 APROXIMACIÓN DE RESORTE RÍGIDO EN LAS SIMULACIONES DE SMD	97
III.6 EL PROBLEMA DE LA ESTIMACIÓN DE LA ENERGÍA LIBRE A TRAVÉS DE LA ECUACIÓN DE JARZYNSKI	100
III.7 OBJETIVOS Y LINEAMIENTOS	105
III.8 RESULTADOS	106
III.8.1 <i>Estimación de máxima-verosimilitud para distribuciones Gaussianas</i>	106
III.8.2 <i>Simulaciones numéricas de los estimadores de Jarzynski y de máxima verosimilitud</i>	107
III.8.3 <i>Distribuciones de trabajo Gaussianas</i>	109
III.8.4 <i>Distribuciones de trabajo no-Gaussianas</i>	110
III.8.5 <i>Caso ejemplar: migración de superóxido en MnSOD humana.</i>	111
III.9 DISCUSIÓN	113
III.10 REFERENCIAS	118
IV. REACTIVIDAD QUÍMICA DE ACIDO SULFÍDRICO (H₂S) EN HEMOPROTEÍNAS.	121
IV.1 INTRODUCCIÓN	121
IV.2 METODOLOGÍA ESPECÍFICA	123
IV.2.1 <i>Preparación de los sistemas y parámetros para la simulación de Dinámica Molecular (MD)</i>	123

IV.2.2	<i>Cálculos de Muestreo de Ligando Implícito (ILS).</i>	126
IV.2.3	<i>Cálculos de Dinámica Molecular Guiada (sMD).</i>	127
IV.2.3.1	Criterio de convergencia de los cálculos de Dinámica Molecular Guiada.	128
IV.2.4	<i>Caracterización de los sitios de hidratación (Water Sites, WS).</i>	129
IV.2.5	<i>Cálculos QM/MM: optimizaciones restringidas</i>	130
IV.2.6	<i>Cálculos QM/MM: dinámica molecular.</i>	131
IV.3	RESULTADOS	132
IV.3.1	<i>Estabilidad de las dinámicas moleculares</i>	132
IV.3.2	<i>Caracterización de canales internos de migración.</i>	133
IV.3.3	<i>Perfil de energía libre para la migración de especies de sulfuro (H_2S vs HS^-)</i>	135
IV.3.4	<i>Formación de sitios de hidratación (WS)</i>	138
IV.3.5	<i>Primer paso de desprotonación: de $Fe(III)-SH_2$ a $Fe(III)-SH^-$</i>	140
IV.3.6	<i>Segundo paso de desprotonación: de $Fe(III)-SH^-$ al $Fe(III)-S^{2-}$</i>	142
IV.4	DISCUSIÓN	143
IV.5	REFERENCIAS	146
V.	EFFECTO DE LA TEMPERATURA EN LA MIGRACIÓN Y UNIÓN DE LIGANDOS A HEMOPROTEÍNAS	152
V.1	INTRODUCCIÓN	152
V.2	METODOLOGÍA ESPECÍFICA	153
V.3	RESULTADOS	158
V.3.1	<i>Evolución temporal de las trazas de re-unión de ligando CO</i>	158
V.3.2	<i>Obtención de las constantes cinéticas de asociación (k_{on}) de CO.</i>	161
V.3.3	<i>Mediciones de k_{on} en función de la temperatura y el tratamiento de Eyring.</i>	163
V.4	DISCUSIÓN	168
V.5	REFERENCIAS	169
VI.	CONCLUSIONES GENERALES	171
VII.	PUBLICACIONES	173
VII.1	PUBLICACIONES ENMARCADAS EN LOS TEMAS DE LA PRESENTE TESIS	173
VIII.	AGRADECIMIENTOS	174

I. INTRODUCCIÓN GENERAL

I.1 Proteínas

Las biomoléculas más importantes dentro de todos los organismos vivos son las proteínas (del griego πρωτεῖος, “prōteîos”, prominente), llamadas así por su carácter sobresaliente en la biología de cualquier organismo vivo. Desde su descubrimiento y a partir de la caracterización de su gran ubicuidad, se las han descrito como máquinas transportadoras, bio-reactores y sensores, replicadores, como material estructural, reguladores de metabolismo, y también como priones y componentes fundamentales en virus.^[1-8]

En términos estructurales, todas las proteínas tienen a sus aminoácidos constitutivos conformando lo que se conoce como “secuencia” o cadena aminoacídica. El aminoácido, es la unidad mínima en esa cadena y se une al siguiente o posterior aminoácido mediante un enlace de tipo amida (enlace peptídico). En algunas proteínas, además de aminoácidos, también se encuentran otras especies químicas (azúcares, lípidos, etc. unidas covalentemente o no, a la cadena) o incluso átomos o metálicos (como cobre, manganeso, hierro, zinc, etc).^[4]

Cada proteína estará entonces constituida por una secuencia de un largo y un orden determinados a partir de una combinación de veinte aminoácidos. A nivel espacial, estas características determinarán la estructura tridimensional que adopta cada proteína, es decir, la forma en que está plegada la proteína. A la disposición espacial de la cadena aminoacídica se la suele llamar “plegamiento” de la proteína.

En conclusión, la relevancia extrema que las proteínas adquieren en diversos procesos biológicos, su generalidad estructural, así como también su especificidad funcional, motiva a la comunidad científica al estudio profundo de la relación entre estructura y función de las proteínas. Esta relación, de carácter íntimo, es la responsable de la modulación y regulación, de los efectos de la proteína en los procesos biológicos estudiados.

I.1.1 Hemoproteínas

Un grupo especial de proteínas, las llamadas hemoproteínas, están presentes en todos los seres vivos y han adquirido una gran importancia a partir de la determinación de la estructura

de tridimensional de la hemoglobina y mioglobina^[9,10], y debido a la enorme diversidad de funciones encontradas en otras hemoproteínas.^[11–14]

Este conjunto de proteínas se caracteriza por poseer todas una molécula adicional a la cadena aminoacídica llamada grupo hemo, cual está unido (covalentemente o no) a algunos aminoácidos de la proteína.

Para las hemoproteínas se han descripto funciones muy variadas como transporte, almacenamiento y sensado de moléculas pequeñas (como las hemoglobinas, mioglobinas, nitroforinas, guanilato ciclasa), catálisis enzimática (como las peroxidasas, catalasas, óxido nítrico sintasa, flavohemoglobinas, hemoglobinas truncadas), transferencia de electrones (como los citocromos), entre otras^[15–25]. En casi todos los casos, el grupo hemo es el actor principal que caracteriza funcionalmente a la hemoproteína.

Existen en la naturaleza varios tipos de grupo hemo, los más habituales son los denominados a, b y c.

En todos ellos, la unidad básica es el anillo porfirínico o porfirina (ver Figura 1). Esta estructura, es un anillo heterocíclico de 5 miembros (4 átomos de carbono y 1 de nitrógeno) llamados pirroles, unidos entre sí por medio de puentes carbonados.

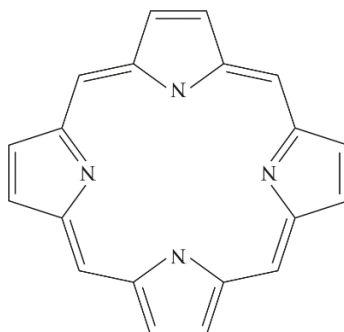


Figura 1: estructura de la porfirina, los átomos de hidrógeno y los grupos laterales fueron omitidos para mayor claridad.

Los tres tipos de hemo mencionados se distinguen entre sí por tener distintos grupos laterales unidos a la porfirina. Por ejemplo, el hemo c presenta uniones covalentes azufre-azufre con aminoácidos pertenecientes a la proteína.

El más habitual en la naturaleza es el grupo hemo b, que no está unido covalentemente a la proteína y tiene grupos laterales denominados como metilénicos, vinílicos y propiónicos como grupos laterales (ver Figura 2).

La orientación espacial de los cuatro átomos de nitrógeno de los anillos pirrólicos determina una cavidad, llamada sitio de coordinación, cuya arreglo espacial se denomina plano

cuadrado (ligandos ecuatoriales), en donde puede alojarse un metal (hierro, manganeso, zinc, etc). Para el caso de las hemoproteínas, este centro suele ser un átomo de hierro (Fe), el cual se presenta típicamente en los estados de oxidación ferroso (formalmente como Fe(II)) o férrico (formalmente como Fe(III)), aunque se han descrito otros estados de oxidación posibles (observados durante poco tiempo), como por ejemplo Fe(IV).^[26–28]

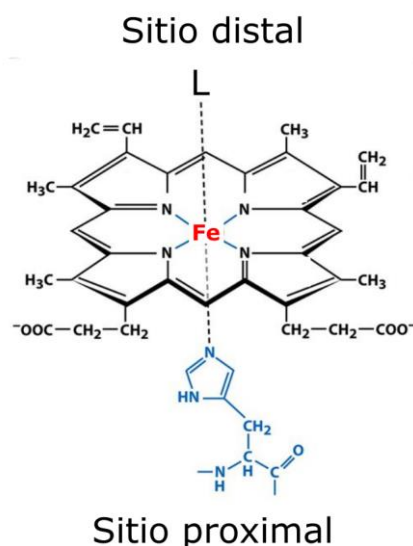


Figura 2: estructura general del grupo hemo b con el átomo de Fe coordinado por un aminoácido en la cara proximal (histidina) y un ligando genérico L, en la cara distal (además de los cuatro átomos de nitrógeno de los anillos pirrólicos). Se muestran también los grupos laterales metilénicos, vinílicos y propiónicos unidos a la porfirina.

El átomo de hierro central entonces permite dos posiciones axiales adicionales de coordinación de ligandos. Se define entonces un entorno o “sitio” proximal a la zona que contiene al denominado quinto sitio de coordinación y un entorno o “sitio” distal a la zona que contiene al denominado sexto sitio de coordinación (ver Figura 2).

En general, en las hemoproteínas el quinto sitio de coordinación (ubicado en el sitio proximal) es ocupado por un aminoácido. Por ejemplo en la mioglobina, la leghemoglobina y la neuroglobina, el aminoácido proximal es una histidina^[19,29,30] (como se muestra en la Figura 2). En otros casos, como en el citocromo P450 o en la óxido nítrico sintasa es una cisteína^[31], mientras que en la catalasa de *E. coli* se ha encontrado una tirosina en esta posición.^[32]

El sexto sitio de coordinación (ubicado en el sitio distal) suele corresponder al sitio donde pequeños ligandos (como CO, NO, O₂, etc) pueden coordinarse (cara distal en Figura 2). En algunos casos esta posición puede estar ocupada por moléculas de agua o por otro aminoácido (por ejemplo por una histidina en neuroglobina^[33]). Como se verá en las secciones posteriores en este capítulo, pueden bloquear el acceso al hierro de los pequeños ligandos.

Dependiendo de la naturaleza de los ligandos y el estado de oxidación del hierro, la intensidad del enlace entre el hierro y el ligando (Fe-L) será diferente. Por ejemplo, para Fe-CO, Fe-NO y Fe-O₂ el enlace es más fuerte con el hierro en estado ferroso mientras que para Fe-CN⁻, Fe-N³⁻, Fe-SH₂ el enlace es más intenso en el estado férrico. Por otro lado, para los aminoácidos habituales que ocupan el quinto sitio de coordinación, la unión puede tener intensidades muy variables, pero debido a restricciones geométricas que la proteína imprime sobre el aminoácido, este se mantiene unido al hierro (salvo casos excepcionales como en guanilato ciclasa).^[34]

Desde el punto de vista estructural, las hemoproteínas pueden dividirse en dos grandes grupos distinguidos por su plegamiento característico. Existe una familia denominada globinas, que presentan un plegamiento particular llamado “3 sobre 3” (3/3); constituido por 3 hélices α (llamadas hélices A, B, E) enfrentadas a otras 3 hélices α (F, G, H), con el grupo hemo en el centro (ver Figura 3). Este grupo de hemoproteínas, se encuentran presentes en bacterias, arqueas y en organismos eucarióticos. Algunos de los miembros más representativos de esta familia son la mioglobina, la neuroglobina, y la leghemoglobina.^[19,20,35]

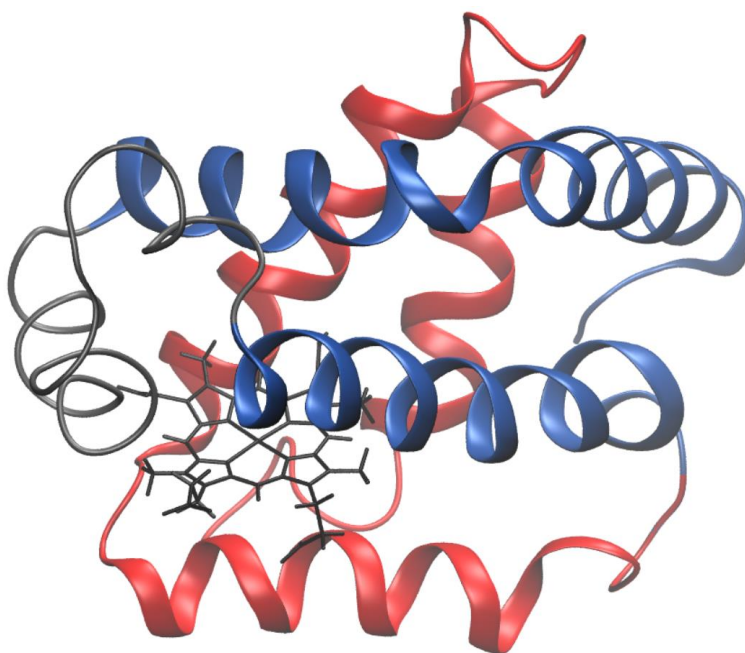


Figura 3: Plegamiento 3/3 típico de globinas (hélices A, B, E en azul y F, G, H en rojo) con el grupo hemo (en negro)

La otra familia de hemoproteínas, descubierta mucho más recientemente^[36], corresponde a las llamadas hemoglobinas truncadas (trHb)^[37], ya que comparten un plegamiento muy similar a las globinas, pero poseen entre 20 y 40 aminoácidos menos en su secuencia y carecen de dos de sus hélices. Estas proteínas se encuentran presentes en general, en eubacterias, cianobacterias, protozoos y plantas^[37]. Al plegamiento de este grupo particular de hemoproteínas se lo llama “2 sobre 2” (2/2, ver Figura 4).

Se les han asignado funciones que van desde sensores de O₂ y NO, transferencia de electrones, catálisis de reacciones redox, entre otras^[36].

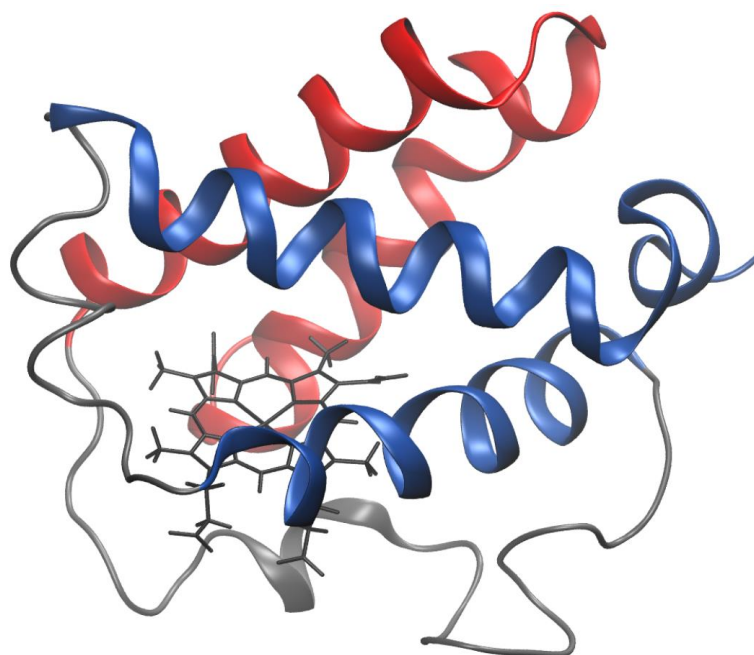


Figura 4: Plegamiento 2/2 típico de hemoglobinas truncadas (hélices B, E en azul y G, H en rojo) con su grupo hemo (en negro)

Para diferenciar las distintas hemoglobinas truncadas que posee un mismo organismo se las identifica con una letra accesoria: N, O ó P; (o I, II ó III) dependiendo del grupo filogenético al que pertenece.^[36]

Una característica fundamental que presentan las hemoglobinas truncadas, es la alta identidad de secuencia intra-grupal (por ejemplo para *M. tuberculosis* O y *M. avium* O la identidad es del 84%) y la baja identidad inter-grupal, (la identidad entre *M. tuberculosis* N y la *M. tuberculosis* O es de 18%).

En las proteínas la estructura se encuentra evolutivamente más conservada que la secuencia, es decir que proteínas con secuencias muy diferentes pueden tener la misma estructura tridimensional e inclusive cumplir las mismas funciones. Por esta razón, a los aminoácidos más conservados en las hemoproteínas se les asigna un nombre y una posición según su ubicación dentro de la estructura proteica, por ejemplo TrpG8, AlaE7, etc. El resto de los aminoácidos se los suele identificar con un número correspondiente a su posición en la secuencia de la proteína, por ejemplo Met92, Tyr36, etc.

Aunque estructuralmente las hemoglobinas son muy similares entre sí, se ha observado que cambios puntuales en aminoácidos del sitio activo modulan el comportamiento de la proteína frente a ligandos diatómicos^[36,37]. Por ejemplo, cambios puntuales en posiciones

específicas (como por ejemplo en B10, CD1, E7, etc) pueden afectar tanto la estabilización de ligandos coordinados al hierro, así como también su capacidad para acceder al sitio activo (por ejemplo afectando los canales de migración de las proteínas). De esta forma, las mutaciones puntuales de aminoácidos pueden ser determinantes en la actividad de la proteína.

Por todo lo anterior, las globinas y las hemoglobinas truncadas resultan ser un excelente sistema de estudio para comprender los factores fisicoquímicos que modulan la actividad de las hemoproteínas. Las conclusiones obtenidas mediante el estudio de casos claves podrán ser muy valiosas para poder extrapolar al resto de los miembros de la familia.

I.2 Unión y disociación de pequeños ligandos.

Muchos procesos biológicos relacionados con las hemoproteínas están vinculados a su reactividad frente a pequeños ligandos, como CO, NO, O₂, H₂S, etc.^[15,18,21,25,38].

La actividad frente a ligandos pequeños está determinada por varios factores: por un lado, la migración del ligando desde el solvente hasta el sitio activo a través de la matriz proteica y la posterior unión del ligando al grupo hemo. Luego, y dependiendo del ligando, de la proteína y del ambiente en que se encuentre (presencia de otros ligandos) el ligando podrá reaccionar o simplemente disociarse del grupo hemo y en última instancia volver al solvente.

La relación entre las velocidades de migración, unión, reactividad y de disociación de los ligandos determinará la actividad global de la hemoproteína frente a estos ligandos.

La migración dependerá de la accesibilidad para llegar al sitio activo a través de la matriz proteica; la unión dependerá de la accesibilidad por alcanzar el hierro del grupo hemo (muchas veces bloqueado por moléculas de agua o aminoácidos de la propia proteína); la reactividad dependerá de diferentes variables tales como las características de la interacción del ligando con el hemo, la presencia de aguas (o aminoácidos) que participan en el proceso, la presencia de otras moléculas de ligando, etc; y la disociación dependerá de la estabilización del ligando unido al grupo hemo e interacciones con aminoácidos cercanos del sitio distal.

Los procesos de migración se ven favorecidos por la presencia de los denominados “canales” de migración^[39–41], definidos como zonas de la proteína de baja densidad electrónica que conectan el sitio activo con el solvente (ver Figura 5), que favorecen el ingreso de los ligandos hasta el sitio activo donde se encuentra el grupo hemo.

Para el caso de las hemoglobinas truncadas por ejemplo, tanto desde un punto de vista experimental como teórico, se han descrito tres canales principales^[36,42–46] (Figura 5): un canal largo (LT) topológicamente posicionado entre las hélices B y E, y dos canales cortos, conocidos como la puerta E7 (E7 gate) y el canal G8 (STG8), llamados así por el rol clave que tienen la histidina (en la posición E7) y el triptófano (en la posición G8), respectivamente.

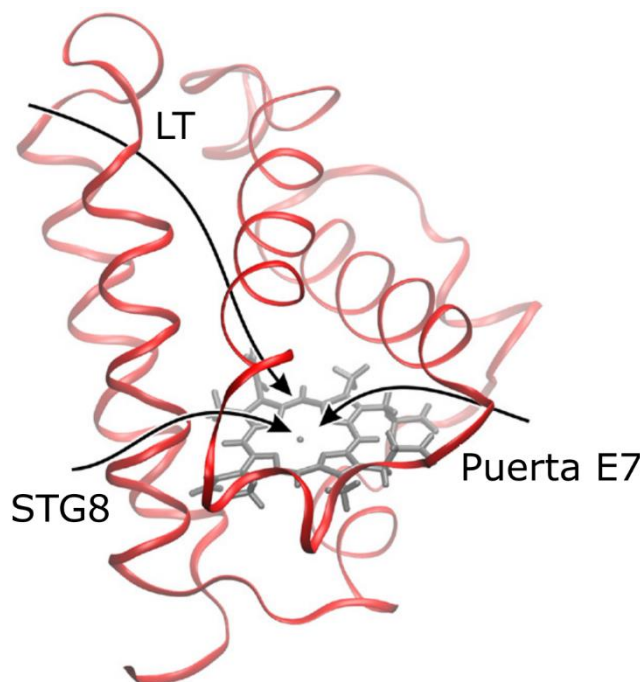


Figura 5: esquema de los tres canales de migración (puerta E7, STG8 y LT) observado en general en las hemoglobinas truncadas.

El canal E7 corresponde al camino altamente conservado y ampliamente estudiado en mioglobina y hemoglobina^[40,41,47]. Se lo ha descrito como un canal polar debido a la cercanía de los grupos propionatos del hemo y porque está ubicado en la cara más expuesta del sitio activo al solvente. El canal STG8 sin embargo es hidrofóbico y en algunos casos, se ha observado que el TrpG8 bloquea la migración de ligandos a través de este canal y del canal LT^[48–50]

Los procesos de migración se pueden ver afectados por diversos factores. En primer lugar, las moléculas de agua pueden jugar un rol fundamental en la velocidad de entrada y de salida de algunos ligandos. Por ejemplo, se ha mostrado que moléculas de agua internas (estabilizadas por residuos proteicos) pueden bloquear la accesibilidad al hemo a través de los canales de migración, retrasando la migración de los ligandos^[51–55]. También, la velocidad de migración se verá afectada, si el ligando requiere de un proceso de desolvatación parcial para el ingreso al sitio activo, sobre todo para ligandos cargados.

Para el proceso de unión, muchas veces el ligando al ingresar al sitio activo requiere de un desplazamiento de moléculas de agua retenidas en posiciones clave del sitio activo, como por ejemplo coordinadas al centro metálico.

Por otro lado, por el eventual requerimiento de un cambio en el estado de spin en la formación o ruptura del enlace Fe-L^[36] puede modular también el proceso de unión y disociación de los ligandos.

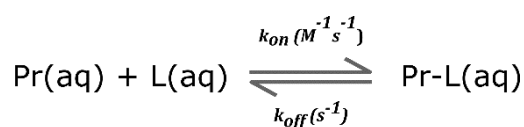
Finalmente, las interacciones que estabilicen (o desestabilicen) el complejo hexacoordinado (o pentacoordinado) modularán el proceso de disociación. Por ejemplo, las más habituales son las interacciones de tipo puente de hidrógeno que estabilizan el complejo hexacoordinado modulando así la disociación^[18,56–58]. Por ejemplo, debido a una interacción específica con el hierro (llamada retrodonación π), el O₂ coordinado al grupo hemo en estado ferroso adquiere una carga parcial negativa permitiéndole interactuar con grupos cargados positivamente de aminoácidos cercanos en el sitio distal.^[56,59]

Se ha demostrado también que las modificaciones electrónicas que podría sufrir el grupo hemo debido a distorsiones en el anillo porfirínico en algunos casos puede ser un factor modulador de la afinidad del ligando.^[60]

En general, es habitual referirse a la “unión” del ligando como al proceso global que engloba a los procesos de migración del solvente al sitio activo y a la unión del ligando al grupo hemo de la hemoproteína. Mientras que la “disociación” suele referirse al proceso que engloba a los procesos de disociación del ligando del grupo hemo y a la posterior migración del ligando del sitio activo al solvente.

I.2.1 Caracterización cinética y termodinámica de los procesos de unión y disociación de pequeños ligandos en hemoproteínas.

Los procesos globales de unión y disociación de un ligando en una proteína en solución acuosa a temperatura constante están gobernados por la siguiente reacción:



Donde Pr(aq), L(aq) y Pr-L(aq) se refieren a la proteína, ligando y el complejo proteína ligando, respectivamente, considerados todos en solución acuosa. Para la reacción como está escrita k_{on} es la constante cinética (bimolecular) de unión del ligando con la proteína y k_{off} es la constante cinética (unimolecular) para la disociación del ligando del complejo Pr-L.

Según la reacción anterior, k_{on} caracterizará la unión del ligando, incluyendo la etapa de migración del ligando desde el solvente hasta el sitio activo y la etapa de formación del enlace Fe-L. A su vez, k_{off} caracterizará la disociación del ligando incluyendo la etapa de ruptura del enlace Fe-L y la etapa de migración del ligando hasta el solvente.

El modelo de Eyring (ecuación (I.1)) es un modelo ampliamente utilizado en cinética química y específicamente en reacciones que involucran proteínas en solución acuosa^[61–64]. Este modelo establece una relación matemática entre las constantes cinéticas de velocidad y la temperatura, a través del cambio de energía libre entre reactivos y el estado de transición ($\Delta G^\ddagger$) de la reacción:

$$k = \left(\frac{k_b T}{h} \right) \cdot e^{-\frac{\Delta G^\ddagger}{RT}} \quad (I.1)$$

En donde T es la temperatura absoluta, R la constante universal de los gases, k_b y h son las constantes de Boltzmann y de Plank, respectivamente. $\Delta G^\ddagger$ es el cambio de energía libre entre reactivos y el estado de transición de la reacción.

Por otro lado, $\Delta G^\ddagger$ representa una medida del trabajo reversible (o de “costo energético”) para que los reactivos arriben a la conformación del estado de transición. En términos coloquiales, se le suele denominar “barrera” de energía libre. Las reacciones que tengan mayores $\Delta G^\ddagger$ serán entonces los que tengan constantes de velocidades más bajas.

En términos generales, los valores de las constantes cinéticas, se verán afectados por cada barrera de energía libre que el sistema debe superar para que suceda la reacción, las barreras de la etapa de migración y de unión en k_{on} y las barreras de la etapa de disociación y migración en k_{off} . Sin embargo, existirán barreras de energía libre de alguna etapa, que resulten significativamente mayor a las del resto de las etapas del proceso global. Para estos casos, la velocidad estará determinada por la etapa que presente mayor barrera de energía libre, es decir, por la etapa más lenta.

Un estudio reciente realizado por nuestro grupo a través de exhaustivo análisis estadístico, ha demostrado que de todos los factores mencionados anteriormente, la migración a través de los canales, y el desplazamiento de las moléculas de agua estabilizadas en el sitio

activo, son los principales factores que determinan el valor de k_{on} , y además, los valores de k_{off} se ven afectados principalmente por el costo energético de romper el enlace Fe-L.^[67,68] El trabajo mencionado muestra que en ambos casos los factores mencionados explican alrededor del 80% de la variabilidad del comportamiento cinético de todas las hemoglobinas truncadas reportadas. Como casos particulares, también se ha encontrado en algunas hemoproteínas que el factor principal que modula k_{on} es solo el bloqueo de aminoácidos en el canal de migración^[18,19,47,65], en otras es el desplazamiento de moléculas de aguas retenidas en el sitio activo^[66,67]. En otros casos, el costo energético de desolvatar el ligando para ingresar al sitio activo puede disminuir k_{on} , especialmente para ligandos cargados.

Por otra parte, cuando la proteína en presencia del ligando están en equilibrio, la velocidad de unión y disociación del ligando son iguales. En estas condiciones, se define la “afinidad” del ligando por la proteína como el valor numérico de la constante de equilibrio que según la ecuación química planteada adquiere la forma $K_d = [PrL]/\{[Pr].[L]\}$, donde la notación entre corchetes implica la concentración de la especie en el equilibrio. Además, en estas condiciones $K_d[M^{-1}] = k_{on}/k_{off}$.

La definición de K_d permite tener una noción más concreta de la “afinidad” del ligando por la hemoproteína y permite una comparación más global entre distintas proteínas y distintos ligandos. Por ejemplo, la hemoglobina I de *Lucina pectinata* en estado férrico tiene un valor de $K_d \approx 10^9 M^{-1}$ ^[70], mientras que $K_d \approx 10^6 M^{-1}$ para *T. fusca*-TrHbO y *B. subtilis*-TrHbO ambas en estado férrico^[70]. Como puede observarse, la afinidad por H₂S en HbI de *Lucina pectinata*-HbI es significativamente mayor que para las otras proteínas, esta característica es general y la distingue de todas las hemoproteínas reportadas hasta el momento.

Los valores de afinidades para la unión de O₂ a las mismas proteínas en los estados ferrosos son $K_d \approx 10^6 M^{-1}$ para *Lucina pectinata*-HbI, $\sim 10^7 M^{-1}$ para *T. fusca*-TrHbO y $\sim 10^9 M^{-1}$ para *B. subtilis*-TrHbO^[70]. Estos valores, destacan la elevada afinidad de *Lucina pectinata*-HbI por el ligando H₂S. La destacable “avidez” de la hemoglobina I de *L. pectinata* por H₂S motivó el estudio del proceso de migración y unión de esta especie en dicha proteína, como se verá en el capítulo IV de la presente tesis.

A título comparativo, el orden de los valores típicos obtenidos para la afinidad en hemoproteínas, permite concluir que las barreras asociadas a k_{on} para un mismo ligando suelen ser menores que las barreras asociadas a k_{off} . De hecho, la diferencia radica en el elevado costo energético que suele requerirse para producir la ruptura del enlace Fe-L, sobre todo para complejos de tipo férricos.

La ecuación (I.1) muestra que la temperatura también resulta ser un parámetro crucial, como modulador de la velocidad de un proceso químico. En general, los procesos asociados a una proteína podrían no responder al modelo de Eyring en todo el amplio rango de temperaturas encontrado en la biósfera (desde aproximadamente $-20\text{ }^{\circ}\text{C}$ a $113\text{ }^{\circ}\text{C}$ ^[69]), aún si las proteínas poseen la adecuada estabilidad para que su plegamiento persista a lo largo todo el rango de temperatura.

En general, para estudiar cómo se modifican los procesos en función de la temperatura, se miden las constantes de velocidad asociadas a distintas temperaturas y evalúa el modelo de Eyring observando la relación entre k y T . Este tipo de análisis resulta muy útil para evaluar el rango de temperatura para el cual la relación esperada cambia significativamente (evidenciada como cambios abruptos en el gráfico de k vs T o $\ln(k)$ vs $1/T$) y por lo tanto para las cuales el mecanismo asociado al proceso de estudio sufre una modificación sustancial.^[61–64]

I.2.2 Mediciones experimentales de las constantes de asociación (k_{on}) y disociación (k_{off})

Para medir experimentalmente las constantes cinéticas k_{on} y k_{off} resulta necesario establecer un parámetro observable que evidencie cambios en el sistema durante todo el rango de tiempo durante el cual se forma (y se disocia) el complejo hexacoordinado.

Los experimentos de tipo Stopped-Flow (flujo-detenido, SF), se basan en la mezcla directa (y rápida) de los reactivos y permiten la observación de los cambios de absorbancia de la solución durante el proceso de interés. En este caso, la medición de k_{on} se realiza a partir de la mezcla rápida entre soluciones de ligando y de la proteína^[71–77], observando la evolución temporal de la absorbancia de la solución a medida que se forma el complejo hexacoordinado.

De manera análoga, para la medición de k_{off} se suele usar experimentos denominados de “desplazamiento”; en los cuales un segundo ligando desplaza al ligando de interés del complejo hexacoordinado. Se suele partir del complejo hexacoordinado (formado con el ligando de interés) y se observan los cambios de absorbancia durante la formación del nuevo complejo hexacoordinado producto de la reacción de desplazamiento. Es requerimiento que el segundo ligando tenga una mayor afinidad por la proteína.

Estas técnicas resultan muy útiles para caracterizar eventos que suceden a partir de tiempos mayores a los pocos milisegundos y tienen la ventaja principal de usar poco volumen

de reactivos. La desventaja principal resulta en el tiempo de mezclado (llamado “tiempo muerto”), que es el tiempo mínimo para que se mezclen totalmente los reactivos y pueda medirse correctamente la absorbancia.^[78] Por lo tanto durante el tiempo muerto (cuyo rango típico es 1-5 ms) no es posible una detección adecuada de la absorbancia.

Otra técnica experimental que suele utilizarse es la llamada láser flash fotólisis (LFP).^[79-83] En esta técnica, se parte de soluciones en donde conteniendo el complejo hexacoordinado con el ligando de interés y mediante la incidencia de un láser, se produce una fotólisis del complejo, generando una población de complejo pentacoordinado y ligando no unido la hemo.

Finalmente, a lo largo del tiempo el ligando se re-unirá hasta restablecer la concentración inicial del complejo hexacoordinado. La cinética de interés se obtiene observando el tiempo requerido para la re-unión del ligando, midiendo las variaciones de la absorbancia de la solución. En general, en estos experimentos suele medirse la diferencia de absorbancias entre el complejo pentacoordinado y hexacoordinado.

La ventaja principal de utilizar experimentos de LFP consiste fundamentalmente en que se pueden caracterizar eventos que suceden durante tiempos menores (100ns-0.1s) que las requeridas en la técnica de SF. Además, como se parte del complejo hexacoordinado, se evita la presencia de tiempos muertos asociados a la mezcla que establecen una cota inferior para el tiempo de medición.

Finalmente, es importante mencionar, que ambas técnicas permiten la medición de k_{on} y k_{off} a diferentes temperaturas, mediante dispositivos que permiten mantener la temperatura constante durante todo el experimento.

En ambos tipos de experimentos existen esquemas de medición que aprovechan el equilibrio dinámico entre la unión y la disociación del complejo para la medición conjunta de k_{on} y k_{off} , sin embargo bajo determinadas condiciones se pueden medir independientemente cada una de las constantes.

Por ejemplo, para la medición de k_{on} supongamos que se mezcla una solución de concentración inicial conocida de proteína (Pr_0) con una solución de concentración inicial conocida de ligando (L_T), y que además es mucho mayor que la de la proteína. La temperatura durante toda la reacción se considera constante.

La evolución temporal de la concentración de complejo ($Pr-L(t)$) y de proteína ($Pr(t)$) estará dada entonces por las siguientes relaciones:

$$Pr-L(t) = \frac{Pr_0 \cdot k_{on} \cdot L_T \cdot (1 - e^{-(k_{off} + k_{on} \cdot L_T)t})}{k_{on} \cdot L_T + k_{off}} \quad (I.2)$$

$$Pr(t) = \frac{Pr_0 \cdot (k_{off} - k_{on} \cdot L_T \cdot e^{-(k_{off} + k_{on} \cdot L_T)t})}{k_{on} \cdot L_T + k_{off}} \quad (I.3)$$

En donde se supuso que como la concentración de ligando es mucho mayor que la de la proteína, esta no cambia con el tiempo (es decir, L_T es constante).

Ahora bien, como se mencionó en la sección anterior los valores de k_{on} suelen mucho más grandes que los valores de k_{off} . Por ejemplo, los valores de k_{on} para ligandos y proteínas más habitualmente estudiados (como CO/O₂ y mioglobina/citoglobina/neuroglobina^[84], respectivamente), rondan el rango de $10^4 - 10^7 \text{ M}^{-1} \text{ s}^{-1}$, mientras que los valores de k_{off} se encuentran en el rango de $0.001-0.014 \text{ s}^{-1}$ para esos sistemas.^[84]

Por otro lado, los valores de solubilidad de los ligandos en condiciones normales de presión, en un rango de temperatura entre los 10 y 50 °C, rondan valores cercanos a 10^{-3} M .

Bajo este análisis cualitativo entonces, el valor de $k_{on} \cdot L_T$ es mucho más grande que el valor de k_{off} , por lo tanto las expresiones anteriores adquieren la siguiente forma:

$$Pr-L(t) = Pr_0 \cdot (1 - e^{-t \cdot (k_{on} \cdot L_T)}) \quad (I.4)$$

$$Pr(t) = Pr_0 \cdot (\frac{1}{K_d} - e^{-t \cdot (k_{on} \cdot L_T)}) \quad (I.5)$$

Lo anterior muestra que la evolución temporal de las especies intervinientes ($Pr-L(t)$ y $Pr(t)$ o equivalentemente, complejo hexacoordinado y pentacoordinado, respectivamente) estarán determinadas por el valor de $k_{on} \cdot L_T$ exclusivamente. Este valor se llama constante “aparente” de la reacción. Por lo tanto, a partir de la medición de la evolución temporal de la absorbancia (aparición de complejo hexacoordinado, desaparición del complejo pentacoordinado, o la diferencia entre ellos) y el valor de L_T , es posible entonces obtener el valor de k_{on} .

I.3 Importancia de la temperatura en la actividad de proteínas.

Otro de los aspectos fundamentales de la fisicoquímica de proteínas que consideraremos en esta tesis doctoral, es la influencia de la temperatura en las propiedades

dinámicas, estructurales y funcionales de las proteínas, específicamente en las hemoproteínas, en las que nos hemos concentrado.

En este sentido, la sensibilidad a la temperatura es un factor crucial que determina la viabilidad de los organismos en todos los hábitats y está determinada por la susceptibilidad de todos los procesos bioquímicos del organismo frente a cambios de temperatura, incluyendo la actividad de las proteínas.

En general, las proteínas requieren una cierta flexibilidad de la estructura para llevar a cabo el proceso que define su actividad de manera óptima. Estos dos factores se ven modulados por variaciones en la temperatura.^[85,86] Por ejemplo, para las proteínas cuya actividad requiere de la unión de ligandos, un aumento de la temperatura hace que las interacciones que estabilizan (tanto en la proteína como en el complejo proteína-ligando) se formen y se re-formen más rápidamente, además permite que los cambios conformacionales necesarios para la actividad se alcancen más rápidamente. Estos efectos producen un aumento en la velocidad del proceso de unión.

Sin embargo, si la temperatura continúa aumentando, las interacciones se van rompiendo y la proteína no puede mantener una conformación con la geometría tridimensional necesaria y, en última instancia, la estructura se despliega perdiendo la actividad^[87], en un proceso que se llama “desnaturalización” de la proteína. En relación a esto último, la temperatura para la cual se desnaturaliza el 50% de la proteína se la llama T_m , y los valores de T_m que presentan las proteínas caracterizan su termoestabilidad.

Para el caso contrario, si la temperatura disminuye del rango óptimo de actividad, las interacciones tiendan a fortalecerse con respecto a la energía cinética disponible, incluso las que inicialmente eran débiles. Se llega entonces a un estado sobre-estabilizado, incluso en las regiones que inicialmente eran móviles. La proteína en estos casos, suele mantener una estructura que le permita la unión de los ligandos pero como los grandes cambios conformacionales están limitados, la podría bajar.^[87]

Se postula entonces, que la temperatura juega un rol crucial como modulador del balance entre termoestabilidad y flexibilidad de las proteínas.^[88]

Desde un punto de vista evolutivo, se cree que la selección natural actúa en mantener este balance.^[85-87] Se observa que las mutaciones de aminoácidos operan en función de una situación de compromiso entre termoestabilidad y flexibilidad, de manera tal de mejorar al máximo la actividad de las proteínas en el entorno particular que el organismo habita^[85,89,90],

sin una disminución considerable en la termoestabilidad) en el¹ Como consecuencia, no siempre es esperable que una proteína presente una alta termoestabilidad y tenga una función proteica óptima, en función de las características de su secuencia aminoacídica y estructura.

Sin embargo, se han encontrado enzimas homólogas de organismos que viven en distintas condiciones de temperatura, que comparten los mismos mecanismos catalíticos, una alta identidad de secuencia y tienen un alto parecido estructural. Estas similitudes permiten pensar en que la adaptación diferencial a la temperatura no siempre es predecible a priori desde un punto de vista netamente estructural. En algunos casos, pueden existir otros factores (como concentración de ligandos, fuerza iónica del medio, capacidad de agregación y difusión proteica, entre otros) también afectados por la temperatura que deben tenerse en cuenta, sobre todo para una correcta interpretación de la relación entre la adaptación de un organismo y sus proteínas.

Lo mencionado anteriormente, demuestra que la discusión acerca de la relación entre la flexibilidad, estabilidad y función proteica sigue abierta.

De hecho, los principales factores fisicoquímicos que modulan la actividad de las proteínas, todavía son cuestiones claves que no están resueltas en el *estado del arte* de la fisicoquímica de proteínas. En los últimos tiempos, todas estas cuestiones resultaron tener interés creciente para el diseño racional de proteínas/enzimas modificadas, para la utilización en procesos biotecnológicos, mejorando por ejemplo la termoestabilidad.

El estudio de estos procesos en hemoproteínas pertenecientes a organismos adaptados a diferentes condiciones térmicas, podrá revelar comportamientos diferenciales frente a cambios de temperatura.

Por esta razón, en esta tesis hemos seleccionado tres hemoglobinas truncadas de la misma familia pertenecientes a una bacteria psicrófila (TrHbO de *P. haloplanktis*), una mesófila (TrHbO de *B. subtilis*), y una termófila (TrHbO de *T. fusca*), las que hemos caracterizado desde un punto de vista experimental para tratar de comprender las bases moleculares de la adaptación a condiciones extremas.

I.4 Un nuevo gasotransmisor: ácido sulfhídrico (H₂S).

Dentro de los ligandos de relevancia fisiológica (como O₂, NO, HNO y CO) las especies de sulfuro han adquirido una importancia creciente en relación a las hemoproteínas, debido al descubrimiento de su producción endógena en mamíferos^[91–98] y plantas.^[99–102]

Además, se le atribuyeron diversas funciones biológicas relacionadas con una reactividad asociada a compuestos endógenos de tipo tioles de bajo peso molecular^[103,104], a grupos tioles proteicos^[105–107], a metaloproteínas^[108] y en particular a hemoproteínas^[109–111], cuya interacción ha demostrado que impacta directamente en la síntesis de ATP y en la función de canales de potasio, con sus consecuentes efectos sobre el tono muscular y la actividad neuronal en mamíferos^[112–115].

En esta tesis hemos seleccionado el estudio del comportamiento de hemoproteínas frente al ligando sulfuro de hidrógeno (H₂S). Esto está motivado, además de sus implicancias biológicas, por el interés en comprender desde una perspectiva fisicoquímica, las bases moleculares de la enorme diferencia de reactividad frente a esta especie en diferentes hemoproteínas.

Varios estudios experimentales muestran que al mezclar soluciones reguladas con concentraciones equivalentes de H₂S y proteína se evidencia la formación de un complejo férrico hexacoordinado estable de bajo spin y las constantes de afinidad para los complejos tipo Fe(III)-(sulfuro)^[38,72,110] descritas hasta la actualidad van desde 10⁹ M⁻¹ (hemoglobina I de *L. pectinata*, HbI)^[38] a 10⁴ M⁻¹ (microperoxidasa 11, un undecapéptido derivado de la hidrólisis parcial del citocromo c y ampliamente empleado como sistema modelo, por carecer de interacciones distales)^[71] a pH fisiológico.

En particular, la afinidad por las especies de sulfuro, ha sido descrita extensamente por un lado, en términos de mecanismos de estabilización distal del ligando sulfuro coordinado al hemo. Mientras que por otro lado, el modelo de hemo microperoxidasa, que no permite una estabilización por interacción con aminoácidos distales, pone también en relevancia el rol de la histidina proximal para la estabilización del complejo, tanto en solución acuosa como en solventes orgánicos de menor polaridad^[121,122].

Por otro lado, se ha estudiado también la reactividad ante excesos de sulfuro en soluciones que contienen hemoproteínas en condiciones oxidativas y no oxidativas^{[70,72,109,116–}

^{120]}. En general, los resultados experimentales muestran que en ambas condiciones, a tiempos de reacción largos (del orden de horas) se observa la formación del estado ferroso de la proteína y la presencia de azufre coloidal suspendido en la solución.^[70,71,109]

Por un lado, la unión preferida por las especies de sulfuro a hemos férricos constituye una característica relevante que lo distingue de CO y NO (ambos poseen una unión favorecida a hemos de tipo ferrosos). Por otro lado, es indispensable destacar que además de los factores ya mencionados para la modulación de la migración y unión en hemoproteínas, resulta necesario tener en cuenta una característica adicional del H₂S que lo distingue del resto de los ligandos pequeños de relevancia fisiológica: su comportamiento ácido-base.

El diagrama de especiación de la Figura 6, muestra que H₂S presenta dos pKa (7 y 12, para la cupla H₂S/HS⁻ y HS⁻/S²⁻ respectivamente).

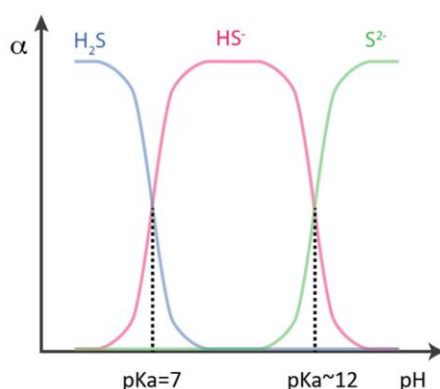


Figura 6: diagrama de especiación del H₂S en solución acuosa

El primer pKa se encuentra dentro del rango de pH que se considera biológicamente relevante (pH sanguíneo 7,35-7,45). En este rango de pH existirá en solución acuosa una fracción aproximada de H₂S y HS⁻ de 0.5. Con lo cual, si se estudia la interacción de las especies de sulfuro con hemoproteínas a en condiciones de pH en el rango de 6-8 unidades se deberá tener en cuenta en las interpretaciones la concentración apreciable de ambas especies.

Por otro lado, el pKa de la especie coordinada podría diferir de la forma libre debido a la interacción con el átomo de hierro del grupo hemo, y debido a la interacción con aminoácidos distales, moléculas de agua, etc.

Más aún, la especie coordinada podría ser diferente de la especie que migra desde el solvente hasta el sitio activo, ya que la migración y la unión del ligando son eventos que pueden ser considerados de manera independiente.

Algunos autores afirman que es H_2S la forma unida a la porfirina ferrosa/férrica de un modelo de citocromo C oxidasa^[123], así como también a la Hbl de *L. Pectinata*^[72]. Otros autores postulan que es HS^- la especie principal que se une a la hemoglobina de *T. fusca*^[124]. En todos los casos las evidencias experimentales que permiten identificar la especie coordinada resultan insuficientes.

El escenario entonces resulta intrigante, la correcta identificación de la especie de sulfuro (H_2S , HS^- o S^{2-}) que migra y permanece coordinada al Fe(III) debería estar bajo continuo escrutinio ya que resulta fundamental para comprender la afinidad y la subsecuente reactividad de estas especies en hemoproteínas.

I.5 Simulaciones computacionales para la obtención de información química.

Estudios en conjunto de métodos experimentales y teóricos, se abordan día a día para comprender la íntima relación entre estructura y función de las proteínas. Los aportes experimentales abordados en sistemas *in-vitro*, dan luz sobre aspectos estructurales y funcionales que resultan clave como punto de partida para interpretar, en última instancia, fenómenos asociados en sus sistemas *in-vivo* correspondientes.

Por otro lado, una completa dilucidación y comprensión de la relación estructura-función, requieren del uso complementario técnicas de simulación computacional que ayuden a interpretar, en un nivel más detallado, los fenómenos observados experimentalmente. Este enfoque también requiere, además del conocimiento técnico, un conocimiento adicional (de alcances y limitaciones) acerca de las herramientas físico-químicas que se disponen actualmente, así como también del desarrollo de nuevos modelos o interpretaciones derivados de ellas.

Dentro del conjunto de técnicas de simulación computacional, la rama de la química computacional se encarga de estudiar cualquier proceso químico sobre sistemas relevantes mimetizando condiciones experimentales habituales (como presión, temperatura, etc) asumiendo un dado modelo matemático para describir las interacciones presentes en el sistema.

En cualquier proceso químico de interés, resulta absolutamente relevante tanto estudiar su factibilidad (la termodinámica asociada al sistema) así como también la velocidad

con la que sucede dicho proceso (la cinética del sistema). Ambos aspectos son cruciales desde el punto de vista biológico: la correcta afinación entre ellos determinan la viabilidad de los organismos.

Por otro lado, el conocimiento sobre el cambio de energía libre mientras asociado al proceso (en general, descrito en función de alguna coordenada de transformación asociada), permite definir un perfil de energía libre (ver Figura 7). Esta medida provee la información necesaria sobre la factibilidad (cambio de energía libre negativo o positivo) y la velocidad (análisis de $\Delta G^\ddagger$) de manera conjunta como muestra la Figura 7. Por lo tanto, el cálculo del perfil de energía libre es uno de los objetivos fundamentales de la química computacional.

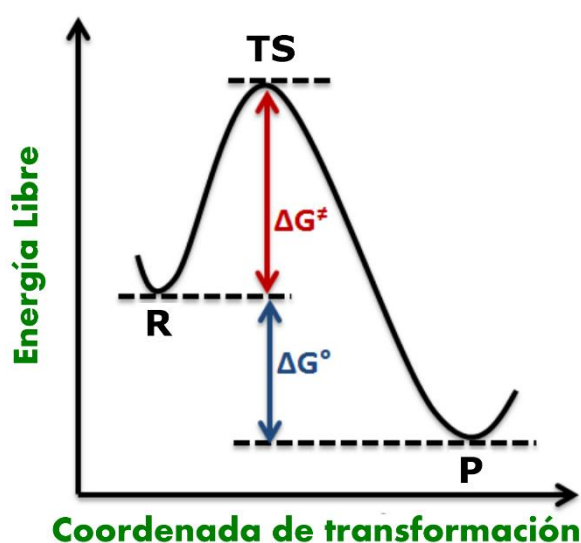


Figura 7: diagrama general de un perfil de energía libre para el pasaje de reactivos (R) a productos (P), a través de un estado de transición (TS)

Dentro de la química computacional, existen distintos niveles de cálculo que permiten una mejor descripción y generalidad, dependiendo del tipo de proceso químico que se desea simular. Para biomoléculas, como las proteínas por ejemplo, se pueden utilizar los métodos de simulación clásica (por ejemplo dinámica molecular) para estudiar procesos en donde no haya formación o ruptura de enlaces covalentes. En caso contrario, se han desarrollado en los últimos años métodos de tipo multi-escala, por ejemplo QM/MM en el cual se selecciona una “porción reactiva” del sistema (región en donde se forman y se rompen enlaces) que se describe con un nivel de cálculo mecano-cuántico, mientras que el resto del sistema se describe con un nivel clásico.

Por otro lado, en ambos conjuntos de métodos, la gran cantidad de grados de libertad que tiene el sistema de interés (por ejemplo un sistema típico que incluye proteína, ligandos y moléculas de solvente circundante) resultaría imposible obtener resultados con una simulación

convencional para tiempos de cómputo accesibles. Esto obliga a desarrollar técnicas que restrinjan el cálculo de energía libre a un sub-espacio del espacio configuracional del sistema.

Lo anterior es la idea principal de los métodos de simulación denominados “sesgados”, que son la base fundamental de los métodos de estimación de energía libre dentro de la simulación computacional de biomoléculas.

Los métodos sesgados se dividen fundamentalmente en dos grandes grupos: 1) Métodos de equilibrio, 2) Métodos de no-equilibrio.

Los primeros evalúan la energía libre suponiendo que el proceso global simulado (separado y analizado en sucesivos pasos intermedios del proceso global) se encuentra termalizado y equilibrado en el ensamble termodinámico deseado. Ejemplos de este tipo son los llamados de Integración Termodinámica^[125], Umbrella Sampling^[126], entre otros. Para estos casos, la energía libre de cada etapa simulada se evalúa a través del factor de Boltzmann asociado a cada método, teniendo en consideración el sesgo impuesto.

Los segundos evalúan el trabajo (W_i) requerido por el sistema para recorrer la coordenada de transformación mediante la aplicación de una fuerza externa. A partir de N medidas de trabajo (partiendo de distintas condiciones iniciales) es posible obtener el cambio de energía libre asociado a la transformación, empleando por ejemplo, la igualdad de Jarzynski^[127] (ecuación **(I.6)**).

$$e^{[-\beta\Delta G]} = \lim_{N \rightarrow \infty} \langle e^{[-\beta W_i]} \rangle_N \quad (\text{I.6})$$

Ejemplos de este conjunto de métodos son Perturbaciones de Energía Libre (FEP), Dinámica Molecular Guiada (sMD), entre otros^[128].

Para el caso de los procesos de unión y disociación de ligandos pequeños en hemoproteínas, se puede hacer uso de la técnica de muestreo de ligando implícito (ILS)^[65,129,130], en donde se considera al sistema en equilibrio y se describe a los ligandos como una pequeña perturbación del sistema. Este método calcula la energía libre en distintas posiciones del ligando en el sistema, considerando la presencia del ligando de manera implícita en el modelo matemática que describe las interacciones del sistema, de modo que utiliza una simulación de dinámica molecular del sistema en ausencia de ligando.

La técnica de ILS permite definir de manera cualitativa canales de migración, como las zonas de baja energía libre, aunque tienen la desventaja que no permiten diferenciar cuantitativamente entre ligandos. En este último caso, el uso de técnicas como de dinámica

molecular guiada (SMD) suelen dar resultados más esclarecedores. Sin embargo, la técnica de ILS es muy utilizada como método cualitativo para la identificación de canales de migración.

I.6 Objetivos y lineamientos de la presente tesis

En la presente tesis, se abordarán tres diferentes tipos de enfoque para el estudio de la migración y unión de ligandos en hemoproteínas, combinando una variedad de estrategias teóricas, de simulación computacional, y experimentales.

En primer lugar, la importancia que tienen los modelos matemáticos subyacentes a las técnicas de simulación computacional, obliga al continuo desarrollo de nuevos modelos o puntos de vista que mejoran el uso de esas técnicas. En el capítulo III se abordará un estudio teórico sobre una herramienta extensamente utilizada para el cálculo de perfiles de energía libre a partir de medidas de trabajo irreversible, la técnica de dinámica molecular guiada acoplada a la igualdad de Jarzynski. En particular, se estudiará las formas en que los perfiles de energía libre en este contexto son estimados. Se mostrará el impacto que tienen los resultados obtenidos aplicados al cálculo del perfil de energía libre asociado a la migración de O_2^- (anión superóxido) en manganeso superóxido dismutasa (MnSOD), la enzima más importante en la detoxificación de este anión. Para este sistema los resultados de simulación reportados mostraban un perfil de energía libre de migración del ligando inconsistente con resultados experimentales conocidos.

Por otro lado, dada la importancia biológica creciente que adquiere el H_2S como gasotransmisor y su interacción muchas veces íntimamente ligada a hemoproteínas, en el capítulo III se estudiará, utilizando técnicas de simulación computacional, la reactividad de especies de sulfuro de hidrógeno (H_2S y sus análogos deprotonados) en hemoproteínas. En particular, se estudiará la proteína Hbl de *L. pectinata*, una proteína que presenta una elevada afinidad por las especies de sulfuro. Por otro lado, se utilizarán también los resultados obtenidos en el capítulo IV para el estudio de los perfiles de migración de las especies de sulfuro en esta proteína.

Finalmente en el capítulo V se estudiará a través de mediciones de laser flash fotólisis, el efecto de la temperatura en la migración y unión de ligandos en hemoproteínas que pertenecen a organismos adaptados y no adaptados a la temperatura. Se tomarán como casos de estudio tres hemoglobinas truncadas (trHb): la trHb perteneciente al microorganismo termófilo *Thermobifida fusca* (Tf-trHbO), una trHb perteneciente al organismo mesófilo *Bacillus*

subtilis (Bs-trHbO) y una trHb perteneciente a un organismo psicrófilo, *Pseudoalteromonas haloplanktis* (Ph-trHbO).

I.7 Referencias

- (1) McLachlan, A. D. Protein Structure and Function. *Annual Review of Physical Chemistry* **1972**, 23 (1), 165–192.
- (2) Kühlbrandt, W. Structure and Function of Mitochondrial Membrane Protein Complexes. *BMC Biology* **2015**, 13 (1).
- (3) *Molecular Cell Biology*, 5th ed.; Lodish, H. F., Ed.; W.H. Freeman: New York, 2008.
- (4) *Molecular Biology of the Cell*, 4th ed.; Alberts, B., Ed.; Garland Science: New York, 2002.
- (5) Prusiner, S. B. Prion Diseases and Central Nervous System Degeneration. *Clinical Research* **1987**, 35 (3), 177–191.
- (6) Aguzzi, A.; Calella, A. M. Prions: Protein Aggregation and Infectious Diseases. *Physiological Reviews* **2009**, 89 (4), 1105–1152.
- (7) Lucas, W. Viral Capsids and Envelopes: Structure and Function. In *Encyclopedia of Life Sciences*; John Wiley & Sons, Ltd, Ed.; John Wiley & Sons, Ltd: Chichester, UK, 2010.
- (8) Shors, T. *Understanding Viruses*; Jones and Bartlett Publishers: Sudbury, Mass, 2009.
- (9) Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C. A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature* **1958**, 181 (4610), 662–666.
- (10) Perutz, M. F.; Rossmann, M. G.; Cullis, A. F.; Muirhead, H.; Will, G.; North, A. C. Structure of Haemoglobin: A Three-Dimensional Fourier Synthesis at 5.5-Å Resolution, Obtained by X-Ray Analysis. *Nature* **1960**, 185 (4711), 416–422.
- (11) Bredt, D. S.; Snyder, S. H. Nitric Oxide: A Physiologic Messenger Molecule. *Annual Review of Biochemistry* **1994**, 63 (1), 175–195.
- (12) Schulz, H.; Hennecke, H.; Thöny-Meyer, L. Prototype of a Heme Chaperone Essential for Cytochrome c Maturation. *Science* **1998**, 281 (5380), 1197–1200.
- (13) Hou, S.; Larsen, R. W.; Boudko, D.; Riley, C. W.; Karatan, E.; Zimmer, M.; Ordal, G. W.; Alam, M. Myoglobin-like Aerotaxis Transducers in Archaea and Bacteria. *Nature* **2000**, 403 (6769), 540–544.
- (14) Qi, Z.; Hamza, I.; O'Brian, M. R. Heme Is an Effector Molecule for Iron-Dependent Degradation of the Bacterial Iron Response Regulator (Irr) Protein. *Proceedings of the National Academy of Sciences* **1999**, 96 (23), 13056–13061.
- (15) Epstein, F. H.; Hsia, C. C. W. Respiratory Function of Hemoglobin. *New England Journal of Medicine* **1998**, 338 (4), 239–248.
- (16) Poulos, T. L. Heme Enzyme Structure and Function. *Chemical Reviews* **2014**, 114 (7), 3919–3962.
- (17) Paoli, M.; Marles-Wright, J.; Smith, A. Structure–Function Relationships in Heme-Proteins. *DNA and Cell Biology* **2002**, 21 (4), 271–280.
- (18) Birukou, I.; Schweers, R. L.; Olson, J. S. Distal Histidine Stabilizes Bound O₂ and Acts as a Gate for Ligand Entry in Both Subunits of Adult Human Hemoglobin. *Journal of Biological Chemistry* **2010**, 285 (12), 8840–8854.
- (19) Scott, E. E.; Gibson, Q. H.; Olson, J. S. Mapping the Pathways for O₂ Entry Into and Exit from Myoglobin. *Journal of Biological Chemistry* **2001**, 276 (7), 5177–5188.
- (20) Vinogradov, S. N.; Moens, L. Diversity of Globin Function: Enzymatic, Transport, Storage, and Sensing. *Journal of Biological Chemistry* **2008**, 283 (14), 8773–8777.
- (21) Andersen, J. F.; Ding, X. D.; Balfour, C.; Shokhireva, T. K.; Champagne, D. E.; Walker, F. A.;

- Montfort, W. R. Kinetics and Equilibria in Ligand Binding by Nitrophorins 1–4: Evidence for Stabilization of a Nitric Oxide–Ferriheme Complex through a Ligand-Induced Conformational Trap[†]. *Biochemistry* **2000**, *39* (33), 10118–10131.
- (22) Martí, M. A.; Lebrero, M. C. G.; Roitberg, A. E.; Estrin, D. A. Bond or Cage Effect: How Nitrophorins Transport and Release Nitric Oxide. *Journal of the American Chemical Society* **2008**, *130* (5), 1611–1618.
- (23) Boon, E. M.; Marletta, M. A. Ligand Discrimination in Soluble Guanylate Cyclase and the H-NOX Family of Heme Sensor Proteins. *Current Opinion in Chemical Biology* **2005**, *9* (5), 441–446.
- (24) Perutz, M. F.; Paoli, M.; Lesk, A. M. Fix L, a Haemoglobin That Acts as an Oxygen Sensor: Signalling Mechanism and Structural Basis of Its Homology with PAS Domains. *Chemistry & Biology* **1999**, *6* (11), R291–R297.
- (25) Ibrahim, M.; Kuchinkas, M.; Youn, H.; Kerby, R. L.; Roberts, G. P.; Poulos, T. L.; Spiro, T. G. Mechanism of the CO-Sensing Heme Protein CooA: New Insights from the Truncated Heme Domain and UVRR Spectroscopy. *Journal of Inorganic Biochemistry* **2007**, *101* (11–12), 1776–1785.
- (26) Gumiero, A.; Metcalfe, C. L.; Pearson, A. R.; Raven, E. L.; Moody, P. C. E. Nature of the Ferryl Heme in Compounds I and II. *Journal of Biological Chemistry* **2011**, *286* (2), 1260–1268.
- (27) Oertling, W. A.; Kean, R. T.; Wever, R.; Babcock, G. T. Factors Affecting the Iron-Oxygen Vibrations of Ferrous Oxy and Ferryl Oxo Heme Proteins and Model Compounds. *Inorganic Chemistry* **1990**, *29* (14), 2633–2645.
- (28) Hohenberger, J.; Ray, K.; Meyer, K. The Biology and Chemistry of High-Valent Iron–oxo and Iron–nitrido Complexes. *Nature Communications* **2012**, *3*, 720.
- (29) Sawai, H.; Makino, M.; Mizutani, Y.; Ohta, T.; Sugimoto, H.; Uno, T.; Kawada, N.; Yoshizato, K.; Kitagawa, T.; Shiro, Y. Structural Characterization of the Proximal and Distal Histidine Environment of Cytoglobin and Neuroglobin. *Biochemistry* **2005**, *44* (40), 13257–13265.
- (30) Pesce, A.; Nardini, M.; Ascenzi, P.; Geuens, E.; Dewilde, S.; Moens, L.; Bolognesi, M.; Riggs, A. F.; Hale, A.; Deng, P.; Nienhaus, G. U.; Olson, J. S.; Nienhaus, K. Thr-E11 Regulates O₂ Affinity in Cerebratulus Lacteus Mini-Hemoglobin. *Journal of Biological Chemistry* **2004**, *279* (32), 33662–33672.
- (31) Rowland, P.; Blaney, F. E.; Smyth, M. G.; Jones, J. J.; Leydon, V. R.; Oxbrow, A. K.; Lewis, C. J.; Tennant, M. G.; Modi, S.; Eggleston, D. S.; Chenery, R. J.; Bridges, A. M. Crystal Structure of Human Cytochrome P450 2D6. *Journal of Biological Chemistry* **2006**, *281* (11), 7614–7622.
- (32) Melik-Adamyan, W.; Bravo, J.; Carpena, X.; Switala, J.; Maté, M. J.; Fita, I.; Loewen, P. C. Substrate Flow in Catalases Deduced from the Crystal Structures of Active Site Variants of HP11 from Escherichia Coli. *Proteins* **2001**, *44* (3), 270–281.
- (33) Kakar, S.; Hoffman, F. G.; Storz, J. F.; Fabian, M.; Hargrove, M. S. Structure and Reactivity of Hexacoordinate Hemoglobins. *Biophysical Chemistry* **2010**, *152* (1–3), 1–14.
- (34) Bikiel, D. E.; Boechi, L.; Capece, L.; Crespo, A.; De Biase, P. M.; Di Lella, S.; González Lebrero, M. C.; Martí, M. A.; Nadra, A. D.; Perissinotti, L. L.; Scherlis, D. A.; Estrin, D. A. Modeling Heme Proteins Using Atomistic Simulations. *Physical Chemistry Chemical Physics* **2006**, *8* (48), 5611–5628.
- (35) Vinogradov, S. N.; Hoogewijs, D.; Bailly, X.; Mizuguchi, K.; Dewilde, S.; Moens, L.; Vanfleteren, J. R. A Model of Globin Evolution. *Gene* **2007**, *398* (1–2), 132–142.
- (36) Milani, M.; Pesce, A.; Nardini, M.; Ouellet, H.; Ouellet, Y.; Dewilde, S.; Bocedi, A.; Ascenzi, P.; Guertin, M.; Moens, L. Structural Bases for Heme Binding and Diatomic Ligand Recognition in Truncated Hemoglobins. *Journal of Inorganic Biochemistry* **2005**, *99* (1), 97–109.
- (37) Wittenberg, J. B.; Bolognesi, M.; Wittenberg, B. A.; Guertin, M. Truncated Hemoglobins: A New Family of Hemoglobins Widely Distributed in Bacteria, Unicellular Eukaryotes, and

- Plants. *Journal of Biological Chemistry* **2002**, 277 (2), 871–874.
- (38) Kraus, D. W.; Wittenberg, J. B. Hemoglobins of the *Lucina Pectinata*/bacteria Symbiosis. I. Molecular Properties, Kinetics and Equilibria of Reactions with Ligands. *Journal of Biological Chemistry* **1990**, 265 (27), 16043–16053.
 - (39) Milani, M. Mycobacterium Tuberculosis Hemoglobin N Displays a Protein Tunnel Suited for O₂ Diffusion to the Heme. *The EMBO Journal* **2001**, 20 (15), 3902–3909.
 - (40) Elber, R. Ligand Diffusion in Globins: Simulations versus Experiment. *Current Opinion in Structural Biology* **2010**, 20 (2), 162–167.
 - (41) Perutz, M. F.; Mathews, F. S. An X-Ray Study of Azide Methaemoglobin. *Journal of Molecular Biology* **1966**, 21 (1), 199–202.
 - (42) Boron, I.; Bustamante, J. P.; Davidge, K. S.; Singh, S.; Bowman, L. A.; Tinajero-Trejo, M.; Carballal, S.; Radi, R.; Poole, R. K.; Dikshit, K.; Estrin, D. A.; Marti, M. A.; Boechi, L. Ligand Uptake in Mycobacterium Tuberculosis Truncated Hemoglobins Is Controlled by Both Internal Tunnels and Active Site Water Molecules. *F1000Research* **2015**.
 - (43) Crespo, A.; Martí, M. A.; Kalko, S. G.; Morreale, A.; Orozco, M.; Gelpi, J. L.; Luque, F. J.; Estrin, D. A. Theoretical Study of the Truncated Hemoglobin HbN: Exploring the Molecular Basis of the NO Detoxification Mechanism. *Journal of the American Chemical Society*. **2005**, 127 (12), 4433–4444.
 - (44) Fabozzi, G.; Ascenzi, P.; Renzi, S. D.; Visca, P. Truncated Hemoglobin GlbO from Mycobacterium Leprae Alleviates Nitric Oxide Toxicity. *Microbial Pathogenesis* **2006**, 40 (5), 211–220.
 - (45) Boechi, L.; Mañez, P. A.; Luque, F. J.; Marti, M. A.; Estrin, D. A. Unraveling the Molecular Basis for Ligand Binding in Truncated Hemoglobins: The trHbO *Bacillus Subtilis* Case. *Proteins: Structure, Function, and Bioinformatics* **2010**, 78 (4), 962–970.
 - (46) Bidon-Chanal, A.; Martí, M. A.; Crespo, A.; Milani, M.; Orozco, M.; Bolognesi, M.; Luque, F. J.; Estrin, D. A. Ligand-Induced Dynamical Regulation of NO Conversion in Mycobacterium Tuberculosis Truncated Hemoglobin-N. *Proteins: Structure, Function, and Bioinformatics* **2006**, 64 (2), 457–464.
 - (47) Boechi, L.; Arrar, M.; Marti, M. A.; Olson, J. S.; Roitberg, A. E.; Estrin, D. A. Hydrophobic Effect Drives Oxygen Uptake in Myoglobin via Histidine E7. *Journal of Biological Chemistry* **2013**, 288 (9), 6754–6762.
 - (48) Ouellet, H.; Milani, M.; LaBarre, M.; Bolognesi, M.; Couture, M.; Guertin, M. The Roles of Tyr(CD1) and Trp(G8) in Mycobacterium Tuberculosis Truncated Hemoglobin O in Ligand Binding and on the Heme Distal Site Architecture [†], [‡]. *Biochemistry* **2007**, 46 (41), 11440–11450.
 - (49) Boechi, L.; Martí, M. A.; Milani, M.; Bolognesi, M.; Luque, F. J.; Estrin, D. A. Structural Determinants of Ligand Migration in Mycobacterium Tuberculosis Truncated Hemoglobin O. *Proteins: Structure, Function, and Bioinformatics* **2008**, 73 (2), 372–379.
 - (50) Guallar, V.; Lu, C.; Borrelli, K.; Egawa, T.; Yeh, S.-R. Ligand Migration in the Truncated Hemoglobin-II from Mycobacterium Tuberculosis: the role of G8 tryptophan. *Journal of Biological Chemistry* **2009**, 284 (5), 3106–3116.
 - (51) Li, Z.; Lazaridis, T. The Effect of Water Displacement on Binding Thermodynamics: Concanavalin A. *The Journal of Physical Chemistry B* **2005**, 109 (1), 662–670.
 - (52) Li, Z.; Lazaridis, T. Thermodynamic Contributions of the Ordered Water Molecule in HIV-1 Protease. *Journal of the American Chemical Society* **2003**, 125 (22), 6636–6637.
 - (53) Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. Prediction of the Water Content in Protein Binding Sites. *The Journal of Physical Chemistry B* **2009**, 113 (40), 13337–13346.
 - (54) Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. Role of the Active-Site Solvent in the Thermodynamics of Factor Xa Ligand Binding. *Journal of the American Chemical Society* **2008**, 130 (9), 2817–2831.
 - (55) Schmidtke, P.; Luque, F. J.; Murray, J. B.; Barril, X. Shielded Hydrogen Bonds as Structural Determinants of Binding Kinetics: Application in Drug Design. *Journal of the American*

- Chemical Society* **2011**, 133 (46), 18903–18910.
- (56) Arroyo Mañez, P.; Lu, C.; Boechi, L.; Martí, M. A.; Shepherd, M.; Wilson, J. L.; Poole, R. K.; Luque, F. J.; Yeh, S.-R.; Estrin, D. A. Role of the Distal Hydrogen-Bonding Network in Regulating Oxygen Affinity in the Truncated Hemoglobin III from *Campylobacter Jejuni*. *Biochemistry* **2011**, 50 (19), 3946–3956.
 - (57) Martí, M. A.; Capece, L.; Bikiel, D. E.; Falcone, B.; Estrin, D. A. Oxygen Affinity Controlled by Dynamical Distal Conformations: The Soybean Leghemoglobin and the Paramecium Caudatum Hemoglobin Cases. *Proteins: Structure, Function, and Bioinformatics* **2007**, 68 (2), 480–487.
 - (58) Boubeta, F. M.; Bari, S. E.; Estrin, D. A.; Boechi, L. Access and Binding of H₂S to Hemeproteins: The Case of Hbl of *Lucina Pectinata*. *The Journal of Physical Chemistry B* **2016**, 120 (36), 9642–9653.
 - (59) Arcon, J. P.; Rosi, P.; Petruk, A. A.; Marti, M. A.; Estrin, D. A. Molecular Mechanism of Myoglobin Autoxidation: Insights from Computer Simulations. *The Journal of Physical Chemistry B* **2015**, 119 (5), 1802–1813.
 - (60) Bikiel, D. E.; Forti, F.; Boechi, L.; Nardini, M.; Luque, F. J.; Martí, M. A.; Estrin, D. A. Role of Heme Distortion on Oxygen Affinity in Heme Proteins: The Protoglobin Case. *The Journal of Physical Chemistry B* **2010**, 114 (25), 8536–8543.
 - (61) Mayor, U.; Johnson, C. M.; Daggett, V.; Fersht, A. R. Protein Folding and Unfolding in Microseconds to Nanoseconds by Experiment and Simulation. *Proceedings of the National Academy of Sciences* **2000**, 97 (25), 13518–13522.
 - (62) Sharma, G.; First, E. A. Thermodynamic Analysis Reveals a Temperature-Dependent Change in the Catalytic Mechanism of Bacillus Stearotherophilus Tyrosyl-tRNA Synthetase. *Journal of Biological Chemistry* **2009**, 284 (7), 4179–4190.
 - (63) Kwon, K.; Streaker, E. D.; Beckett, D. Binding Specificity and the Ligand Dissociation Process in the E. Coli Biotin Holoenzyme Synthetase. *Protein Science* **2009**, 11 (3), 558–570.
 - (64) Gul, S.; Mellor, G. W.; Thomas, E. W.; Brocklehurst, K. Temperature-Dependences of the Kinetics of Reactions of Papain and Actinidin with a Series of Reactivity Probes Differing in Key Molecular Recognition Features. *Biochemical Journal* **2006**, 396 (1), 17–21.
 - (65) Forti, F.; Boechi, L.; Estrin, D. A.; Marti, M. A. Comparing and Combining Implicit Ligand Sampling with Multiple Steered Molecular Dynamics to Study Ligand Migration Processes in Heme Proteins. *Journal of Computational Chemistry* **2011**, 32 (10), 2219–2231.
 - (66) Bustamante, J. P.; Radusky, L.; Boechi, L.; Estrin, D. A.; ten Have, A.; Martí, M. A. Evolutionary and Functional Relationships in the Truncated Hemoglobin Family. *PLOS Computational Biology* **2016**, 12 (1), e1004701.
 - (67) Bustamante, J. P.; Szretter, M. E.; Sued, M.; Martí, M. A.; Estrin, D. A.; Boechi, L. A Quantitative Model for Oxygen Uptake and Release in a Family of Hemeproteins. *Bioinformatics* **2016**, btw083.
 - (68) Ghosh, A. *The Smallest Biomolecules: Diatomics and Their Interactions with Heme Proteins*; Elsevier: Amsterdam; London, 2008.
 - (69) D'Amico, S. Activity-Stability Relationships in Extremophilic Enzymes. *Journal of Biological Chemistry* **2003**, 278 (10), 7891–7896.
 - (70) Pietri, R.; Román-Morales, E.; López-Garriga, J. Hydrogen Sulfide and Hemeproteins: Knowledge and Mysteries. *Antioxidants & Redox Signaling* **2011**, 15 (2), 393–404.
 - (71) Bieza, S. A.; Boubeta, F.; Feis, A.; Smulevich, G.; Estrin, D. A.; Boechi, L.; Bari, S. E. Reactivity of Inorganic Sulfide Species toward a Heme Protein Model. *Inorganic Chemistry* **2015**, 54 (2), 527–533.
 - (72) Pietri, R.; Lewis, A.; León, R. G.; Casabona, G.; Kiger, L.; Yeh, S.-R.; Fernandez-Alberti, S.; Marden, M. C.; Cadilla, C. L.; López-Garriga, J. Factors Controlling the Reactivity of Hydrogen Sulfide with Hemeproteins. *Biochemistry* **2009**, 48 (22), 4881–4894.
 - (73) Su, J.; Groves, J. T. Mechanisms of Peroxynitrite Interactions with Heme Proteins.

- Inorganic Chemistry* **2010**, 49 (14), 6317–6329.
- (74) Owens, C. P.; Du, J.; Dawson, J. H.; Goulding, C. W. Characterization of Heme Ligation Properties of Rv0203, a Secreted Heme Binding Protein Involved in *Mycobacterium Tuberculosis* Heme Uptake. *Biochemistry* **2012**, 51 (7), 1518–1531.
 - (75) Winter, M. B.; Herzik, M. A.; Kuriyan, J.; Marletta, M. A. Tunnels Modulate Ligand Flux in a Heme Nitric Oxide/oxygen Binding (H-NOX) Domain. *Proceedings of the National Academy of Sciences* **2011**, 108 (43), E881–E889.
 - (76) Zhao, Y.; Brandish, P. E.; Ballou, D. P.; Marletta, M. A. A Molecular Basis for Nitric Oxide Sensing by Soluble Guanylate Cyclase. *Proceedings of the National Academy of Sciences* **1999**, 96 (26), 14753–14758.
 - (77) Lange, R.; Heiber-Langer, I.; Bonfils, C.; Fabre, I.; Negishi, M.; Balny, C. Activation Volume and Energetic Properties of the Binding of CO to Hemoproteins. *Biophysical Journal* **1994**, 66 (1), 89–98.
 - (78) Kalidas, C. *Chemical Kinetic Methods: Principles of Fast Reaction Techniques and Applications*; New Age International: New Delhi, 2005.
 - (79) Abbruzzetti, S.; Tilleman, L.; Bruno, S.; Viappiani, C.; Desmet, F.; Van Doorslaer, S.; Coletta, M.; Ciaccio, C.; Ascenzi, P.; Nardini, M.; Bolognesi, M.; Moens, L.; Dewilde, S. Ligation Tunes Protein Reactivity in an Ancient Haemoglobin: Kinetic Evidence for an Allosteric Mechanism in Methanosarcina Acetivorans Protoglobin. *PLoS ONE* **2012**, 7 (3), e33614.
 - (80) Tilleman, L.; Abbruzzetti, S.; Ciaccio, C.; De Sanctis, G.; Nardini, M.; Pesce, A.; Desmet, F.; Moens, L.; Van Doorslaer, S.; Bruno, S.; Bolognesi, M.; Ascenzi, P.; Coletta, M.; Viappiani, C.; Dewilde, S. Structural Bases for the Regulation of CO Binding in the Archaeal Protoglobin from Methanosarcina Acetivorans. *PLOS ONE* **2015**, 10 (6), e0125959.
 - (81) Giordano, D.; Boron, I.; Abbruzzetti, S.; Van Leuven, W.; Nicoletti, F. P.; Forti, F.; Bruno, S.; Cheng, C.-H. C.; Moens, L.; di Prisco, G.; Nadra, A. D.; Estrin, D.; Smulevich, G.; Dewilde, S.; Viappiani, C.; Verde, C. Biophysical Characterisation of Neuroglobin of the Icefish, a Natural Knockout for Hemoglobin and Myoglobin. Comparison with Human Neuroglobin. *PLoS ONE* **2012**, 7 (12), e44508.
 - (82) Marcelli, A.; Abbruzzetti, S.; Bustamante, J. P.; Feis, A.; Bonamore, A.; Boffi, A.; Gellini, C.; Salvi, P. R.; Estrin, D. A.; Bruno, S.; Viappiani, C.; Foggi, P. Following Ligand Migration Pathways from Picoseconds to Milliseconds in Type II Truncated Hemoglobin from Thermobifida Fusca. *PLoS ONE* **2012**, 7 (7), e39884.
 - (83) Gabba, M.; Abbruzzetti, S.; Spyraakis, F.; Forti, F.; Bruno, S.; Mozzarelli, A.; Luque, F. J.; Viappiani, C.; Cozzini, P.; Nardini, M.; Germani, F.; Bolognesi, M.; Moens, L.; Dewilde, S. CO Rebinding Kinetics and Molecular Dynamics Simulations Highlight Dynamic Regulation of Internal Cavities in Human Cytoglobin. *PLoS ONE* **2013**, 8 (1), e49770.
 - (84) Taguchi, S. Binding of Oxygen and Carbon Monoxide to a Heme-Regulated Phosphodiesterase from Escherichia Coli: kinetics and infrared spectra of the full-length wild-type enzyme, isolated pas domain, and met-95 mutants. *Journal of Biological Chemistry* **2003**, 279 (5), 3340–3347.
 - (85) Fields, P. A. Review: Protein Function at Thermal Extremes: Balancing Stability and Flexibility. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* **2001**, 129 (2–3), 417–431.
 - (86) Somero, G. N. The Physiology of Climate Change: How Potentials for Acclimatization and Genetic Adaptation Will Determine “Winners” and “Losers.” *Journal of Experimental Biology* **2010**, 213 (6), 912–920.
 - (87) Fields, P. A.; Dong, Y.; Meng, X.; Somero, G. N. Adaptations of Protein Structure and Function to Temperature: There Is More than One Way to “Skin a Cat.” *Journal of Experimental Biology* **2015**, 218 (12), 1801–1811.
 - (88) Feller, G. Protein Stability and Enzyme Activity at Extreme Biological Temperatures. *Journal of Physics: Condensed Matter* **2010**, 22 (32), 323101.
 - (89) Jaenicke, R. Protein Stability and Molecular Adaptation to Extreme Conditions. *European*

- Journal of Biochemistry* **1991**, 202 (3), 715–728.
- (90) Berezovsky, I. N.; Shakhnovich, E. I. Physics and Evolution of Thermophilic Adaptation. *Proceedings of the National Academy of Sciences* **2005**, 102 (36), 12742–12747.
 - (91) Warenycia, M. W.; Goodwin, L. R.; Benishin, C. G.; Reiffenstein, R. J.; Francom, D. M.; Taylor, J. D.; Dieken, F. P. Acute Hydrogen Sulfide Poisoning. Demonstration of Selective Uptake of Sulfide by the Brainstem by Measurement of Brain Sulfide Levels. *Biochemical Pharmacology* **1989**, 38 (6), 973–981.
 - (92) Goodwin, L. R.; Francom, D.; Dieken, F. P.; Taylor, J. D.; Warenycia, M. W.; Reiffenstein, R. J.; Dowling, G. Determination of Sulfide in Brain Tissue by Gas Dialysis/Ion Chromatography: Postmortem Studies and Two Case Reports. *Journal of Analytical Toxicology* **1989**, 13 (2), 105–109.
 - (93) Furne, J.; Saeed, A.; Levitt, M. D. Whole Tissue Hydrogen Sulfide Concentrations Are Orders of Magnitude Lower than Presently Accepted Values. *AJP: Regulatory, Integrative and Comparative Physiology* **2008**, 295 (5), R1479–R1485.
 - (94) Ishigami, M.; Hiraki, K.; Umemura, K.; Ogasawara, Y.; Ishii, K.; Kimura, H. A Source of Hydrogen Sulfide and a Mechanism of Its Release in the Brain. *Antioxidants & Redox Signaling* **2009**, 11 (2), 205–214.
 - (95) Wintner, E. A.; Deckwerth, T. L.; Langston, W.; Bengtsson, A.; Leviten, D.; Hill, P.; Insko, M. A.; Dumpit, R.; VandenEckart, E.; Toombs, C. F.; Szabo, C. A Monobromobimane-Based Assay to Measure the Pharmacokinetic Profile of Reactive Sulphide Species in Blood: Monobromobimane Assay for Sulphide Detection. *British Journal of Pharmacology* **2010**, 160 (4), 941–957.
 - (96) Abe, K.; Kimura, H. The Possible Role of Hydrogen Sulfide as an Endogenous Neuromodulator. *Journal of Neuroscience* **1996**, 16 (3), 1066–1071.
 - (97) Hosoki, R.; Matsuki, N.; Kimura, H. The Possible Role of Hydrogen Sulfide as an Endogenous Smooth Muscle Relaxant in Synergy with Nitric Oxide. *Biochemical and Biophysical Research Communications* **1997**, 237 (3), 527–531.
 - (98) Moore, P. K.; Whiteman, M. *Chemistry, Biochemistry and Pharmacology of Hydrogen Sulfide*; 2015.
 - (99) Zhang, H.; Tang, J.; Liu, X.-P.; Wang, Y.; Yu, W.; Peng, W.-Y.; Fang, F.; Ma, D.-F.; Wei, Z.-J.; Hu, L.-Y. Hydrogen Sulfide Promotes Root Organogenesis in *Ipomoea Batatas*, *Salix Matsudana* and *Glycine Max*. *Journal of Integrative Plant Biology* **2009**, 51 (12), 1086–1094.
 - (100) Zhang, H.; Hu, L.-Y.; Hu, K.-D.; He, Y.-D.; Wang, S.-H.; Luo, J.-P. Hydrogen Sulfide Promotes Wheat Seed Germination and Alleviates Oxidative Damage against Copper Stress. *Journal of Integrative Plant Biology* **2008**, 50 (12), 1518–1529.
 - (101) Jin, Z.; Pei, Y. Physiological Implications of Hydrogen Sulfide in Plants: Pleasant Exploration behind Its Unpleasant Odour. *Oxidative Medicine and Cellular Longevity* **2015**, 2015, 1–6.
 - (102) Papenbrock, J.; Riemenschneider, A.; Kamp, A.; Schulz-Vogt, H. N.; Schmidt, A. Characterization of Cysteine-Degrading and H₂S-Releasing Enzymes of Higher Plants - From the Field to the Test Tube and Back. *Plant Biology* **2007**, 9 (5), 582–588.
 - (103) Ida, T.; Sawa, T.; Ihara, H.; Tsuchiya, Y.; Watanabe, Y.; Kumagai, Y.; Suematsu, M.; Motohashi, H.; Fujii, S.; Matsunaga, T.; Yamamoto, M.; Ono, K.; Devarie-Baez, N. O.; Xian, M.; Fukuto, J. M.; Akaike, T. Reactive Cysteine Persulfides and S-Polythiolation Regulate Oxidative Stress and Redox Signaling. *Proceedings of the National Academy of Sciences* **2014**, 111 (21), 7606–7611.
 - (104) Kimura, H. Physiological Role of Hydrogen Sulfide and Polysulfide in the Central Nervous System. *Neurochemistry International* **2013**, 63 (5), 492–497.
 - (105) Greiner, R.; Pálincás, Z.; Bäsell, K.; Becher, D.; Antelmann, H.; Nagy, P.; Dick, T. P. Polysulfides Link H₂S to Protein Thiol Oxidation. *Antioxidants & Redox Signaling* **2013**, 19 (15), 1749–1765.
 - (106) Filipovic, M. R. Persulfidation (S-Sulfhydration) and H₂S. In *Chemistry, Biochemistry and*

- Pharmacology of Hydrogen Sulfide*; Moore, P. K., Whiteman, M., Eds.; Springer International Publishing: Cham, 2015; Vol. 230, pp 29–59.
- (107) Ono, K.; Akaike, T.; Sawa, T.; Kumagai, Y.; Wink, D. A.; Tantillo, D. J.; Hobbs, A. J.; Nagy, P.; Xian, M.; Lin, J.; Fukuto, J. M. Redox Chemistry and Chemical Biology of H₂S, Hydropersulfides, and Derived Species: Implications of Their Possible Biological Activity and Utility. *Free Radical Biology and Medicine* **2014**, *77*, 82–94.
 - (108) Nishimori, I.; Vullo, D.; Minakuchi, T.; Scozzafava, A.; Osman, S. M.; AlOthman, Z.; Capasso, C.; Supuran, C. T. Anion Inhibition Studies of Two New β -Carbonic Anhydrases from the Bacterial Pathogen *Legionella Pneumophila*. *Bioorganic & Medicinal Chemistry Letters* **2014**, *24* (4), 1127–1132.
 - (109) Kabil, O.; Banerjee, R. Redox Biochemistry of Hydrogen Sulfide. *Journal of Biological Chemistry* **2010**, *285* (29), 21903–21907.
 - (110) Kajimura, M.; Fukuda, R.; Bateman, R. M.; Yamamoto, T.; Suematsu, M. Interactions of Multiple Gas-Transducing Systems: Hallmarks and Uncertainties of CO, NO, and H₂S Gas Biology. *Antioxidants & Redox Signaling* **2010**, *13* (2), 157–192.
 - (111) Li, Q.; Lancaster, J. R. Chemical Foundations of Hydrogen Sulfide Biology. *Nitric Oxide* **2013**, *35*, 21–34.
 - (112) Cooper, C. E.; Brown, G. C. The Inhibition of Mitochondrial Cytochrome Oxidase by the Gases Carbon Monoxide, Nitric Oxide, Hydrogen Cyanide and Hydrogen Sulfide: Chemical Mechanism and Physiological Significance. *Journal of Bioenergetics and Biomembranes* **2008**, *40* (5), 533–539.
 - (113) Hughes, M. N.; Centelles, M. N.; Moore, K. P. Making and Working with Hydrogen Sulfide. *Free Radical Biology and Medicine* **2009**, *47* (10), 1346–1353.
 - (114) Szabó, C. Hydrogen Sulphide and Its Therapeutic Potential. *Nat Rev Drug Discov* **2007**, *6* (11), 917–935.
 - (115) Wagner, F.; Asfar, P.; Calzia, E.; Radermacher, P.; Szabó, C. Bench-to-Bedside Review: Hydrogen Sulfide – the Third Gaseous Transmitter: Applications for Critical Care. *Critical Care* **2009**, *13* (3), 213.
 - (116) Vitvitsky, V.; Yadav, P. K.; Kurthen, A.; Banerjee, R. Sulfide Oxidation by a Noncanonical Pathway in Red Blood Cells Generates Thiosulfate and Polysulfides. *Journal of Biological Chemistry* **2015**, *290* (13), 8310–8320.
 - (117) Mishanina, T. V.; Libiad, M.; Banerjee, R. Biogenesis of Reactive Sulfur Species for Signaling by Hydrogen Sulfide Oxidation Pathways. *Nature Chemical Biology* **2015**, *11* (7), 457–464.
 - (118) Pálincás, Z.; Furtmüller, P. G.; Nagy, A.; Jakopitsch, C.; Pirker, K. F.; Magierowski, M.; Jasnos, K.; Wallace, J. L.; Obinger, C.; Nagy, P. Interactions of Hydrogen Sulfide with Myeloperoxidase: Sulfide Is a Substrate and Inhibitor of Myeloperoxidase. *British Journal of Pharmacology* **2015**, *172* (6), 1516–1532.
 - (119) Michelson, E. H.; DuBois, L. Intraspecific Variations in the Hemolymph of *Biomphalaria Glabrata*, a Snail Host of *Schistosoma Mansoni*. *Malacologia* **1975**, *15* (1), 105–111.
 - (120) Wang, D.; Liu, L.; Wang, H.; Xu, H.; Chen, L.; Ma, L.; Li, Z. Clues for Discovering a New Biological Function of *Vitreoscilla* Hemoglobin in Organisms: Potential Sulfide Receptor and Storage. *FEBS Letters* **2016**, *590* (8), 1132–1142.
 - (121) Watanabe, K.; Suzuki, T.; Kitagishi, H.; Kano, K. Reaction between a Haemoglobin Model Compound and Hydrosulphide in Aqueous Solution. *Chemical Communications* **2015**, *51* (19), 4059–4061.
 - (122) Pavlik, J. W.; Noll, B. C.; Oliver, A. G.; Schulz, C. E.; Scheidt, W. R. Hydrosulfide (HS⁻) Coordination in Iron Porphyrinates. *Inorganic Chemistry* **2010**, *49* (3), 1017–1026.
 - (123) Collman, J. P.; Ghosh, S.; Dey, A.; Decreau, R. A. Using a Functional Enzyme Model to Understand the Chemistry behind Hydrogen Sulfide Induced Hibernation. *Proceedings of the National Academy of Sciences* **2009**, *106* (52), 22090–22095.
 - (124) Nicoletti, F. P.; Comandini, A.; Bonamore, A.; Boechi, L.; Boubeta, F. M.; Feis, A.; Smulevich, G.; Boffi, A. Sulfide Binding Properties of Truncated Hemoglobins. *Biochemistry* **2010**, *49*

- (10), 2269–2278.
- (125) Kirkwood, J. G. Statistical Mechanics of Fluid Mixtures. *The Journal of Chemical Physics* **1935**, 3 (5), 300–313.
- (126) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. THE Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *Journal of Computational Chemistry* **1992**, 13 (8), 1011–1021.
- (127) Jarzynski, C. Nonequilibrium Equality for Free Energy Differences. *Physical Review Letters* **1997**, 78 (14), 2690–2693.
- (128) Cossins, B. P.; Foucher, S.; Edge, C. M.; Essex, J. W. Assessment of Nonequilibrium Free Energy Methods. *The Journal of Physical Chemistry B* **2009**, 113 (16), 5508–5519.
- (129) Cohen, J.; Olsen, K. W.; Schulten, K. Finding Gas Migration Pathways in Proteins Using Implicit Ligand Sampling. In *Methods in Enzymology*; Elsevier, 2008; Vol. 437, pp 439–457.
- (130) Cohen, J.; Arkhipov, A.; Braun, R.; Schulten, K. Imaging the Migration Pathways for O₂, CO, NO, and Xe Inside Myoglobin. *Biophysical Journal* **2006**, 91 (5), 1844–1857.

II. METODOLOGÍA GENERAL

II.1 Porqué usar simulación computacional en química

La simulación computacional es una herramienta que ha aparecido y se ha desarrollado con el advenimiento de las computadoras, y que permite, mediante modelos matemáticos plasmados en programas, realizar predicciones e interpretaciones acerca de las propiedades o el comportamiento de los sistemas.

En el caso de la química, el área más importante de simulación computacional se basa en predecir el comportamiento termodinámico y cinético a partir de las interacciones interatómicas. Además del poder predictivo, la simulación computacional en química permite estudiar el comportamiento de los sistemas a un nivel atómico, aportando información relevante que no siempre puede ser abordada desde un punto de vista experimental.

Para tal fin, se puede considerar que son dos los componentes cruciales de una simulación computacional: en primer lugar, en las formas más habituales para describir las interacciones entre los átomos presentes en los sistemas químicos (Sección A). En segundo lugar, el otro componente son las diversas estrategias que se utilizan para obtener información relevante. (Sección B)

II.2 Métodos de simulación computacional

Actualmente el modelo que ha resultado más exitoso en predecir las interacciones químicas es la Mecánica Cuántica. Los llamados “métodos cuánticos”, se utilizan para el estudio propiedades de los sistemas que se obtienen a partir de conocimiento del detalle de la estructura electrónica (por ejemplo, para el estudio reacciones químicas).

Hasta mediados del siglo XX, la aplicación de estos métodos se limitaba al estudio de las propiedades de átomos aislados o moléculas pequeñas, debido al alto costo computacional que representaba. Posteriormente, el desarrollo de la tecnología informática permitió a los cálculos de mecánica cuántica la resolución de una gran diversidad de problemas químicos en sistemas más complejos.

Sin embargo, aún en la actualidad, resulta inviable este tipo de cálculos en sistemas de miles de átomos (como por ejemplo una proteína), no solo por el tiempo de cómputo, sino también debido a que muchos resultados que se obtienen presentan errores asociados que dependen fuertemente con el tamaño del sistema. Por otro lado, tampoco tendría sentido invertir costosos tiempos de cómputos para obtener información detallada electrónica de las regiones no involucradas directamente en fenómenos reactivos.

Debido a lo anterior, en paralelo a los métodos cuánticos, se han desarrollado metodologías basadas en la mecánica clásica (métodos clásicos), y métodos multi-escala cuántico-clásicos (métodos híbridos).

Los métodos clásicos son típicamente usados en macromoléculas, por ejemplo proteínas. Estos métodos utilizan una descripción de la energía basada en modelos enmarcados en la física clásica.

Estos métodos que no permiten la descripción de procesos en los que haya formación o ruptura de enlaces químicos, ya que no describen explícitamente a los electrones del sistema. Tiene la ventaja de que los tiempos de cómputo asociados permiten la observación de eventos temporales del orden de los nano-microsegundos, por lo cual se pueden estudiar cambios conformacionales, migración de ligandos, etc.

Los métodos de multi-escala aprovechan las ventajas de uno y otro método, es decir permiten la formación y ruptura de enlaces químicos en la porción reactiva del sistema y además aprovechan la posibilidad de observar eventos temporales mayores a los que los métodos cuánticos puros permiten.

SECCIÓN A

DESCRIPCIÓN DE LAS INTERACCIONES INTERATÓMICAS

II.2.1 Métodos de estructura electrónica

II.2.1.1 Breve reseña histórica

La aurora del siglo XX alumbró una era, sin precedentes, de trastrocamientos y revaluaciones de las teorías clásicas que habían guiado a la física desde tiempos pre-newtonianos. El 14 de diciembre del año 1900 (curiosamente, el mismo día en que esta introducción fue terminada), Max Planck proclamó, ante el auditorio de la Sociedad Física de

Alemania, que era posible eliminar las paradojas que infestaban la teoría clásica de la emisión y de la absorción de luz por la materia, con solo asumir que esa energía radiante podía ser absorbida o emitida solo en forma de “paquetes” discretos. Plank bautizó a esos “paquetes” con el nombre de cuantos de luz.

Cinco años más tarde, Albert Einstein logró aplicar, con gran éxito, esta idea a las leyes empíricas del efecto fotoeléctrico. Varios años después, Arthur Compton efectuó su hoy clásica experiencia mostrando que la dispersión de rayos-X por electrones libres obedece la misma ley que la colisión entre dos esferas elásticas. De este modo, en unos pocos años, esa idea novedosa de la cuantificación de la energía radiante, se estableció con firmeza tanto en la física teórica como en la experimental.

En 1913, un investigador danés, Niels Bohr, extendió estas ideas para describir la energía mecánica de los electrones dentro de los átomos. Introdujo las “reglas de cuantización” específicas para sistemas atómicos y consiguió solucionar de manera coherente y lógica algunas incongruencias que el modelo atómico planetario de Ernest Rutherford tenía. Además, sus ideas permitieron, la interpretación de las líneas espectrales del átomo de hidrógeno y algunos más pesados.

El primer trabajo de Bohr, acerca de la teoría cuántica, desencadenó todo un cataclismo científico. Muchísimas propiedades ópticas, magnéticas y químicas de muchos elementos fueron interpretadas, pero a pesar del gran avance, no podía resolver algunas cuestiones atómicas bien conocidas, tales como la intensidad de los espectros de emisión atómicos.

En 1925 un físico francés, Louis de Broglie, planteó una interpretación totalmente inesperada: la naturaleza ondulatoria de las orbitas cuantificadas de Bohr (las llamó “ondas piloto”), cuya velocidad de propagación y longitud de onda dependían de la velocidad de traslación del electrón respectivo. De esta manera, pudo demostrar que las varias órbitas cuánticas permitidas por el modelo de Bohr eran aquellas en las cuales podía “acomodarse” un número entero de ondas piloto. De esta manera, el modelo atómico adquirió una visión parecida a un instrumento musical, en donde su tono básico era la órbita inferior (la de menor energía), y los varios armónicos eran las órbitas más y más exteriores (las de mayor energía).

Un año después, las ideas de Broglie hallaron en el físico austriaco Erwin Schrödinger mayor extensión y formalidad matemática, el nombre que recibieron estos desarrollos fue de mecánica cuántica ondulatoria.

Esta teoría fue la que dio razón a todos los fenómenos que fueron explicados por Bohr y además resolvió lo que no se había podido resolver. Además, su formulación, incluía fenómenos nuevos (como difracción de un haz de electrones, etc)^[1]

A partir de todos estos desarrollos, la teoría de la mecánica cuántica se posicionó como la teoría necesaria para la descripción de partículas de dimensiones atómicas.

II.2.1.2 Principios básicos

Uno de los postulados de la mecánica cuántica establece la existencia de una función $|\Psi\rangle$, llamada función de onda, que describe el estado de un sistema mecano-cuántico y contiene toda la información física que puede conocerse del sistema.

En particular, esta función se puede proyectar en el espacio de coordenadas de todas las partículas intervinientes (r) y el tiempo (t), adquiriendo la forma matemática más habitual de $\Psi(r, t)$. Esta dependencia espacio-temporal permite obtener numerosas propiedades espaciales y temporales del sistema (como densidad de carga, momentos dipolar y magnético, probabilidades espaciales de encontrar alguna partícula, etc)

La función $\Psi(r, t)$ se obtiene de la resolución de la ecuación fundamental de la Mecánica Cuántica, la ecuación de Schrödinger dependiente del tiempo:

$$-\frac{\hbar}{i} \frac{\partial \Psi(r, t)}{\partial t} = -\frac{\hbar^2}{2m} \nabla^2 \Psi(r, t) + V(r, t) \Psi(r, t) \quad (II.1)$$

donde m es la masa de la partícula, y $V(r, t)$ es la función de energía potencial del sistema.

Ahora bien, suponiendo que el potencial $V(r, t) = V(r)$, es decir, para sistemas bajo fuerzas externas constantes en el tiempo, y que la función de onda $\Psi(r, t)$ se puede desacoplar en dos funciones, una que solo involucre coordenadas espaciales y otra que involucre coordenadas temporales (es decir $\Psi(r, t) = \psi(r) \cdot f(t)$), la ecuación (II.1) se transforma en:

$$-\frac{\hbar^2}{2m} \nabla^2 \psi(r) + V(r) \psi(r) = E \cdot \psi(r) \quad (II.2)$$

Se define entonces el operador Hamiltoniano $\hat{H}$ del sistema como:

$$\hat{H} = -\frac{\hbar^2}{2m} \nabla^2 + V(r) \quad (II.3)$$

el primer término en H es el operador energía cinética y el segundo es la función potencial, resultando entonces en

$$\hat{H}|\psi(r)\rangle = E|\psi(r)\rangle \quad (II.4)$$

Donde E es la energía del sistema.

Resolver las ecuaciones (II.3) y (II.4) implica la resolución de un problema de autovalores y autofunciones asociados al operador $\hat{H}$. La resolución dará entonces una autofunción $\psi(r)$ con autovalor asociado E.

La mecánica cuántica para sistemas moleculares, debe entonces resolver esta ecuación para un sistema de núcleos y electrones, para de esta manera conocer la función de onda $\psi(r)$ y por lo tanto todas las propiedades espaciales del sistema.

Suponiendo que los núcleos y los electrones son masas puntuales, y despreciando efectos relativistas, el Hamiltoniano para una molécula de M núcleos y N electrones resulta, en unidades atómicas:

$$H = -\frac{1}{2} \sum_{i=1}^N \nabla_i^2 - \frac{1}{2} \sum_{A=1}^M \frac{1}{m_A} \nabla_A^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} + \sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{R_{AB}} \quad (II.5)$$

donde i, j denotan electrones, A y B denotan núcleos, los valores de Z_A y m_A corresponden a los números atómicos y a las masas de los núcleos, respectivamente; r_{iA} , r_{ij} y R_{AB} representan las distancias electrón-núcleo, electrón-electrón y núcleo-núcleo.

En unidades atómicas, la unidad de masa es la masa del electrón, m_e , la unidad de carga es la carga del protón, e' , y la unidad del momento angular es $\hbar$, simplificándose la expresión del Hamiltoniano. Se definen las unidades de energía, hartree: $1\text{hartree} = E_h \equiv e'^2 / a_0 = e^2 / 4\pi\epsilon_0 a_0$ y de longitud, Bohr: $1\text{bohr} \equiv \hbar^2 / m_e e'^2$

Observando la ecuación (II.5) puede verse que el primer y segundo término corresponden a los operadores energía cinética de los electrones y los núcleos, mientras que los tres últimos constituyen el operador energía potencial. El mismo es la suma de la interacción Coulómbica atractiva entre los núcleos y los electrones (tercer término), la interacción repulsiva entre los electrones (cuarto término) y la repulsión entre las cargas de los núcleos (quinto término).

Para los sistemas de interés químico, resulta de gran utilidad la aproximación de Born-Oppenheimer.^[2,3] La misma supone que los núcleos son mucho más pesados que los electrones, por lo tanto, éstos últimos se moverán mucho más rápidamente que los primeros. Esto permite

desacoplar el movimiento de núcleos y electrones, de manera tal de resolver una ecuación de Schrödinger para los electrones del sistema considerando fijas las posiciones de los núcleos.

En términos matemáticos, desacoplar los movimientos de núcleos y electrones, permite escribir la función de onda del sistema como producto de funciones de onda, una asociada a los electrones (ψ_{el}) que dependa solo de coordenadas asociadas a los electrones (y spin), y otra función asociada a los núcleos (ψ_N), que dependa solo de coordenadas asociadas a los núcleos.

Bajo esta suposición, en la ecuación (II.5) desaparece el término energía cinética de los núcleos, y el término de repulsión núcleo-núcleo es una constante.

Puede establecerse entonces un nuevo Hamiltoniano, llamado Hamiltoniano electrónico ($\hat{H}_{el}$) asociado a la función de onda electrónica:

$$\hat{H}_{el}|\psi_{el}\rangle = E_{el}|\psi_{el}\rangle \quad (II.6)$$

$$\hat{H}_{el} = -\frac{1}{2}\sum_{i=1}^N \nabla_i^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} \quad (II.7)$$

Por cada una de las posiciones de los núcleos podemos resolver la ecuación (II.6), por lo tanto la función de onda electrónica y la energía dependerán paramétricamente de las posiciones de los núcleos.

Lo anterior constituye posiblemente una de las ideas más poderosas de la fisicoquímica y de la simulación computacional, dado que permite definir lo que conocemos como superficie o hipersuperficie de energía potencial (SEP). Esta hipersuperficie de energía potencial contiene en principio, toda la información química relevante (como se ha presentado en la introducción de esta tesis).

Finalmente, luego de resolver una ecuación equivalente para los núcleos, la energía total resultará la suma de la contribución electrónica y la nuclear.

La primera aproximación realizada para resolver un sistema de N electrones (descrito por las ecuaciones (II.6) y (II.7)), consistió en despreciar la interacción electrónica (término dependiente de r_{ij}), esto permite considerar a $\hat{H}_{el}$ como suma de Hamiltonianos monoelectrónicos $\hat{h}_i$ sin términos que “cruzan” las coordenadas de un electrón i con los de un electrón j.

$$\hat{H}_{el} = \sum_{i=1}^N \hat{h}_i \quad (II.8)$$

En estos casos, $\hat{h}_i$ tendrá su propio conjunto de autovalores ϵ_i y autovectores χ_i , llamados spin-orbitales, siendo la función de onda total el producto de las funciones de onda monoeléctricas, a este producto se lo denominó producto de Hartree:

$$\psi^{HP}(x_1, x_2, \dots, x_N) = \chi_i(x_1)\chi_j(x_2) \dots \chi_k(x_N) \quad (II.9)$$

Donde $(x_1, x_2, \dots, x_N)$ representan las coordenadas espaciales y de spin de los electrones $(1, 2, \dots, N)$.

Sin embargo, debe tenerse en cuenta que los electrones son fermiones, es decir partículas con spin semientero, y por lo tanto la función de onda debe ser antisimétrica respecto al intercambio de coordenadas de dos partículas cualesquiera. Es decir, debe cumplir que

$$\psi(x_1, x_2, \dots, x_i, \dots, x_j, \dots, x_N) = -\psi(x_1, x_2, \dots, x_j, \dots, x_i, \dots, x_N)$$

donde $x_{i,j}$ representa las coordenadas de cada electrón, incluyendo las coordenadas de spin

El producto de Hartree no satisface la condición de antisimetría de la función de onda electrónica, y además hace a los electrones distinguibles entre ellos, ya que identifica cada electrón con un spin-orbital determinado. En 1929, Slater demostró que esta condición puede cumplirse utilizando como función de onda un determinante, construido de la siguiente forma:

$$\psi(x_1, x_2, \dots, x_N) = \left(\frac{1}{N!}\right)^{\frac{1}{2}} \begin{bmatrix} \chi_i(x_1) & \chi_j(x_1) & \cdots & \chi_k(x_1) \\ \chi_i(x_2) & \chi_j(x_2) & \cdots & \chi_k(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_i(x_N) & \chi_j(x_N) & \cdots & \chi_k(x_N) \end{bmatrix} \quad (II.10)$$

Este determinante contiene N electrones ocupando N spin-orbitales χ_i , constituyendo una función de onda antisimétrica ante cualquier intercambio de pares de coordenadas y hace indistinguibles a los electrones.

Si bien el producto de Hartree considera la interacción electrónica de manera promedio, como se mencionó no cumple con la condición de considerar a los electrones como fermiones indistinguibles. El determinante de Slater, si bien sigue utilizando las funciones monoeléctricas, incorpora los llamados efectos (o correlación) de intercambio, es decir, la correlación entre el movimiento de dos electrones con el mismo estado de spin, de manera que si dos electrones se encuentran en el mismo orbital, la función de onda se anula (ya que el

determinante de Slater se anula), por lo que tal evento tiene probabilidad nula de ocurrir. Sin embargo, la correlación entre electrones de spin opuesto permanece sin considerarse.

II.2.1.3 El Teorema Variacional y el método variacional lineal

En este punto resulta necesario encontrar un método que permita hallar la función de onda (o las funciones de onda monoeléctricas).

En este contexto, gran cantidad de métodos que resuelven la ecuación de Schrödinger se apoyan en el Teorema Variacional.

Como se mencionó, la ecuación que se desea resolver es una ecuación de autovectores/autovalores de la forma:

$$\hat{H}|\Phi\rangle = E|\Phi\rangle \quad (II.11)$$

donde $\hat{H}$, el operador Hamiltoniano.

Otro postulado de la mecánica cuántica establece que para cualquier sistema físico, $\hat{H}$ es un operador hermítico, por lo tanto posee un conjunto completo de funciones de base $\{|\Phi_\alpha\rangle\}$ (ortonormales), que lo diagonalizan y además sus autovalores asociados E_α son números reales. Es decir:

$$\begin{aligned} \hat{H}|\Phi_\alpha\rangle &= E_\alpha|\Phi_\alpha\rangle \\ \alpha &= 0, 1, \dots \text{ y } E_\alpha \in \mathbb{R} \\ \langle\Phi_\alpha|\Phi_\beta\rangle &= \delta_{\alpha\beta} \end{aligned} \quad (II.12)$$

Por lo tanto, como $\{|\Phi_\alpha\rangle\}$ es un conjunto de funciones de base del espacio de soluciones de las ecuaciones (II.12), cualquier función $|\tilde{\Phi}\rangle$ normalizada, que llamaremos función de prueba, podrá escribirse como combinación lineal del conjunto de funciones $\{|\Phi_\alpha\rangle\}$, es decir

$$|\tilde{\Phi}\rangle = \sum_{\alpha} c_{\alpha}|\Phi_{\alpha}\rangle \quad (II.13)$$

El teorema variacional establece que dada una función de prueba normalizada $|\tilde{\Phi}\rangle$, el valor medio del Hamiltoniano para esa función debe ser mayor o igual que el autovalor más bajo (es decir la energía del estado fundamental del sistema, E_0) de $\hat{H}$ en la base $\{|\Phi_\alpha\rangle\}$. Esto es:

$$\langle\tilde{\Phi}|\hat{H}|\tilde{\Phi}\rangle \geq E_0 \quad (II.14)$$

En términos prácticos, el teorema variacional nos permite saber que la energía obtenida para cualquier función de prueba, será siempre mayor o igual a la energía real del estado fundamental E_0 . En otras palabras, cuanto menor sea la energía asociada a una función de prueba usada, mejor será esa función como aproximación de la solución real para el estado fundamental. Por lo tanto, la resolución del problema consiste en hallar una función de prueba $|\tilde{\Phi}\rangle$ que minimice la energía del sistema.

Tomemos entonces una función de prueba expresada como una expansión lineal (con coeficientes reales) de N funciones conocidas $|\psi_i\rangle$ (llamadas habitualmente como “funciones de base”).

$$|\tilde{\Phi}\rangle = \sum_i c_i |\psi_i\rangle \quad (II.15)$$

Es importante destacar que las funciones $|\psi_i\rangle$ no son necesariamente el conjunto de autofunciones de $\hat{H}$ (en cuyo caso, la resolución sería trivial).

Observando la definición (II.15), el problema se traslada entonces a encontrar los coeficientes c_i que minimizan la energía. Por lo tanto, evaluando el valor medio de $\hat{H}$ se obtiene que

$$\langle \tilde{\Phi} | \hat{H} | \tilde{\Phi} \rangle = \sum_{i,j} c_i c_j \langle \psi_i | \hat{H} | \psi_j \rangle \quad (II.16)$$

La representación del operador Hamiltoniano en la base $\{|\psi_i\rangle\}$ corresponde a una matriz $\mathbf{H}$: $N \times N$ cuyos elementos cumplen las siguientes relaciones

$$H_{ij} = \langle \psi_i | H | \psi_j \rangle \quad (II.17)$$

$$\langle \tilde{\Phi} | H | \tilde{\Phi} \rangle = \sum_{i,j} c_i c_j \langle \psi_i | H | \psi_j \rangle = \sum_{i,j} c_i c_j H_{ij} \quad (II.18)$$

Dado que la función de prueba $|\tilde{\Phi}\rangle$ está normalizada y las funciones de base $\{|\psi_i\rangle\}$ son un conjunto ortonormal, los coeficientes están sujetos a la siguiente relación:

$$\langle \tilde{\Phi} | \tilde{\Phi} \rangle = \sum_{i,j} c_i c_j \langle \psi_i | \psi_j \rangle = \sum_i c_i^2 = 1 \quad (II.19)$$

Entonces el problema consiste en encontrar un conjunto de parámetros $\{c_i\}$ que minimicen la expresión (II.14), sujetos a la restricción (II.20).

Este problema puede resolverse utilizando el método de los multiplicadores de Lagrange, planteando el siguiente Lagrangiano.

$$\lambda(c_1, c_2, \dots, c_N, E) = \langle \tilde{\Phi} | H | \tilde{\Phi} \rangle - E \cdot (\langle \tilde{\Phi} | \tilde{\Phi} \rangle - 1) \quad (\text{II.20})$$

y según las ecuaciones (II.18), (II.19) y (II.20) el Lagrangiano a minimizar es

$$\lambda(c_1, c_2, \dots, c_N, E) = \sum_{i,j} c_i c_j H_{ij} - E \left(\sum_i c_i^2 - 1 \right) \quad (\text{II.21})$$

Derivando la expresión anterior respecto de cada uno de los coeficientes c_i e igualando a cero para buscar el mínimo, se obtiene el siguiente conjunto de ecuaciones

$$\frac{\partial \lambda}{\partial c_k} = 0 = \sum_j c_j H_{kj} + \sum_i c_i H_{ik} - 2E c_k \quad k = 1, \dots, N \quad (\text{II.22})$$

Dado que la matriz $\mathbf{H}$ es simétrica, es decir $H_{ij} = H_{ji}$, resulta:

$$\sum_i H_{ij} - E c_i = 0 \quad i = 1, \dots, N \quad (\text{II.23})$$

La expresión (II.23) puede escribirse también en forma matricial, definiendo un vector $\mathbf{c}$ de elementos c_i , resultando:

$$\mathbf{H} \cdot \mathbf{c} = E \cdot \mathbf{c} \quad (\text{II.24})$$

Resolviendo la ecuación (II.24), se obtienen entonces N vectores ortonormales c^α y sus correspondientes autovalores E_α . Una vez encontrados los coeficientes c^α podemos escribir las autofunciones $|\tilde{\Phi}_\alpha\rangle$ como:

$$|\tilde{\Phi}_\alpha\rangle = \sum_i c_i^\alpha |\psi_i\rangle \quad \alpha = 0, 1, \dots, N-1 \quad (\text{II.25})$$

Puede verse finalmente que los valores de E_α corresponden a los valores medios de Hamiltoniano para cada una de las funciones $|\tilde{\Phi}_\alpha\rangle$:

$$\begin{aligned} \langle \tilde{\Phi}_\beta | H | \tilde{\Phi}_\alpha \rangle &= \sum_{i,j} c_i^\beta c_j^\alpha \langle \tilde{\Phi}_\beta | H | \tilde{\Phi}_\alpha \rangle = \sum_{i,j} c_i^\beta H_{ij} c_j^\alpha = (c^\beta)^T H c^\alpha \\ &= E_\alpha (c^\beta)^T c^\alpha = E_\alpha \delta_{\alpha\beta} \end{aligned} \quad (\text{II.26})$$

En particular, el menor autovalor, que denominaremos E_α , corresponde a la mejor aproximación al estado fundamental del sistema descrito por el Hamiltoniano $\hat{H}$ dentro del espacio de funciones generado por el conjunto $\{|\psi_i\rangle\}$.

Puede demostrarse también que los siguientes autovalores corresponden a cotas superiores a las energías de los distintos estados excitados.

En resumen, el método variacional lineal, basado en el teorema variacional, permite encontrar las mejores soluciones a la ecuación de Schrödinger representando al operador $\hat{H}$ en una base finita conocida $\{|\psi_i\rangle\}$ y diagonalizando la matriz resultante $\mathbf{H}$.

II.2.1.4 El método de Hartree-Fock (HF)

El método HF fue uno de los primeros métodos de cálculo de estructura electrónica. Si bien no ha sido utilizado en la presente tesis, se describirá brevemente ya que sobre este método están desarrollados una gran cantidad de metodologías desarrolladas posteriormente.

Por otro lado, este método tiene una gran relevancia porque solo en el contexto de HF cobra sentido la utilización de energías orbitales, concepto al cual se recurre cotidianamente en la química.

En HF la función de onda se describe como un determinante de Slater (ecuación (II.10)), escrito también en términos de las funciones χ_i monoeléctricas, que llamaremos spin-orbitales.

Aplicando el teorema variacional, la función de onda del sistema se hallará encontrando al conjunto de spin-orbitales χ_i que minimicen la energía.

Retornando a la expresión del Hamiltoniano electrónico:

$$\hat{H}_{el} = -\frac{1}{2} \sum_{i=1}^N \nabla_i^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} \quad (II.27)$$

y definiendo a los operadores de Coulomb ($\hat{J}$), de Intercambio ($\hat{K}$), y el operador $\hat{h}$ según:

$$\hat{J}_b \chi_a(1) = \left\{ \int dx_2 \chi_b^*(2) r_{12}^{-1} \chi_b(2) \right\} \chi_a(1) \quad (II.28)$$

$$\hat{K}_b \chi_a(1) = \left\{ \int dx_2 \chi_b^*(2) r_{12}^{-1} \chi_a(2) \right\} \chi_b(1) \quad (II.29)$$

$$\hat{h} = -\frac{1}{2} \nabla_1^2 - \sum_{\alpha} \frac{Z_{\alpha}}{r_{1\alpha}} \quad (II.30)$$

La expresión (II.28) corresponde al término de interacción Coulómbica entre los electrones 1 y 2, el cual se calcula promediando sobre todas las coordenadas espaciales y de spin del electrón 2.

La expresión (II.29), surge debido a la antisimetrización de la función de onda, no pudiéndose asociarle un significado físico simple.

Por otra parte, $\hat{h}$ corresponde al operador de un electrón suma de los operadores energía cinética e interacción con los núcleos.

La expresión de la energía que se desea minimizar resulta entonces:

$$E = \langle \psi | \hat{H} | \psi \rangle = \sum_{i=1}^N h_{ii} + \sum_{i=1}^N \sum_{j=1}^N (J_{ij} - K_{ij}) \quad (II.31)$$

Minimizando esta expresión (derivando respecto a los spin-orbitales) sujeta a la condición de ortonormalidad de los spin-orbitales χ_i , se obtienen las llamadas ecuaciones de Hartree-Fock:

$$\hat{f}|\chi_i(1)\rangle = \sum_j \varepsilon_{ij}|\chi_j(1)\rangle \quad (II.32)$$

y los valores de ε_{ij} corresponden a los multiplicadores de Lagrange asociados a los N spin-orbitales $\chi_j(1)$ y $\hat{f}$, es el operador de Fock, definido como:

$$\hat{f} = \hat{h} + \sum_b (\hat{J}_b - \hat{K}_b) \quad (II.33)$$

Haciendo un cambio de base de $\{|\chi_a\rangle\}$ a $\{|\chi'_a\rangle\}$ puede diagonalizarse la matriz de multiplicadores de Lagrange, resultando:

$$\hat{f}|\chi'_i(1)\rangle = \varepsilon'_i|\chi'_i(1)\rangle \quad (II.34)$$

donde los $\{|\chi'_i(1)\rangle\}$ corresponden al conjunto de orbitales moleculares canónicos y los $\{\varepsilon'_i\}$ a sus energías asociadas. De aquí en adelante, notaremos χ_i a los orbitales canónicos y ε_i a sus energías orbitales asociadas.

La resolución de estas ecuaciones, es decir la obtención del conjunto de orbitales canónicos y sus correspondientes autovalores, tiene como problema adicional que el operador $\hat{f}$ depende sus propias autofunciones. Por lo tanto debe resolverse el problema de forma iterativa.

Se realiza en principio la aproximación de expandir al conjunto de orbitales canónicos en una base finita, $\{|\Phi_\mu\rangle\}$ en la que cada orbital tendrá un conjunto de coeficientes $\{C_{\mu i}\}$. Resulta entonces que:

$$\sum_{\mu=1}^M C_{\mu i} \hat{f} |\Phi_{\mu}\rangle = \varepsilon_i \sum_{\mu=1}^M C_{\mu i} \hat{f} |\Phi_{\mu}\rangle \quad (II.35)$$

donde M es el número de funciones en la base elegida.

Aplicando $\langle \Phi_v |$ a ambos lados de la ecuación anterior, se obtiene un sistema de ecuaciones homogéneas:

$$\sum_{\mu=1}^M C_{\mu i} \langle \Phi_v | \hat{f} | \Phi_{\mu} \rangle - \varepsilon_i \langle \Phi_v | \Phi_{\mu} \rangle = 0 \quad \mu = 1, \dots, M \quad (II.36)$$

Esta ecuación puede escribirse de forma matricial, definiendo $f_{\mu\nu} = \langle \Phi_v | \hat{f} | \Phi_{\mu} \rangle$ como los elementos de matriz **F** y $S_{\mu\nu} = \langle \Phi_v | \Phi_{\mu} \rangle$ como los elementos de matriz de **S**:

$$\mathbf{FC} = \mathbf{SC}\boldsymbol{\varepsilon} \quad (II.37)$$

siendo **C** la matriz de coeficientes $C_{\mu i}$ y **ε** la matriz diagonal de los coeficientes ε_i .

En este punto, el problema de determinar los orbitales moleculares $\{\psi_i\}$ y las energías orbitales ε_i consiste en resolver la ecuación matricial (II.37). Sin embargo, puede verse que la matriz de Fock **F** depende de los coeficientes, con lo que la solución de esta ecuación debe realizarse iterativamente.

En un sistema simple de capa cerrada donde cada orbital molecular está ocupado por dos electrones, la densidad electrónica queda determinada por:

$$\rho(r) = 2 \sum_a^{N/2} |\psi_a(r)|^2 \quad (II.38)$$

de manera que $\rho(r)dr$ representa la probabilidad de encontrar un electrón dentro de un radio dr de r . La integral de la densidad electrónica en todo el espacio resulta el número total de electrones.

Insertando en esta ecuación la expansión de la función de onda en orbitales resulta:

$$\rho(r) = 2 \sum_a^{N/2} \psi_a^*(r) \psi_a(r) = 2 \sum_a^{N/2} \sum_v C_{va}^* \Phi_v^*(r) \sum_{\mu} C_{\mu a} \Phi_{\mu}(r) \quad (II.39)$$

$$\rho(r) = \sum_{\mu\nu} \left[2 \sum_a^{\frac{N}{2}} c_{\mu a} c_{\nu a}^* \right] \Phi_{\mu}(r) \Phi_{\nu}^*(r) \quad (II.40)$$

$$\rho(r) = \sum_{\mu\nu} P_{\mu\nu} \Phi_{\mu}(r) \Phi_{\nu}^*(r) \quad (II.41)$$

En la ecuación (II.40), se introduce la denominada matriz de densidad ($P_{\mu\nu}$).

$$P_{\mu\nu} = 2 \sum_a^{\frac{N}{2}} c_{\mu a} c_{\nu a}^* \quad (II.42)$$

Para la gran mayoría de las aplicaciones en química, resulta de utilidad obtener una medida de la carga sobre cada átomo del sistema. Si bien no existe una única manera de definir la cantidad de electrones “asociados” a cada átomo, una manera de tener una representación de la distribución de carga sobre el sistema consiste en lo que se denomina análisis poblacional de Mülliken.

Considerando la definición de la matriz de densidad $\mathbf{P}$ y de la matriz de solapamiento $\mathbf{S}$, puede verse que el número total de electrones N corresponde a:

$$N = \sum_{\mu} \sum_{\nu} P_{\mu\nu} S_{\mu\nu} = \sum_{\mu} (\mathbf{PS})_{\mu\mu} = \text{tr}(\mathbf{PS}) \quad (II.43)$$

Puede interpretarse entonces $(\mathbf{PS})_{\mu\mu}$ como la cantidad de electrones asociada al orbital Φ_{μ} . Si los orbitales utilizados están centrados en los átomos, la cantidad de electrones asociados a cada átomo corresponderá a la suma de los electrones asociados a los orbitales de cada átomo. La carga neta sobre cada átomo se calcula a partir de la ecuación (II.43). Este cálculo resta a la carga nuclear del átomo A , el número de electrones asociados (es decir, $q_A = Z_A - \sum_{\mu \in A} (\mathbf{PS})_{\mu\mu}$). En donde Z_A corresponde a la carga nuclear del átomo A y el índice indica la suma sobre todos los orbitales centrados en el átomo A .

Por último, la calidad de los resultados del método HF dependerá fuertemente de la calidad de la base usada para expandir los orbitales. Sin embargo, por más que se expanda infinitamente la base, en el método de HF hay aparejado un error intrínseco del método. El mismo se origina en el tratamiento de la repulsión electrón-electrón, el cual se realiza en forma promediada, a partir de un potencial de campo medio, sin tener en cuenta las repulsiones instantáneas. Este error asociado recibe el nombre de energía de correlación y se define como la diferencia entre la energía no relativista exacta, dentro de la aproximación de Bohr-

Oppenheimer (E_0) y la energía límite de HF, correspondiente a la energía calculada con una base completa, es decir $E_{corr} = E_0 - E_{HF}^{lim}$

Existe una gran cantidad de métodos que permiten obtener resultados que estiman la energía de correlación. Entre ellos pueden mencionarse el método de interacción de configuraciones, los métodos basados en la teoría de perturbaciones de Moller-Plesset y la Teoría del funcional de la densidad (DFT).

En la próxima sección, se desarrollarán los principios básicos de DFT, dado que es la teoría de elección en los cálculos cuánticos realizados en esta tesis.

II.2.1.5 Teoría del Funcional de la Densidad (DFT)

La teoría del funcional de la densidad (DFT), en vez de intentar encontrar la función de onda que minimiza la energía del sistema, basa su formulación en la búsqueda de la densidad electrónica.

Esta teoría resulta más atractiva en este sentido, dado que la función de onda carece de un significado físico directo, mientras que la densidad electrónica sí lo tiene, ya que está relacionada con la probabilidad de encontrar un electrón. Otra gran ventaja es que la densidad electrónica depende solamente de las coordenadas espaciales (x, y, z) y no de coordenadas de spin de los electrones.

Las primeras ideas acerca de esta teoría fueron tomadas de los trabajos de Thomas y Fermi alrededor de 1920, en donde usaban consideraciones estadísticas para aproximar la distribución de electrones en un átomo^[4,5]. Estos trabajos mostraron que en un átomo, la energía depende explícitamente de la densidad electrónica. Como estos modelos no mostraron una mejora significativa respecto a los modelos que se conocían previamente, se los consideró como modelos simplificados de casos específicos ya que estos modelos no eran aplicables a moléculas

En 1964, Pierre Hohenberg y Walter Kohn tomaron y extendieron estas ideas primigenias demostrando que para moléculas con un estado fundamental no degenerado, la energía molecular del estado fundamental, la función de onda y todas las demás propiedad electrónicas, están determinadas unívocamente por la densidad de probabilidad electrónica del estado fundamental $\rho_0(x, y, z)$.^[6]

Puede verse también, que el potencial externo $v(r_i)$, es decir el potencial actuante sobre el electrón i producido por las cargas de los núcleos, quedará también determinado por la densidad electrónica ρ_0 .

Se dice entonces que la energía del estado fundamental E_0 es un funcional de ρ_0 y se escribe $E_0[\rho_0]$. El funcional energía puede escribirse entonces de la siguiente manera:

$$E[\rho] = T[\rho] + V_{Ne}[\rho] + V_{ee}[\rho] \quad (II.44)$$

donde $T[\rho]$ representa el funcional de energía cinética de los electrones, $V_{Ne}[\rho]$ el funcional de interacción entre los núcleos y los electrones, y $V_{ee}[\rho]$ el funcional de interacción entre electrones.

Por otro lado, los mismos autores demostraron el teorema que se conoce como Teorema Variacional de Hohenberg-Kohn. El mismo es análogo al Teorema Variacional mencionado anteriormente, y establece que para toda función densidad de prueba $\rho_{pr}(r)$ que satisface

$$\int \rho_{pr}(r)dr = n \quad \rho_{pr} \geq 0 \quad \forall r \quad E(\rho_{pr}) \geq E_0 \quad (II.45)$$

donde n es el número total de electrones.

Es decir, ρ_0 minimiza el funcional de la energía, al igual que la función de onda del estado fundamental minimiza la energía del sistema.

La teoría del funcional de la densidad carecía de utilidad práctica hasta que Kohn y Sham desarrollaron un método que permite calcular eficientemente la energía del estado fundamental y la densidad electrónica.^[7]

El desarrollo de Kohn y Sham se basó en la idea de considerar un sistema ficticio de n electrones no interactuantes, todos ellos experimentando un potencial externo $v_s(r_i)$ tal que la densidad electrónica del sistema no interactuante $\rho_{pr}(r)$ es igual a la densidad del estado fundamental del sistema real $\rho_0(r)$.

Entonces, el Hamiltoniano del sistema de electrones no interactuantes puede escribirse como:

$$\hat{H}_s = \sum_{i=1}^n \left[-\frac{1}{2} \nabla_i^2 + v_s(r_i) \right] = \sum_{i=1}^N \hat{h}_i^{KS} \quad (II.46)$$

donde $\hat{h}_i^{KS}$ es el Hamiltoniano de un electrón de Kohn-Sham. La función de onda del Hamiltoniano de electrones no interactuantes puede calcularse como un determinante de Slater de las autofunciones de $\hat{h}_i^{KS}$, es decir los spin-orbitales χ_i^{KS} tal que:

$$\hat{h}_i^{KS} \chi_i^{KS} = \varepsilon_i \chi_i^{KS} \quad (II.47)$$

Kohn y Sham definieron también:

$$\Delta T[\rho] = T[\rho] - T_s[\rho] \quad (II.48)$$

$$V_{ee}[\rho] = V_{ee}[\rho] - \frac{1}{2} \int \int \frac{\rho(r_1)\rho(r_2)}{r_{12}} dr_1 dr_2 \quad (II.49)$$

ΔT y ΔV_{ee} corresponden a la diferencia en la energía cinética y en la parte no clásica (expresada en el segundo término en la expresión de $\Delta V_{ee}[\rho]$) entre el sistema real y el sistema no interactuante. La suma de ambos términos define un nuevo funcional, llamado funcional de intercambio y correlación ($E_{xc}[\rho]$) es decir:

$$E_{xc}[\rho] = \Delta T[\rho] + \Delta V_{ee}[\rho] \quad (II.50)$$

$E_{xc}[\rho]$ contiene entonces las siguientes contribuciones: la energía cinética de correlación ($\Delta T[\rho]$), la energía de intercambio (asociada al requerimiento de antisimetría), la energía de correlación Coulómbica, asociada a las repulsiones interelectrónicas, y por último una corrección de autointeracción (self-interaction correction, SIC). Esta última contribución se considera en el segundo término de la ecuación (II.49), la cual permite la interacción de un electrón con su propia densidad de carga.

Si bien los términos de autointeracción se cancelaban en el método de HF, (a partir de las definiciones de los operadores de Coulomb (II.28) e Intercambio (II.29) se observa que $J_{ii} = K_{ii}$ con lo cual los términos $i=j$ se anulan en (II.31)), en la formulación original de Kohn y Sham no se consideró explícitamente este efecto, con lo cual es agregado de manera ad-hoc. En algunos funcionales de intercambio y correlación el término de SIC también es agregado explícitamente, mientras que en otros no (cuando el funcional de intercambio alcanza a minimizar el error debido a la autointeracción)

Continuando, la ecuación (II.44) puede reescribirse de la siguiente manera:

$$E[\rho] = T_s[\rho] + V_{ne}[\rho] + \frac{1}{2} \int \int \frac{\rho(r_1)\rho(r_2)}{r_{12}} dr_1 dr_2 + E_{xc}[\rho] \quad (II.51)$$

donde:

$$T_s[\rho] = \frac{1}{2} \langle \chi_i^{KS} | \sum_i \nabla_i^2 | \chi_i^{KS} \rangle \quad (II.52)$$

Y

$$V_{Ne}[\rho] = \int v(r)\rho(r)dr \quad (II.53)$$

Para encontrar la densidad electrónica que minimiza la energía debe calcularse la siguiente derivada funcional e igualar a cero:

$$\frac{\partial E[\rho(r)]}{\partial \rho(r)} = 0 \quad (II.54)$$

El procedimiento es análogo al procedimiento para obtener las ecuaciones de HF, en este caso las ecuaciones resultantes se conocen como las ecuaciones de Kohn y Sham (KS):

$$\left[-\frac{1}{2}\nabla_i^2 - \sum_{\alpha} \frac{Z_{\alpha}}{r_{1\alpha}} + \int \frac{\rho(r_2)}{r_{12}} dr_2 + v_{xc}(1) \right] \chi_i^{KS}(1) = \varepsilon_i^{KS} \chi_i^{KS}(1) \quad (II.55)$$

En la expresión (II.56), la expresión entre corchetes constituye el operador $\hat{h}_i^{KS}$ y el término $v_{xc}(1)$ corresponde al potencial de intercambio y correlación, es decir la derivada funcional de E_{xc} .

Los orbitales de Kohn y Sham se expresan como una combinación lineal en un conjunto de funciones de base $\{|\Phi_{\mu}\rangle\}$, lo cual, en forma similar a lo explicado para HF, permite escribir el problema en forma matricial de la siguiente manera:

$$\mathbf{h}^{KS} \mathbf{C} = \mathbf{S} \mathbf{C} \boldsymbol{\varepsilon} \quad (II.56)$$

donde $\mathbf{h}^{KS}$ constituye la matriz formada por los elementos de matriz $h_{\mu\nu}^{KS} = \langle \Phi_{\nu} | \hat{h}^{KS} | \Phi_{\mu} \rangle$ y $\mathbf{C}$ es la matriz de coeficientes de los orbitales de KS en la base elegida. Los elementos de $\mathbf{h}^{KS}$ dependen de la densidad y por lo tanto de $\mathbf{C}$, por lo que la resolución de esta ecuación debe realizarse en forma iterativa. Para ello se parte de un conjunto preliminar de coeficientes que permiten calcular la densidad, con esa densidad se construye $\mathbf{h}^{KS}$ y luego se resuelve la ecuación 2.1.47 hasta encontrar una matriz $\mathbf{C}$ autoconsistente.

II.2.1.5.1 Funcionales de Intercambio y correlación

La calidad de los cálculos de DFT responde como puede esperarse a la calidad del funcional utilizado para modelar la energía de intercambio y correlación.

El primer funcional de intercambio y correlación fue propuesto por Hohenberg y Kohn.^[6] El mismo se expresa de la siguiente manera:

$$E_{xc}[\rho] = \int \rho(r) \varepsilon_{xc}(\rho) dr \quad (II.57)$$

donde $\varepsilon_{xc}(\rho)$ corresponde a la energía de intercambio y correlación por electrón en un gas homogéneo de electrones con densidad ρ . La energía de intercambio se calcula en este caso como:

$$\varepsilon_i = -\frac{3}{4} \left(\frac{3}{\pi} \right)^{1/3} (\rho(r))^{1/3} \quad (II.58)$$

y la energía de correlación puede obtenerse a través de una función derivada por Vosko, Wilk y Nusair.^[8] Este funcional se denomina aproximación de densidad local (LDA), o LSDA en la implementación para sistemas de capa abierta.

Los modelos LDA y LSDA están basados en un gas uniforme de electrones, por lo que resulta una buena aproximación en sistemas donde la densidad electrónica cambia lentamente. Sin embargo, esto es difícil que ocurra en un sistema químico real.

Para incorporar las variaciones de la densidad electrónica con la posición, se desarrollaron los funcionales corregidos por gradiente (denominados por sus siglas en inglés GGA: Generalized Gradient Approximation). En estos funcionales, se incorpora el $\nabla\rho(r)$ en la expresión del funcional de intercambio y correlación.

Como ejemplo puede mencionarse el funcional de intercambio y correlación de Perdew-Burke-Ernzerhof (PBE)^[9]

Al igual que en LDA, estos funcionales pueden dividirse generalmente en un término de intercambio y uno de correlación. Ejemplos de funcionales de intercambio son el funcional desarrollado por Perdew y Wang (PW86)¹⁰ y el funcional de Becke (B)^[11], que contiene parámetros empíricos. Como ejemplos de funcionales de correlación pueden mencionarse el funcional de Lee-Yang-Parr (LYP)^[12], y el de Perdew y Wang (PW91).^[13]

Debido a que el método de HF permite obtener el intercambio exacto, se desarrollaron funcionales llamados funcionales híbridos, que incluyen una porción del intercambio de HF con funcionales de intercambio y correlación de GGA. Como ejemplo podemos mencionar el funcional B3LYP y el funcional B3PW91.^[14]

La utilización de la teoría del funcional de la densidad con funcionales GGA permite obtener generalmente buenos resultados para geometrías y otras propiedades moleculares, energías de enlace, frecuencias de vibración, etc.

Sin embargo, a diferencia de la teoría de HF, DFT presenta una limitación severa y es que debido a la inclusión de un funcional de intercambio y correlación aproximado, la teoría deja de ser variacional (es decir, el valor de energía podría ser menor que el verdadero) y los valores de energía dependerán de la elección del funcional elegido. Sin embargo, un aumento en el tamaño de las funciones de base darán valores menores de energía calculada, para cálculos realizados con un mismo funcional, lo que permite en ese caso mantener la idea variacional.

Es habitual entonces, validar los resultados del cálculo realizado, utilizando distintos funcionales de intercambio-correlación (además de funciones de base) y cálculos ab-initio para compararlos entre sí.

Sin embargo, DFT (y otros métodos ab-initio de más alto nivel) sigue siendo actualmente la teoría de elección por los químicos computacionales cuando se desean realizar cálculos de estructura electrónica, dado que es la teoría que ofrece la mejor relación entre la calidad de sus resultados y el costo computacional requerido, especialmente en sistemas conteniendo metales de transición.

En la próxima sección se describirá la implementación de DFT denominada SIESTA, utilizada en algunos cálculos en la presente tesis.

II.2.2 El método SIESTA

En esta sección se presentará el método SIESTA (Spanish Initiative for Electronic Simulations of Thousands of Atoms).^[15]

Este método fue originalmente desarrollado por físicos de materia condensada, por lo que está diseñado para calcular la estructura electrónica de sistemas periódicos. Sin embargo, si se eligen convenientemente los parámetros de la celda, es posible realizar cálculos que corresponden a sistemas no periódicos, que es lo que hemos hecho en este trabajo de tesis.

A diferencia de la mayoría de los esquemas de cálculo basados en la teoría del funcional de la densidad que utilizan funciones de base descritas en forma analítica, este método utiliza funciones bases numéricas descritas en una grilla tridimensional.^[16–18]

La utilización de las mismas otorga varias ventajas: en primer lugar, permite utilizar cualquier tipo de funciones de base, a diferencia de los cálculos realizados con bases analíticas que normalmente utilizan funciones Gaussianas para simplificar la realización de las integrales asociadas. En segundo lugar, las funciones de base se encuentran acotadas en la grilla, es decir

que son forzadas a valer cero más allá de un cierto radio de corte, lo que permite acelerar significativamente los cálculos.

Si bien en este contexto podría utilizarse cualquier tipo de funciones de base, se observó que resulta conveniente la utilización de funciones de base llamadas orbitales pseudoatómicos (PAOs) que consisten en las autofunciones de los átomos con los correspondientes pseudopotenciales. El confinamiento de estas funciones a través del uso de un radio de corte, provoca un incremento en la energía asociada a estos PAOs. Este incremento puede variarse a través de la elección del parámetro ΔE_{PAO} (pseudoatomic orbital energy shift) modificando así la calidad de los resultados obtenidos. Típicamente, valores de ΔE_{PAO} entre 100 y 200 meV, otorgan resultados adecuados, sin embargo en caso de que el sistema contenga metales de transición se recomienda establecer este valor entre 20 y 30 meV.

Por otra parte, el usuario puede determinar la precisión del cálculo indicando la separación entre puntos de la grilla, lo que se indica con el parámetro *Energy Cut Off*, y corresponde a la energía de la onda plana de mayor frecuencia que es posible incluir en la grilla. En cuanto a los funcionales de intercambio y correlación, en el programa se encuentran implementados LDA, LSDA y PBE (GGA).

Para el tratamiento de los núcleos y los electrones internos se utilizan *pseudopotenciales*. La utilización de los mismos permite por un lado, eliminar el costo computacional asociado al cálculo de los electrones internos, y por otro, permitir obtener una densidad de carga suave en toda la grilla, permitiendo además incluir a muy bajo costo los efectos relativistas, que suelen ser relevantes en átomos pesados.

En este contexto, el Hamiltoniano electrónico resulta:

$$H = T + \sum_{\alpha} [V_{\alpha}^{local}(r) + V_{\alpha}^{nl}(r)] + V^H(r) + V^{xc}(r) \quad (II.59)$$

donde α corresponde a los núcleos. La suma en la ecuación (II.59) corresponde al pseudopotencial, el cual posee una contribución local y otra no local.

Con el objeto de evitar el cómputo del término $V_{\alpha}^{local}(r)$, que resulta de difícil convergencia en un arreglo infinito de cargas (debido a la naturaleza periódica del cálculo), se define $V_{\alpha}^{atom}(r)$, el cual se suma al $V_{\alpha}^{local}(r)$, de manera de generar un potencial de átomo neutro (NA). El potencial agregado $V_{\alpha}^{atom}(r)$ corresponde al potencial creado por la densidad electrónica de cada átomo aislado ρ_{α}^{atom} . Podemos definir entonces:

$$V_{\alpha}^{NA} = V_{\alpha}^{local} + V_{\alpha}^{atom} \quad (II.60)$$

$$\rho^{atom}(r) = \sum_{\alpha} \rho_{\alpha}^{atom}(r) \quad (II.61)$$

$$\delta\rho(r) = \rho(r) - \rho^{atom}(r) \quad (II.62)$$

En la ecuación (II.62), $\delta\rho(r)$ corresponde a la diferencia entre la densidad electrónica autoconsistente y la de los átomos aislados, la cual es mucho más pequeña que $\rho(r)$. Siendo $\delta V^H(r)$ el potencial creado por $\delta\rho(r)$, podemos describir el Hamiltoniano electrónico total como:

$$\hat{H} = \hat{T} + \sum_{\alpha} V_{\alpha}^{nl}(r) + \sum_{\alpha} V_{\alpha}^{NA}(r) + \delta V^H(r) + V^{xc}(r) \quad (II.63)$$

De esta manera, el $V_{\alpha}^{local}(r)$ se reemplaza por una suma de potenciales de átomo neutro, más una pequeña corrección, que se agrega para corregir el efecto del agregado de $V_{\alpha}^{atom}(r)$.

Los primeros dos términos de la ecuación (II.63) involucran integrales de dos centros y se calculan al inicio del cálculo, en el espacio recíproco, tabulándose en función de la distancia interatómica. Los términos restantes son calculados en la grilla tridimensional.

La densidad electrónica se calcula de la siguiente manera: En primer lugar, supongamos que $\{\psi_i(r)\}$ son la parte espacial de los spin-orbitales, las cuales pueden describirse como combinación lineal de las funciones de base:

$$\psi_i(r) = \sum_{\mu} \phi_{\mu}(r) C_{\mu i} \quad (II.64)$$

Entonces, la densidad electrónica resulta:

$$\rho(r) = \sum_i n_i |\psi_i(r)|^2 = \sum_{\mu\nu} \rho_{\mu\nu} \phi_{\nu}^*(r) \phi_{\mu}(r) \quad (II.65)$$

donde n_i corresponde al número de ocupación de ψ_i y $\rho_{\mu\nu}$ a los elementos de matriz de la matriz densidad, dados por:

$$\rho_{\mu\nu} = \sum_i C_{\mu i} n_i C_{i\nu} \quad (II.66)$$

Los pasos autoconsistentes establecidos en el programa SIESTA pueden resumirse en:

1. Definición de la base $\{\phi_i\}$ a utilizar y cálculo de los elementos de matriz no

autoconsistentes correspondientes a los términos $\hat{T}$ y V_{α}^{nl} .

2. Estimación inicial de los coeficientes de la matriz de densidad (por ejemplo asumiendo $\rho(r) = \rho^{NA}(r)$)

Obtención de la densidad electrónica $\rho(r) = \sum_{\mu\nu} \rho_{\mu\nu} \phi_{\nu}^*(r) \phi_{\mu}(r)$ y de $\delta\rho(r) = \rho(r) - \rho^{NA}(r)$

Cálculo de $\delta V^H(r)$ a través de la ecuación de Poisson ($\nabla^2 V = -\rho/\epsilon_0$).

3. Cálculo de V^{xc} en la grilla.
4. Evaluación de las integrales asociadas a $\delta V^H(r)$, V_{α}^{NA} y V^{xc} .
5. Resolución del problema de autovalores para encontrar nuevos coeficientes de la matriz densidad.
6. Construcción de una nueva matriz densidad con los nuevos coeficientes y comparación con la matriz anterior. En caso de no verificarse la autoconsistencia, se vuelve al ítem 3.
7. Cálculo de la energía.
8. Cálculo de las fuerzas sobre los átomos.

En resumen, el método SIESTA permite realizar, a través de la utilización de bases numéricas, el uso de pseudopotenciales para los electrones internos, y los funcionales de intercambio y correlación LDA, LSDA o PBE, cálculos de estructura electrónica en forma altamente eficiente para sistemas de mediana cantidad de átomos, basados en la teoría del funcional de la densidad.

II.2.3 Métodos basados en la Mecánica Clásica

Muchos sistemas de interés en química y biología son sistemas moleculares de gran cantidad de átomos, como por ejemplo las proteínas, membranas, etc. Para este tipo de sistemas como se mencionó, aún con el poder computacional existente hoy en día, es 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, lo que se denomina mecánica molecular.

En estos métodos 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 a menudo para el desarrollo de los distintos parámetros requeridos para el cálculo (que se explicarán en la próxima sección). También, muchos de estos parámetros son obtenidos a partir del ajuste de potenciales clásicos de manera tal de reproducir resultados experimentales conocidos.

II.2.3.1 Campos de fuerza clásicos – Simulaciones atomísticas

En los métodos de mecánica molecular atomística, la energía potencial viene dada por una función que se denomina “campo de fuerza”, que no es más que una expresión matemática para la energía potencial dependiente de las coordenadas de los núcleos y de una serie de parámetros.^[19]

Existen también campos de fuerza asociados a descripciones de tipo “grano grueso”, en donde la expresión matemática involucra una energía potencial dependiente de coordenadas de grupos de átomos (como una entidad).

La expresión del campo de fuerza atomístico de AMBER^[20]— el cual fue utilizado para los cálculos realizados en esta tesis — contiene dos clases de contribuciones a la energía potencial, las contribuciones de unión y las de no unión.

Por un lado, las de unión, se computan para átomos que están formando un enlace covalente con otro o como máximo a dos átomos de distancia, e incluyen penalidades energéticas para las desviaciones respecto de su valor de equilibrio de distancias de enlace, ángulos y ángulos diedros. Mientras que las dos primeras se representan a través de un potencial armónico centrado en la distancia de equilibrio, los ángulos diedros o torsiones están descriptos con una función periódica (ver primeros tres términos en la ecuación (II.67)).

Por otro lado, las contribuciones de no unión, se computan entre átomos que están en diferentes moléculas o a más de tres enlaces de distancia, y se dividen en interacciones electrostáticas basadas en la ley de Coulomb, y potenciales de tipo Lennard-Jones, que describe en forma aproximada las interacciones dispersivas y la repulsión originada por el principio de exclusión de Pauli a distancias muy cortas. La expresión de la energía potencial resulta:^[21]

$$\begin{aligned}
E(r^N) = & \sum_{\text{enlaces}} \frac{k_{b,i}}{2} (r_i - r_{i0}) + \sum_{\text{angulos}} \frac{k_{\theta,i}}{2} (\theta_i - \theta_{i0}) \\
& + \sum_{\text{torsiones}} \frac{V_n}{2} \cos(n\varpi - \gamma) \\
& + \sum_{i=1}^{N-1} \sum_{j=i+1}^N \left\{ \varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right\}
\end{aligned} \tag{II.67}$$

El primer término refiere a la contribución de estiramiento de los enlaces, el cual se describe a través de un potencial armónico. En ésta expresión, $k_{b,i}$ corresponde a la constante de fuerza asociada a la unión i y r_{i0} a su distancia de equilibrio.

Si bien una mejor descripción de un enlace químico se obtendría con un potencial como por ejemplo el sugerido por Morse^[22], este contiene un parámetro adicional, que agrega un costo computacional adicional, por lo tanto no se suele utilizar en cálculos MM. Sin embargo, en algunos casos, suele recurrirse a potenciales de esta característica para modelar la formación o ruptura de enlaces, aunque seguirá siendo de forma aproximada. Además, un potencial armónico resulta una aproximación razonable (a segundo orden) al potencial real en la zona cercana al mínimo de cualquier función potencial.

El segundo término corresponde a las contribuciones dadas por las flexiones angulares. Estas también se representan con un potencial armónico de constante $k_{\theta,i}$ y valor de equilibrio θ_i . 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 $k_{\theta,i}$ suelen ser menores que los de las constantes $k_{b,i}$.

El tercer término, es el asociado a las torsiones o ángulos diedros y 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 (II.67) contiene las contribuciones de no unión, y depende de la distancia entre pares de átomos, r_{ij} . La primera parte, corresponde al potencial de Lennard-Jones y contiene los parámetros ε_{ij} , asociado a la profundidad del pozo de energía, y σ_{ij} , correspondiente a la distancia entre los átomos correspondiente al mínimo de energía. La segunda parte, la parte electrostática, contiene como parámetros las cargas sobre los átomos. En el campo de fuerzas de AMBER, la distribución de cargas es representada a través de cargas puntuales q_i ubicadas en las coordenadas de cada átomo. Estas cargas se obtienen partir de

ajustar las mismas para que reproduzcan el potencial electrostático obtenido con un método cuántico (ajuste RESP, Restrained Electrostatic Potential)

Para disminuir el costo computacional asociado a los términos de no unión, se utiliza un radio de corte, de manera tal de no computar las interacciones con átomos que se encuentran a una distancia mayor que el radio de corte elegido.

Para las interacciones de Lennard-Jones, al ser interacciones cuya intensidad decae a distancias cortas, esto no genera un problema siempre que el radio de corte sea lo suficientemente grande. Sin embargo, para las interacciones Coulómbicas, (que son de mayor alcance), en los casos en que se utilizan condiciones periódicas en la simulación, se utiliza una metodología denominada sumas de Ewald^[19], un método originalmente desarrollado para el estudio de cristales iónicos, en su implementación denominada *Particle Mesh Ewald*.

Esta metodología se utiliza para calcular las interacciones electrostáticas totales entre los átomos presentes en una celda unidad, de forma muy eficiente. La energía electrostática se calcula en dos partes: una primera parte, en la cual a las interacciones electrostáticas asociadas a las cargas puntuales se les agrega una distribución de carga Gaussiana, que neutraliza cada carga. Esta parte se calcula en el espacio real y a diferencia de la suma directa de las interacciones electrostáticas, la expresión resultante converge en forma mucho más rápida. Las correcciones necesarias debido a la incorporación de las densidades “neutralizantes”, se calculan en el espacio recíproco.

Las sumas asociadas a los cálculos en el espacio recíproco también presentan una rápida convergencia. 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.^[19]

Finalmente, como puede observarse en la ecuación **(II.67)**, el campo de fuerzas es constituido además por una gran cantidad de parámetros asociados a cada uno de los sumandos en la descripción de la energía potencial. Para el caso de simulación de proteínas, el campo de fuerza de AMBER contiene un conjunto de parámetros por default asociados a aminoácidos, algunos iones, ADN y ARN, lípidos, etc., incluso para distintos estados de oxidación o protonación.

En este contexto, puede hablarse de la transferibilidad del campo de fuerza, entendida como la posibilidad de la utilización de un mismo conjunto de parámetros en otros sistemas similares. Por ejemplo, los parámetros para un residuo de arginina serán los mismos en todas

las proteínas estudiadas, y no será necesaria su parametrización en cada proteína que se desee estudiar. Ese concepto permite utilizar el mismo conjunto de parámetros por residuo en todas las proteínas estudiadas, sin pérdida de generalidad.

En el campo de fuerza de AMBER, los átomos dentro de cada residuo reciben un nombre de átomo y se le asigna 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, en estos casos los parámetros asociados a las uniones, los ángulos y los diedros serán los mismos.

Si bien en el campo de fuerza se incluyen parámetros para todos los aminoácidos naturales y otras moléculas comúnmente encontradas en los sistemas proteicos, para incluir grupos prostéticos como el grupo hemo en diferentes estados de oxidación y coordinación, por ejemplo, será necesario obtener los parámetros necesarios para realizar la simulación.

II.2.3.1.1 Modelos de agua explícita

Existen numerosos modelos de agua para simular el solvente que rodea a las biomoléculas. En primer lugar, los modelos de solvente implícito, describen al solvente circundante como una “jalea electrostática” caracterizada por una constante dieléctrica. Estos modelos tienen la ventaja de tener menor costo computacional, en detrimento del detalle de eventos relacionados a interacciones específicas con el solvente.

Por otro lado, existen modelos de solvente explícitos, que asignan un campo de fuerzas similar a los pertenecientes a los aminoácidos, iones, etc.

La mayor parte de los modelos de agua (no polarizable) asumen una carga positiva sobre los hidrógenos y una carga negativa junto con una interacción de Lennard-Jones en la posición del oxígeno.

En general, se diferencian en la geometría del enlace (distancia HO-H y ángulo H-O-H), y en la magnitud de la distribución de cargas (todos los modelos localizan la carga positiva sobre los átomos de hidrógeno pero difieren en la localización de la/s carga/s negativa/s y en las propiedades físicoquímicas del agua real que se pretende reproducir y contra las cuales se ajustan los parámetros del modelo).^[23]

Entre los más populares se encuentran el modelo TIP3P (por sus siglas en inglés, transferable three point intermolecular potential), el cual fue utilizado en esta tesis. Este modelo fue propuesto por Jorgensen et al. en 1983 y en el cual la carga negativa está localizada sobre el

átomo de oxígeno y las cargas positivas sobre los átomos de hidrógeno. Además, cada molécula de agua se mantiene en una geometría rígida. Un esquema de este tipo de molécula de agua se encuentra en la Figura 1

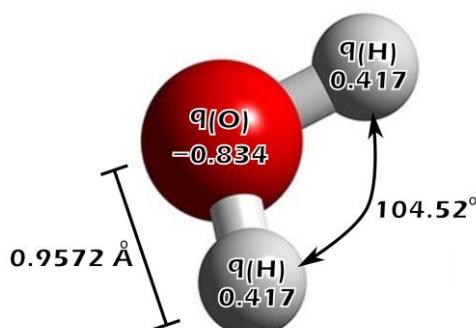


Figura 1: parámetros relevantes (cargas, distancia, y ángulo) del modelo de agua TIP3P

Los parámetros del modelo (por ejemplo, aquellos correspondientes al potencial de Lennard-Jones y las cargas sobre cada átomo) son ajustados de manera tal que reproduzcan el valor de entalpía de vaporización y la densidad del agua líquida en condiciones estándar de presión y temperatura y por lo general este modelo de aguas es el más ampliamente utilizado con campos de fuerzas parametrizados para biomoléculas.^[24]

Si bien existen numerosos modelos de agua, un trabajo de Cerutti et al., en donde se compara la capacidad de los campos de fuerza de biomoléculas para reproducir las propiedades de las estructuras cristalográficas de proteínas en combinación con distintos tipos de modelos de aguas, llega a la conclusión de que la elección del modelo de aguas tiene un efecto despreciable en comparación con la elección del tipo de campo de fuerzas utilizado para la proteína, y ya que estos han sido mayormente parametrizados utilizando un modelo de aguas de tipo TIP3P, es conveniente utilizar este modelo de aguas.^[25]

Otros modelos de agua simples usualmente utilizados son por ejemplo el modelo SPC^[26] o su actualización SPC/E^[27], y el modelo TIP4P^[24], que utiliza cuatro puntos en vez de tres para describir la distribución de cargas. Los valores de los parámetros asociados a los modelos de agua simples se desarrollan de forma de reproducir las propiedades del agua medidas experimentalmente como la densidad, la función de distribución radial, y otras propiedades fisicoquímicas. Como se trata de modelos sencillos, existen gran cantidad de propiedades que no pueden ser descriptas.

Existen modelos más avanzados, 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.

II.2.4 Métodos multi-escala cuántico-clásicos (QM/MM)

En las secciones anteriores se han presentado los aspectos básicos de los métodos basados en la mecánica cuántica y en la mecánica clásica. Los métodos basados en la mecánica cuántica presentan información de la distribución electrónica del sistema y son imprescindibles cuando se desea estudiar procesos reactivos, pero traen aparejado un gran costo computacional. Por otro lado, los métodos llamados de mecánica molecular presentan un costo computacional extremadamente menor, y permiten estudiar fenómenos estructurales asociados a moléculas de gran cantidad de átomos.

Resulta entonces natural pensar que pueden combinarse ambas metodologías para el estudio de un sistema en el que interese una descripción de la densidad electrónica de una cierta región del sistema (llamada región QM, que se debe describir a un nivel mecánico cuántico), y el resto del sistema puede tratarse con mecánica clásica (llamada región MM), como muestra el ejemplo de la Figura 2. Esta idea ha tenido un enorme éxito, lo que ha motivado la adjudicación del premio Nobel de Química en 2013 a Arieh Warshel, Michael Levitt y Martin Karplus por sus contribuciones.

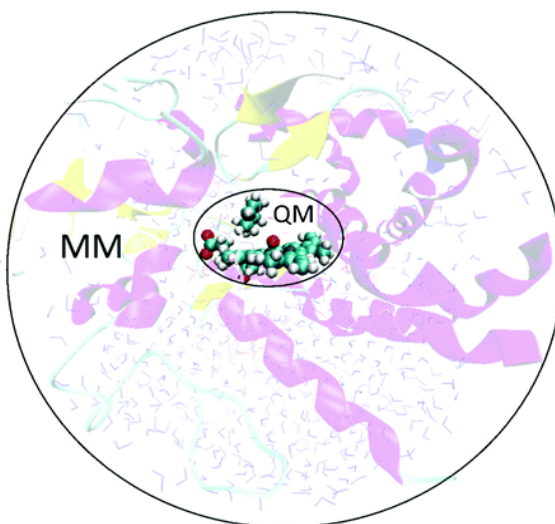


Figura 2: en un cálculo QM / MM de una reacción enzimática, el sistema se divide en dos regiones: la región que se quiere describir con detalle electrónico (región QM, mostrado aquí en representación de bolas y palos) y el resto del sistema (región MM).

Un ejemplo de aplicación de este tipo de metodologías es el estudio de reacciones enzimáticas.^[28–32] En ellas, el sitio activo donde ocurre la reacción debe tratarse necesariamente con mecánica cuántica, pero el resto de la enzima puede tratarse clásicamente. Esto constituye un avance significativo respecto de la utilización de sistemas modelo para tratar este tipo de

problemáticas, en los que solo se incluía el sitio activo y algunos grupos pequeños relevantes. En las metodologías multi-escala se considera el efecto de toda la proteína, en la mayoría de los casos crucial en el desarrollo del fenómeno reactivo.

Para aplicar una metodología QM/MM, en primer lugar es necesario definir la región del sistema que se tratará con mecánica cuántica, que llamaremos subsistema QM. El resto del sistema (subsistema MM) se tratará con mecánica clásica.

En la elección debe tenerse en cuenta que cuanto mayor sea el subsistema QM, mayor calidad tendrá el cálculo, pero también su costo se incrementará significativamente. En el caso de que la simulación a realizar consista en el estudio de una molécula inmersa en un solvente, resulta natural elegir como subsistema QM a la molécula a estudiar y tratar al solvente como subsistema MM. En el caso del estudio de una reacción enzimática, la elección es más compleja. Resulta imprescindible incluir en el subsistema QM todos los átomos que incluyan enlaces que puedan romperse o formarse durante el proceso de interés. Sin embargo, para obtener una buena representación del problema de estudio, normalmente se requiere incluir una mayor cantidad de átomos que simplemente los involucrados en la reacción propiamente dicha.

Dentro de los métodos QM/MM, pueden distinguirse dos clases de metodologías: los esquemas aditivos y los esquemas sustractivos.^[33,34] La metodología utilizada en esta tesis corresponde a un esquema aditivo, por lo que se describirá este tipo de esquema.

En los métodos QM/MM aditivos, el Hamiltoniano consiste en la suma de la contribución del subsistema QM (H_{QM}), el subsistema MM (H_{MM}) y un término de acoplamiento QM/MM ($H_{QM/MM}$):

$$H = H_{QM} + H_{MM} + H_{QM/MM} \quad (II.68)$$

Para el cálculo de la energía E_{QM} (asociada al H_{QM}), debe seleccionarse un nivel de cálculo cuántico. Los cálculos presentados en esta tesis están realizados a nivel de la teoría del funcional de la densidad (DFT). En el Hamiltoniano cuántico no solo se tienen en cuenta las cargas de los núcleos cuánticos, sino que se realiza el cálculo teniendo en cuenta además el potencial electrostático generado por las cargas clásicas, este esquema recibe el nombre de embebimiento electrostático.

La energía E_{MM} se calcula a través del uso de un campo de fuerza. En el caso de la presente tesis, el campo de fuerza utilizado es el campo de fuerza AMBER¹⁸, explicado en la sección II.2.3.

El término clave en el cálculo QM/MM es el término de interacción o coupling: $H_{QM/MM}$. La forma en que se define habitualmente, da cuenta del método QM/MM particular.

En forma general, $H_{QM/MM}$ incluye las interacciones electrostáticas, de van der Waals repulsivos de corto alcance y de unión de los átomos en la frontera, entre los átomos QM y los átomos MM. En el método utilizado en los cálculos realizados en esta tesis, el término de acoplamiento QM/MM consiste en el descrito en la expresión (II.69) para un sistema de A átomos MM y B átomos QM:

$$E_{QM/MM} = \sum_{i=1}^A q_i \int \frac{\rho(r)}{|r - \tau_i|} dr + \sum_{i=1}^A \sum_{j=1}^B \frac{q_i Z_j}{|R_j - \tau_i|} + \sum_{i=1}^A \sum_{j=1}^B 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{|\tau_i - R_j|} \right)^{12} - \left(\frac{\sigma_{ij}}{|\tau_i - R_j|} \right)^6 \right] \quad (II.69)$$

donde, τ_i corresponde a las posiciones de los núcleos MM, R_j a las coordenadas de los núcleos QM.

El primer sumando da cuenta de la interacción electrostática entre la densidad electrónica del subsistema cuántico y las cargas sobre los átomos MM (q_i). El segundo término describe la interacción electrostática entre los núcleos QM, de carga Z_j , y los átomos MM. El tercer término describe las interacciones de van der Waals repulsivos de corto alcance, en la misma forma que lo realiza en campo de fuerza clásico, descrito en la sección II.2.3. Este último término implica que sea requerido obtener los parámetros asociados al potencial de Lennard-Jones (ϵ_{ij} y σ_{ij}) para los átomos del subsistema QM, los que normalmente son obtenidos del campo de fuerza utilizado para representar el subsistema MM.

En la mayoría de los casos en que se desea aplicar una metodología QM/MM, se encuentra el problema adicional de que la frontera entre el subsistema QM y el MM involucra un enlace covalente. Es decir, dos átomos que están unidos covalentemente se encuentran localizados uno en el subsistema QM (A_{QM}) y otro en el MM (A_{MM}). Para el tratamiento de este problema en los cálculos QM/MM realizados en esta tesis se utilizó el método *SPLAM* (*Scaled Position Link Atom Method*).^[35]

En este método, un átomo llamado átomo *Link*, usualmente un átomo de Hidrógeno (H_{link}), se ubica en el enlace por el que atraviesa la frontera. De esta manera, el H_{link} completa la valencia del átomo cuántico ubicado en la frontera. La posición del H_{link} está forzada a mantenerse en el enlace $A_{QM}-A_{MM}$. Para ello, la fuerza sobre el H_{link} se descompone en una

componente perpendicular y otra paralela al enlace $A_{QM}-A_{MM}$. La componente perpendicular se redistribuye entre los átomos A_{QM} y A_{MM} . Por otra parte, para mantener el enlace $A_{QM}-A_{MM}$, los términos de unión, ángulos y diedros involucrados en el enlace son contabilizados a través del campo de fuerza clásico.

Las interacciones de *Lennard-Jones* que involucran a los átomos A_{QM} , H_{link} , A_{MM} son anuladas para pares de átomos separados por menos de tres enlaces.

Por último, para evitar una sobrepolarización de la unión $A_{QM}-H_{link}$ debido a la corta distancia entre el H_{link} y el átomo A_{MM} , se anula la carga sobre el átomo A_{MM} . Esta metodología fue aplicada en diversos trabajos anteriores que incluyen sistemas biológicos.^[31,33,36,37]

SECCIÓN B

METODOS DE EXPLORACION DE LA SEP

II.2.5 Métodos de exploración de la superficie de energía potencial

Como se ha mencionado anteriormente, la superficie de energía potencial (SEP), definida en el contexto de la aproximación de Born-Oppenheimer, es una función escalar que define el valor de la energía potencial del sistema en función de las coordenadas de los núcleos, es decir, en un sistema de N átomos depende de $3N$ variables, asociadas a las coordenadas cartesianas de cada uno de los núcleos. El conocimiento de la superficie de energía potencial permite obtener información cinética y termodinámica de un sistema. Exceptuando sistemas extremadamente simples, la SEP resulta compleja y es imposible conocerla de forma analítica, por lo que se requiere de métodos de simulación computacional para explorarla. En esta sección nos centraremos en distintos métodos desarrollados para explorar la SEP de forma eficiente, que permiten concentrar la atención en el sector de la SEP que se desea explorar para poder responder al interrogante químico de interés. Si bien existe una gran cantidad de metodologías para realizar esta tarea, se desarrollarán solamente las metodologías aplicadas en los cálculos realizados en esta tesis.

II.2.5.1 Optimizaciones de geometría – Búsqueda del mínimo de energía

En cualquier problema de interés químico, siempre va a resultar importante encontrar los puntos mínimos de la superficie de energía potencial, SEP.

En las vecindades de los mínimos de la SEP se encuentran las configuraciones o conformaciones estables, por lo tanto será una región relevante para el sistema de estudio. Para encontrar los mínimos de la SEP, debe aplicarse un algoritmo de minimización.

El problema de la minimización puede verse formalmente de la siguiente manera: se tiene una función $F(x_1, x_2, \dots, x_{3N})$ que corresponde a la SEP, en la cual se quiere encontrar un mínimo. En el mínimo, las derivadas primeras respecto de cada una de las variables $(x_1, x_2, \dots, x_{3N})$ valen cero, y todas las derivadas segundas son positivas:

$$\frac{\partial F}{\partial x_i} = 0 \quad y \quad \frac{\partial^2 F}{\partial x_i^2} > 0 \quad \forall i \in \{1 \dots 3N\} \quad (II.70)$$

Debido a la complejidad con la que varía la SEP con las coordenadas cartesianas, se pueden utilizar diferentes métodos numéricos para obtener los mínimos de la función.

Dentro de los diferentes métodos, pueden distinguirse los métodos derivativos, y los no derivativos. Dentro de los derivativos (que utilizan las derivadas de la función a minimizar), existen métodos que utilizan solo la derivada primera, llamados de primer orden, y los que utilizan además las derivadas segundas, llamados de segundo orden.

Los métodos utilizados en esta tesis corresponden a métodos derivativos de primer orden, en particular los métodos llamados *Conjugate Gradients* y *Steepest Descents*.^[19]

La inclusión de la derivada primera en la determinación del mínimo de energía resulta de mucha utilidad, dado que el gradiente de la SEP, indica la dirección hacia donde se encuentra el mínimo, y su valor absoluto otorga la pendiente en el punto. El gradiente de la SEP corresponde al opuesto de la fuerza, por lo que siguiendo la dirección de la fuerza sobre cada átomo puede gradualmente reducirse la energía del sistema.

El método *Steepest Descents* realiza los movimientos en dirección paralela a la fuerza neta, realizando pasos de una longitud arbitraria, fijada al inicio del cálculo. En caso de que se realice un paso que produzca una reducción de la energía, la longitud del paso se incrementa por un factor. En caso de producirse en contrario un incremento de la energía, se procede a reducir la longitud del paso.

El método *Conjugate Gradients*, por su parte, se diferencia del método de *Steepest Descents*, en que la dirección en la que se realizan el movimiento, corresponde a la dirección de la fuerza neta más la dirección anterior multiplicada por un factor. Este factor se calcula como el cociente entre el módulo del gradiente del paso actual y el paso anterior. De esta manera, no se observa el fenómeno oscilatorio que se observa en el método de *Steepest Descents* en las zonas donde hay mínimos muy angostos.

Por último, debe tenerse en cuenta que estos métodos intentan siempre ir en la dirección de reducción de la energía, por lo que el resultado final dependerá fuertemente de la configuración de partida. En otras palabras, a través simplemente de optimizaciones de geometría como las descritas en esta sección, no se podrán atravesar barreras que conecten distintos mínimos, y el mínimo que se obtenga será el más cercano a la estructura de partida. Resulta entonces imprescindible la utilización de otras metodologías que permitan un muestreo más amplio de la SEP.

II.2.5.2 Búsqueda del camino de transformación

Es de interés en química conocer los aspectos termodinámicos y cinéticos asociados a muchos procesos químicos (por ejemplo, reacciones químicas, cambios conformacionales en proteínas, etc).

Para poder calcular las barreras de activación y moverse por la SEP desde un mínimo a otro, por ejemplo, de reactivos a productos, es necesario desplazarse del mínimo. Esto no puede realizarse con los métodos de minimización presentados en la sección I.2.5.1, dado que, como se mencionó, no permiten atravesar barreras de activación.

Para poder lograr el pasaje entre dos estados estables del sistema, como reactivos y productos, debe definirse lo que se denomina una coordenada de transformación. La misma puede consistir en una distancia, un ángulo, o puede ser muy compleja e involucrar distintos parámetros geométricos del sistema.

Un esquema alternativo es localizar los estados de transición, para lo cual existen algoritmos para tal fin, lo que no requiere de definir a priori una coordenada de transformación.

Sin embargo, esto es una tarea bastante dificultosa, por lo que en esta tesis se emplea el esquema alternativo de mapear el camino de transformación eligiendo una coordenada de reacción adecuada.

Una vez elegida la coordenada de transformación, el objetivo es entonces obtener la variación de la energía potencial en función de la coordenada de transformación, es decir de

esta manera se obtiene el valor de la SEP a través de un camino definido por la coordenada de transformación.

Dependiendo de la elección de la coordenada de transformación, ese camino corresponderá al camino de mínima energía que separa dos mínimos (por ejemplo reactivos y productos de una reacción química, cambios conformacionales de la estructura de la proteína, etc) y el punto máximo de esta curva corresponderá al estado de transición, correspondiente a un punto de ensilladura en la SEP (máximo en la dirección de la coordenada de transformación y mínimo en el resto de las coordenadas)

Para realizar este cálculo, se suele agregar a la energía potencial un término extra, que tiene la siguiente forma:

$$V_R = k (\xi - \xi_0)^2 \quad (II.71)$$

La expresión **(II.71)** corresponde a un potencial armónico, de constante k , en donde ξ_0 corresponde al valor deseado de la coordenada de transformación en un momento de la simulación, y ξ corresponde al valor de la coordenada de transformación.

En la práctica, el valor de ξ_0 se va modificando a lo largo del cálculo, lo que provoca que el valor de ξ se vaya modificando. En cada movimiento de ξ_0 , se realiza una optimización de todo el sistema sujeto a la restricción dada por **(II.71)**, lo que permite obtener el camino de mínima energía. De esta manera se puede obtener el perfil de energía potencial asociado a una transformación de interés.

En este punto resulta relevante destacar que este perfil obtenido mediante minimizaciones, no tiene en cuenta efectos térmicos. Por lo tanto, no provee información termodinámica directamente. Aun así, estos métodos resultan útiles para la búsqueda de caminos de transformación y pueden proveer información cualitativa de los procesos que tienen lugar.

II.2.5.3 Simulaciones de Dinámica Molecular

En esta tesis se realizarán principalmente simulaciones de sistemas proteicos cuya 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 del sistema, sino que existen una serie de mínimos (llamadas conformaciones), que estarán presentes en mayor o menor proporción según sus energías relativas.

Cuando se realiza una medida experimental, todas las conformaciones relevantes estarán representadas en el ensamble de moléculas presente en el sistema.

Para obtener un ensamble se puede hacer uso de la hipótesis ergódica, que supone que a través de una reconstrucción de la evolución temporal de la dinámica de un sistema, el sistema puede evolucionar un tiempo suficiente para que haya podido explorar todas las distintas conformaciones disponibles.

En una simulación de dinámica molecular, 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, es decir, la trayectoria se obtiene a través de la resolución de las ecuaciones diferenciales asociadas a:

$$\frac{\partial^2 x_i}{\partial t^2} = \frac{F_{x_i}}{m_i} \quad (II.72)$$

donde se describe el movimiento de una partícula de masa m_i en la coordenada x_i , con F_{x_i} la fuerza en la dirección x_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. Esto convierte el problema en no resoluble analíticamente por lo que debe recurrirse a métodos numéricos.

El método implementado en el programa AMBER se denomina Algoritmo de Verlet^[38], basado en el método de las diferencias finitas.

En el Algoritmo de Verlet, la integración se parte en pequeños pasos, separados por un intervalo de tiempo fijo δt . Se calcula la fuerza en cada partícula como la suma de las interacciones con el resto de las partículas, y de la fuerza puede calcularse la aceleración, y con ella la nueva posición de las partículas en el siguiente paso.

Durante el δt , la fuerza es considerada constante. En el algoritmo de Verlet, se utilizan las posiciones y aceleraciones al tiempo t ($r(t)$) y las posiciones al tiempo $t - \delta t$, ($r(t - \delta t)$), para determinar las nuevas posiciones a $t + \delta t$, según:

$$r(t + \delta t) = r(t) + \delta t \cdot v(t) + \frac{1}{2} \delta t^2 a(t) + \dots \quad (II.73)$$

$$r(t - \delta t) = r(t) - \delta t \cdot v(t) - \frac{1}{2} \delta t^2 a(t) + \dots \quad (II.74)$$

Sumando las expresiones (II.73) y (II.74), despreciando los términos de orden superior, se obtiene:

$$r(t + \delta t) = 2r(t) - r(t - \delta t) + \delta t^2 a(t) \quad (\text{II.75})$$

Las velocidades no aparecen explícitamente, pero pueden obtenerse de la siguiente manera:

$$v(t) = \frac{r(t + \delta t) - r(t - \delta t)}{2\delta t} \quad (\text{II.76})$$

Debe tenerse especial cuidado en la elección del intervalo de tiempo δt , denominado *time step*. Cuanto menor sea el *time step*, mayor será la cantidad de pasos de simulación que necesitaremos para realizar una simulación de una dada cantidad de tiempo, con lo que incrementaremos el costo computacional. Sin embargo, un intervalo de tiempo demasiado grande dará origen a inestabilidades en el algoritmo de integración.

Por otra parte, debe elegirse un *time step* que sea menor que las escalas de tiempo asociadas a los movimientos que se desean observar y que permita describir el movimiento de los átomos en la molécula.

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 clásicamente, los movimientos más rápidos están asociados a las vibraciones de los enlaces, que se encuentran en el orden de los fs. Por lo tanto, un intervalo de tiempo de 0.1-0.2 fs resulta apropiado. Las vibraciones 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 *time step* y así acelerar los cálculos consiste en utilizar el algoritmo SHAKE^[39], que permite fijar estas uniones en sus posiciones de equilibrio.

II.2.5.3.1 Simulación de condiciones experimentales:

temperatura y presión

La temperatura y la energía cinética (K), están relacionadas de la siguiente manera:

$$K = \sum_{i=1}^N \frac{m_i v_i^2}{2} = \frac{3k_b T}{2} N \quad (\text{II.77})$$

donde v_i y m_i corresponden a la velocidad y la masa de la partícula i y N el número de partículas del sistema. Por lo tanto, una manera simple de mantener la temperatura constante consiste en escalar las velocidades.

Uno de los métodos más utilizados para regular la temperatura en una simulación de dinámica molecular es el algoritmo de Berendsen^[40]. El mismo consiste en acoplar el sistema a un baño externo cuya temperatura se fija en la temperatura deseada, que remueve o entrega calor según sea necesario. Supongamos que a un tiempo dado t el sistema posee una temperatura $T(t)$. Puede verse de la ecuación (II.77), que si las velocidades se escalan por un factor λ , el cambio asociado en la temperatura resulta:

$$\Delta T = (\lambda^2 - 1)T(t) \quad (\text{II.78})$$

En este método, las velocidades se escalan en cada paso, de manera que el cambio en la temperatura es proporcional a la diferencia entre la temperatura del sistema y la temperatura del baño:

$$\frac{\partial T(t)}{\partial t} = \frac{1}{\tau} [T_{\text{baño}} - T(t)] \quad (\text{II.79})$$

donde τ es un parámetro de acoplamiento que determina cuan fuertemente el baño está acoplado al sistema. El cambio de temperatura entre dos pasos sucesivos separados por un *time step* δt , corresponde a:

$$\Delta T = \frac{\delta t}{\tau} [T_{\text{baño}} - T(t)] \quad (\text{II.80})$$

Con esta expresión, el factor de escalamiento de las velocidades λ resulta:

$$\lambda^2 = 1 + \frac{\delta t}{\tau} \left[\frac{T_{\text{baño}}}{T(t)} - 1 \right] \quad (\text{II.81})$$

Cuanto más grande sea este parámetro, más débil será el acoplamiento, por lo que la fluctuación en la temperatura será mayor. Normalmente, la utilización de un factor de acoplamiento entre 100 y 1000 veces más grande que el *time step* elegido otorga buenos resultados para los sistemas estudiados en esta tesis.

Por otra parte, es también deseable realizar simulaciones de dinámica molecular en condiciones de presión constante. Para mantener la presión constante, el sistema debe variar el volumen de la celda en la que se realiza la simulación.

Berendsen propuso entonces un esquema similar al planteado para la regulación de la temperatura, por el cual se acopla el sistema a un baño equilibrado a una cierta presión. El volumen de la celda se escala entonces en un factor λ , lo que equivale a escalar las coordenadas atómicas en $\lambda^{1/3}$. Resulta entonces:

$$\lambda = 1 - \kappa \frac{\delta t}{\tau_p} (P - P_{baño}) \quad (II.82)$$

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.

II.2.6 Análisis avanzados de las simulaciones

II.2.6.1 RMSD

Luego de realizada una simulación de dinámica molecular, 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.

La visualización de la trayectoria a través de algún programa de visualización molecular aporta gran cantidad de información sobre el sistema. Con ella se pueden analizar la evolución de parámetros geométricos claves, como distancias, ángulos, diedros, etc.; interacciones de puente hidrógeno e hidrofóbicas.

Por otro lado, en el caso de una macromolécula, puede observarse la dinámica de la estructura terciaria, aunque los movimientos globales asociados a ella muchas veces requieren un largo tiempo de simulación para ser observados.

El programa de visualización utilizado durante toda la tesis fue el programa Visual Molecular Dynamics^[48] (VMD), el cual presenta una gran cantidad de herramientas muy útiles para el análisis de una simulación de dinámica molecular.

Un parámetro que permite estudiar la estabilidad de la estructura a lo largo de la simulación es la raíz cuadrada de desviación cuadrática media (RMSD) de la estructura a lo largo de la simulación, respecto a una estructura de referencia, la cual se calcula de la siguiente manera:

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N \delta_i^2} \quad (II.83)$$

Donde N corresponde al número de átomos en el sistema, y δ_i a la distancia entre las posiciones del átomo i en dos estructuras distintas, una de ellas la estructura será la estructura de referencia.

Realizando este cálculo para cada una de las “fotos” de la simulación, se obtiene el valor RMSD en función del tiempo. La estructura de referencia puede ser la estructura cristalográfica, por ejemplo, o la estructura promedio, lo que otorga una noción de la magnitud de las fluctuaciones del sistema a lo largo de la dinámica.

II.2.6.2 Identificación de los Sitios de hidratación (WS)

La interacción de una proteína con las moléculas de agua se puede estudiar a partir de la formación de los denominados sitios de hidratación (o WS por sus siglas en inglés “Water Sites”) que se definen como aquellas regiones del espacio con la capacidad de alojar moléculas de agua con mayor probabilidad que el seno del solvente.

Con el objetivo de identificar la presencia de los WS en alguna región espacial de la proteína, se utilizó la siguiente estrategia de clustering adaptada de trabajos anteriores de nuestro grupo^[41,42] e implementada en el módulo WATCLUST^[43] en el programa VMD. El cálculo necesita una trayectoria (conjunto de “fotos”) derivada de una dinámica molecular con moléculas de agua explícitas.

Posteriormente, se realiza un alineamiento de la región de interés (en donde se desean identificar los WS) para eliminar grados de libertad innecesarios para el cálculo.

Solo se considerarán las moléculas de agua interactuando en la región de interés a lo largo de la simulación y solo se consideraran las posiciones de los átomos de oxígeno de esas moléculas de agua interactuantes.

El algoritmo comienza tomando como centro una molécula de agua de la primera foto y realiza los siguientes pasos:

- a) Encuentra todas las moléculas de agua que están cerca de ella (a menos de un valor de corte ajustable, llamado “dist”), en todas las fotos subsiguientes.
- b) Toma luego como centro cada una de las moléculas de agua encontradas en a) (sin tener en cuenta la primera molécula de agua tomada como centro) y encuentra nuevas moléculas de agua.
- c) Si dentro del umbral para estas nuevas moléculas de agua encuentra otras nuevas moléculas de agua, realizará el mismo procedimiento sobre cada una de esas moléculas de agua recién encontradas, no considerando las moléculas de agua tomadas como centro anteriormente.
- d) Finalmente, el proceso de búsqueda concluirá cuando no se encuentren más moléculas de agua dentro del umbral de “dist” en ninguna foto.

Posteriormente, las aguas encontradas son consideradas como formando un cluster. Si el número de aguas pertenecientes al cluster es igual o más grande que el valor del parámetro "watnumbermim", el cluster definirá un WS.

Cada foto puede contribuir entonces con no más de una molécula de agua al cluster.

Obsérvese que el parámetro "dist" debe ser lo suficientemente pequeño para evitar un WS excesivamente disperso, pero lo suficientemente grande como para evitar el hallazgo de muchos WS que están demasiado cerca entre sí. Los valores recomendados están en el rango de 0.17-0.20 Å.

Por otra parte, el parámetro "watnumbermin", es decir, el número mínimo de moléculas de agua que un grupo debe albergar para ser considerado un sitio de agua, representa indirectamente un umbral de tiempo: debe haber moléculas de agua durante un cierto tiempo en la región donde el grupo se encuentra con el fin de establecer un sitio real de agua. Este parámetro suele establecerse de tal manera que hay una molécula de agua encontrada dentro del sitio en al menos 5-15% de las fotos totales (o tiempo de simulación), aunque esta decisión no altera el algoritmo de búsqueda en sí (sólo determina qué clústeres son efectivamente considerado como WS).

Se pueden caracterizar a los WS, con un parámetro llamado (WFP, Water finding probability): probabilidad de encontrar una molécula de agua en la región definida por el WS (utilizando un valor de volumen arbitrario de $1 \text{ \AA}^3$, es decir, un radio de aproximadamente 0,6Å) normalizado con respecto al del agua a bulk. El radio, y por lo tanto el volumen utilizado para calcular el WFP, puede ser modificado por el usuario a través del parámetro "WFRr".

II.2.7 Cálculos de Energía Libre

En la presente tesis se desarrollará en el capítulo III los detalles teóricos y técnicos acerca de un método de cálculo de energía libre, basado en la realización de simulaciones de dinámica molecular guiada. Es por eso que a modo de resumen se explicarán los aspectos más relevantes del cálculo de energía libre en simulación computacional.

Como ya se ha mencionado en la introducción de esta tesis, la información que provee la energía libre y su variación asociada a un proceso, resulta esencial en un sistema químico para conocer la factibilidad y la velocidad de que ocurra el mismo. En particular, si la evolución del

proceso puede estar descrito por un conjunto de coordenadas del sistema, entonces lo que se busca obtener es el perfil de energía libre (o PMF) asociado.

Dependiendo de las características del sistema de estudio será conveniente la utilización de métodos denominados “sesgados” o “no sesgados”.

Los métodos “sesgados”, son técnicas que restringen el sistema a recorrer la coordenada de transformación que describe al proceso de interés. Existen numerosos métodos sesgados^[19], uno de ellos es el método de Dinámica Molecular Guiada (*Steered Molecular Dynamics, sMD*) el cual se basa en la realización de un trabajo sobre el sistema para recorrer la coordenada de transformación. Este método fue el más utilizado y estudiado en la presente tesis para obtener perfiles de energía libre.

Los métodos “no sesgados”, evalúan la probabilidad de obtener cada valor de la coordenada de transformación a lo largo de una simulación. La energía libre suele obtenerse a través de la siguiente expresión:

$$-\beta G(\xi) = \ln[P(\xi)] \quad (II.84)$$

Con $\beta = (k_b T)^{-1}$, donde k_b corresponde a la constante de Boltzman y T a la temperatura absoluta, G es la energía libre de Gibbs, y ξ es una coordenada de transformación asociada a la energía libre que se desea calcular. Un ejemplo de este tipo de métodos es el de Muestreo de Ligando Implícito, descrito en capítulos posteriores.

II.2.7.1 Dinámica Molecular Guiada (sMD)

El método de Dinámica Molecular Dirigida (sMD, por sus siglas en inglés), se basa en el agregado de un término armónico a la energía potencial del sistema.

Típicamente, este potencial es de forma cuadrática, con un centro de la cuadrática que se mueve a lo largo de la simulación, según:

$$V(r) = \frac{k}{2} [\xi - (\xi_0 + v \cdot t)]^2 \quad (II.85)$$

Donde ξ es la coordenada de transformación (asociada al perfil que desea calcularse), ξ_0 es el valor inicial de la coordenada y v corresponde a la velocidad con la que se mueve la coordenada.

De esta manera, el sistema es “obligado” a recorrer la coordenada de transformación. Por lo tanto se realiza un trabajo externo, W sobre el sistema. Partiendo entonces de distintas

condiciones iniciales (de posiciones y velocidades) con ξ_0 fijo, se pueden obtener un conjunto de valores de trabajo acumulado W_i a lo largo de la coordenada de transformación.

Finalmente, a partir del conjunto de trabajos y mediante la relación demostrada por Jarzynski^[44] (ecuación **(II.86)**) se puede obtener el perfil de energía libre asociado a la coordenada de transformación:

$$e^{[-\beta\Delta G]} = \lim_{N \rightarrow \infty} \langle e^{[-\beta W_i]} \rangle_N \quad (\text{II.86})$$

donde W_i , es el i -ésimo trabajo para llevar al sistema a lo largo de todo el proceso.

En la práctica, deben realizarse en primer lugar simulaciones de dinámica molecular del estado inicial. Luego de asegurarse que se ha realizado una exploración eficiente del estado inicial, se obtienen a partir de las trayectorias las estructuras iniciales para el cálculo sMD.

Se realizan entonces un número de simulaciones partiendo del estado inicial hacia el estado final. Debe seleccionarse con cuidado la velocidad v , dado que la trayectoria debe realizarse en forma suave, pero una velocidad extremadamente lenta incrementa significativamente el costo computacional del cálculo.

Se registra en cada una de ellas el trabajo en función de la coordenada, y al finalizar la forma más habitual de proceder es la de realizar el promedio exponencial (asociado a la ecuación **(II.87)** considerando el promedio discreto) para obtener el perfil de energía libre.

Dado que la ecuación de Jarzynski es válida solamente si se realiza un número infinito de simulaciones y que esto es impracticable, debe realizarse una cantidad suficiente de simulaciones en las que pueda considerarse que se ha logrado la convergencia del cálculo, es decir, que la realización de nuevas simulaciones no modifica significativamente el resultado del perfil final. El método sMD se encuentra implementado en el programa AMBER^[45], utilizado para realizar los cálculos de dinámica molecular clásica y perfiles de energía libre presentados en esta tesis. Los aspectos teóricos de la teoría subyacente a esta metodología se describirán con más detalle en el capítulo III.

II.2.7.2 Muestro de Ligando Implícito (ILS)

El enfoque de la técnica de muestreo de ligando implícito (ILS)^[46,47] está basada en el cálculo del PMF de colocar un ligando en diferentes puntos de la matriz de una proteína.

La técnica de ILS resulta muy útil para la determinación de canales de migración, que se entienden como zonas o canales de baja energía libre (o alta probabilidad) para la ubicación de ligandos, que conectan el sitio activo con el solvente.

La metodología se basa en la suposición de que ligandos pequeños sin carga o dipolo importante interactúan débilmente con la proteína y prácticamente no modifican la dinámica de esta. Basado en esta hipótesis, utiliza una MD realizada en ausencia del ligando y luego calcula el PMF considerando de manera implícita al ligando, es decir, como una pequeña perturbación en el sistema.

El valor del PMF se define de la siguiente manera

$$G(r) = -\beta^{-1} \ln[\rho(r)] \quad (II.87)$$

Donde $G(r)$ es el PMF asociado a ubicar un ligando en la posición r , el cual tendrá una probabilidad de $\rho(r)$ y un factor de normalización ρ_0 . $\beta^{-1} = k_B T$.

En el ensamble NPT, $\rho(r)$ vale

$$\rho(r) = \frac{\int dV \int d^{3N} p \int d^{3N} q \int d^3 p' \int d^3 r' e^{-\beta[H(p,q,p',r')+PV]} \delta^3(r' - r)}{\int dV \int d^{3N} p \int d^{3N} q \int d^3 p' \int d^3 r' e^{-\beta[H(p,q,p',r')+PV]}} \quad (II.88)$$

Donde la suma se realiza en todo el volumen y todas las coordenadas (posiciones y momentos) del sistema (p , q refiere a coordenadas de la proteína y solvente, y p' y q' refiere a coordenadas del ligando). H es el Hamiltoniano de todo el sistema.

Separando los grados de libertad del ligando se obtiene que

$$\rho(r) = \frac{\int dV \int d^{3N} p \int d^{3N} q \int dp' \int d\Omega \int d^3 r' e^{-\beta[H(p,q,p',r',\Omega)+PV]} \delta^3(r' - r)}{\int dV \int d^{3N} p \int d^{3N} q \int dp' \int d\Omega \int d^3 r' e^{-\beta[H(p,q,p',r',\Omega)+PV]}} \quad (II.89)$$

Donde r' corresponde a la posición del centro de masa del ligando, p' es el momento del centro de masa y Ω corresponden a los grados de libertad restantes del ligando.

Ahora bien, frente a la presencia del ligando, el Hamiltoniano del sistema se verá modificado una cantidad igual a la energía de interacción proteína ligando $\Delta E(q, r, \Omega)$, la energía cinética $K(p')$ y la energía potencial del ligando $U(\Omega)$.

Para el caso en particular de la presente tesis, en donde no se utilizaron ligandos cargados el término de interacción que se consideró fue de tipo Lennard-Jones.

A partir de lo anterior, se obtiene la siguiente expresión

$$\rho(r) = \frac{\int dV \int d^{3N} p \int d^{3N} q \int d\Omega e^{-\beta[H_0(p,q)+PV]} e^{-\beta[\Delta E(q,r,\Omega)+U(\Omega)]} \int dp' e^{-\beta K(p')}}{\int dV \int d^{3N} p \int d^{3N} q \int d\Omega \int d^3 r' e^{-\beta[H_0(p,q)+PV]} e^{-\beta[\Delta E(q,r,\Omega)+U(\Omega)]} \int dp' e^{-\beta K(p')}} \quad (II.90)$$

Por otro lado, en el ensamble NPT el valor medio de un observable A puede calcularse a partir de

$$\langle A(r) \rangle_{NPT} = \frac{\int dV \int d^{3N} p \int d^{3N} q e^{-\beta[H(p,q)+PV]} A(p, q, r)}{\int dV \int d^{3N} p \int d^{3N} q e^{-\beta[H(p,q)+PV]}} \quad (II.91)$$

Por lo tanto, combinando las ecuaciones (II.90) y (II.91) para la probabilidad se obtiene

$$\rho(r) = \frac{\langle \int d\Omega e^{-\beta[\Delta E(r,\Omega)+U(\Omega)]} \rangle_{NPT}}{\langle \int d^3 r' \int d\Omega e^{-\beta[\Delta E(r,\Omega)+U(\Omega)]} \rangle_{NPT}} \quad (II.92)$$

finalmente el PMF es

$$G(r) = -\beta^{-1} \ln \left[\frac{\langle \int d\Omega e^{-\beta[\Delta E(r,\Omega)+U(\Omega)]} \rangle_{NPT}}{\langle \int d\Omega e^{-\beta U(\Omega)} \rangle_{NPT}} \right] \quad (II.93)$$

En donde se considero que el PMF vale cero cuando el ligando es ubicado en cualquier posición en el vacío (es decir, $\Delta E(q, r, \Omega) = 0$)

Para el caso de ligandos pequeños, como los usados en la presente tesis, los enlaces de unión se mantienen fijos, por lo que los únicos grados de libertad Ω que restan, corresponderán solo a rotaciones del ligando.

En este caso $U(\Omega)$ será constante, por lo que a partir de la ecuación (II.93), el PMF toma la siguiente forma.

$$G(r) = -\beta^{-1} \ln \left[\frac{\langle \int d\Omega e^{-\beta[\Delta E(r,\Omega)]} \rangle_{NPT}}{\langle \int d\Omega \rangle_{NPT}} \right] \quad (II.94)$$

Para el cálculo se utiliza la siguiente expresión que es una versión discretizada de la ecuación (II.94).

$$G(r) = -(\beta^{-1}) \ln \left[\sum_{m=1}^M \sum_{k=1}^C \frac{e^{-\beta[\Delta E(r_m, \Omega_k)]}}{M \cdot C} \right] \quad (II.95)$$

Donde la suma está realizada sobre hasta una cantidad total de M fotos de la dinámica y C distintas orientaciones del ligando.

En conclusión, a partir de la ecuación (II.95) se puede calcular el PMF considerando distintas posiciones y orientaciones al ligando.

El resultado final de este cálculo será una grilla tridimensional de puntos, cada punto tendrá asociado un valor de energía libre. La manera típica de representación es en

isosuperficies de valor de energía libre constante. Variando el valor de corte en esta representación se podrán visualizar los canales de migración presentes en la proteína.

En la práctica el sistema se divide en celdas que forman una red tridimensional cuya resolución se debe escoger según la precisión que se requiera. Normalmente la resolución está en celdas entre 0.5 y 1 Å. Cuanto más fina sea la grilla, más suaves serán las isosuperficies de energía libre que permiten obtener sitios de unión y canales a lo largo de la matriz proteica.

Otro parámetro que debe escogerse es el número de estructuras que se utilizaran para obtener la energía libre en cada punto. Para obtener resultados razonables se utilizan entre 1000 y 10000. Según el ligando para el cual se quiera obtener la energía libre, se deben especificar los parámetros ϵ y σ asociados al mismo. La distancia entre los átomos que lo forman, los valores para ϵ y σ del potencial de L-J. En la presente tesis los cálculos por ILS se realizaron con el programa VMD^[48]

II.3 Métodos experimentales

II.3.1 Técnica de láser flash fotólisis (LFP)

Las hemoproteínas han sido las proteínas más estudiadas en biofísica, y se han convertido en paradigmas para la caracterización de procesos biomoleculares fundamentales como la unión de ligandos, actividad enzimática y transiciones conformacionales.

La presencia del grupo hemo prostético a la matriz proteica da a las hemoproteínas un espectro de absorbancia característico, el cual es ligando sensible en una región convenientemente ubicada en el rango UV-visible. Esta característica ha ofrecido una poderosa y sensible herramienta para la investigación de funciones moleculares asociadas a estas proteínas.

El enlace Fe-Ligando es en mayor o menor medida fotolábil. En particular, los estados ferrosos hexacoordinado (con ligandos como O₂, NO, CO, etc) son muy fotolábiles y el estado pentacoordinado reactivo puede ser generado de manera transiente por incidencia con un láser pulsado.

Posteriormente, los ligandos fotodisociados rápidamente se re-unirán al grupo hemo y el proceso puede ser monitoreado con métodos de absorbancia transiente.

La re-unión de ligandos a hemoproteínas en solución mediante la técnica de láser flash fotólisis ha sido extensamente usada a lo largo de los años para entender los mecanismos moleculares subyacentes a la interacción proteína ligando.

En particular, en la presente tesis se ha utilizado como ligando “sonda” al CO. Una propiedad muy útil del complejo ferroso hemo-CO es que es fácilmente fotodisociable por luz visible, la cual cliva el enlace Fe-Ligando sin reacciones secundarias. La diferencia entre la absorbancia del complejo carboxy y deoxy puede ser explotada para monitorear la evolución temporal de la reacción de re-unión.

Como el ligando rápidamente se re-une al hemo, la fotólisis puede ser ciclada y un promedio sobre la señal puede ser realizado. Después de la disociación del hierro del hemo, los ligandos o bien se reúnen desde las cavidades internas de la matriz proteica en un proceso que se denomina recombinación geminal, o escapan hacia el solvente a través de canales de salida.

Las cinéticas de re-unión de ligandos reflejarán entonces poblaciones en donde el ligando se re-unió desde el sitio activo, cavidades internas de la proteína o incluso desde el seno de la solución (fuera de la matriz proteica)

Un número muy grande de estudios experimentales han caracterizado las cinéticas de re-unión bajo una variedad de condiciones experimentales para mioglobina,^[49–52] hemoglobina^[53–55] y más recientemente, a otras hemoproteínas como por ejemplo neuroglobina.^[56,57]

En particular, en la presente tesis se estudiaron las constantes de asociación bimolecular (k_{on}) del ligando CO en tres proteínas diferentes pertenecientes a organismos adaptados diferencialmente a condiciones de temperatura.

A continuación se describirán los principios básicos de la técnica

II.3.1.1 Principios básicos de los experimentos de LFP

Las técnicas de absorción de nanosegundos resueltas en el tiempo se basan en el uso de un haz de luz sonda que interroga la muestra antes y después de la excitación con un láser de nanosegundo (Laser Flash Fotólisis).^[58–60]

Como haz de luz sonda suele generalmente usarse una fuente luminosa de alta frecuencia (de onda continua, CW) un con un gran ancho de banda, dentro del rango espectral UV-vis-NIR, donde las transiciones electrónicas transientes son sondeadas.

El espectro UV-violeta cercano es de especial interés, corresponde a la zona de absorción del hemo llamada banda de Soret. Esta zona del espectro presenta una elevada sensibilidad al estado de coordinación del hierro del hemo, así como también a los cambios estructurales que lo circundan.

El parámetro experimental es el cambio en la absorbancia (ΔA), que está relacionada con los cambios en la luz transmitida del haz de luz sonda antes $I(t < t_0)$ y después $I(t)$ de la excitación del láser, de acuerdo con las siguientes igualdades:

$$-\log \frac{I(t)}{I(t < t_0)} = -\log \frac{\frac{I(t)}{I(0)}}{\frac{I(t < t_0)}{I(0)}} = -\log \frac{T(t)}{T(t < t_0)} = \Delta A(t) \quad (II.96)$$

Donde $T(t < t_0)$ y $T(t)$ corresponden a la transmitancia antes y después de la excitación del láser, respectivamente. $I(0)$ corresponde a la intensidad de luz en ausencia de muestra.

Por otro lado, en general para el caso de hemoproteínas se parte de una muestra con el complejo hexacoordinado formado entre la proteína y el ligando, y mediante el destello del láser se produce la fotólisis del complejo, liberando el complejo del sitio de coordinación, por lo tanto la diferencia de absorbancia puede asociarse a la diferencia entre las absorbancias del complejo pentacoordinado y la del hexacoordinado (ecuación (II.97))

$$\Delta A(t) = A(5c) - A(6c) \quad (II.97)$$

II.3.1.1.1 Esquema experimental del equipo de Laser Flash

Fotólisis

El esquema experimental típico se presenta en la Figura 3A, este tipo de esquema permite una detección versátil y sensible de trazas de absorbancias transientes a una longitud de onda o espectros resueltos en el tiempo con resolución temporal limitada por el láser.

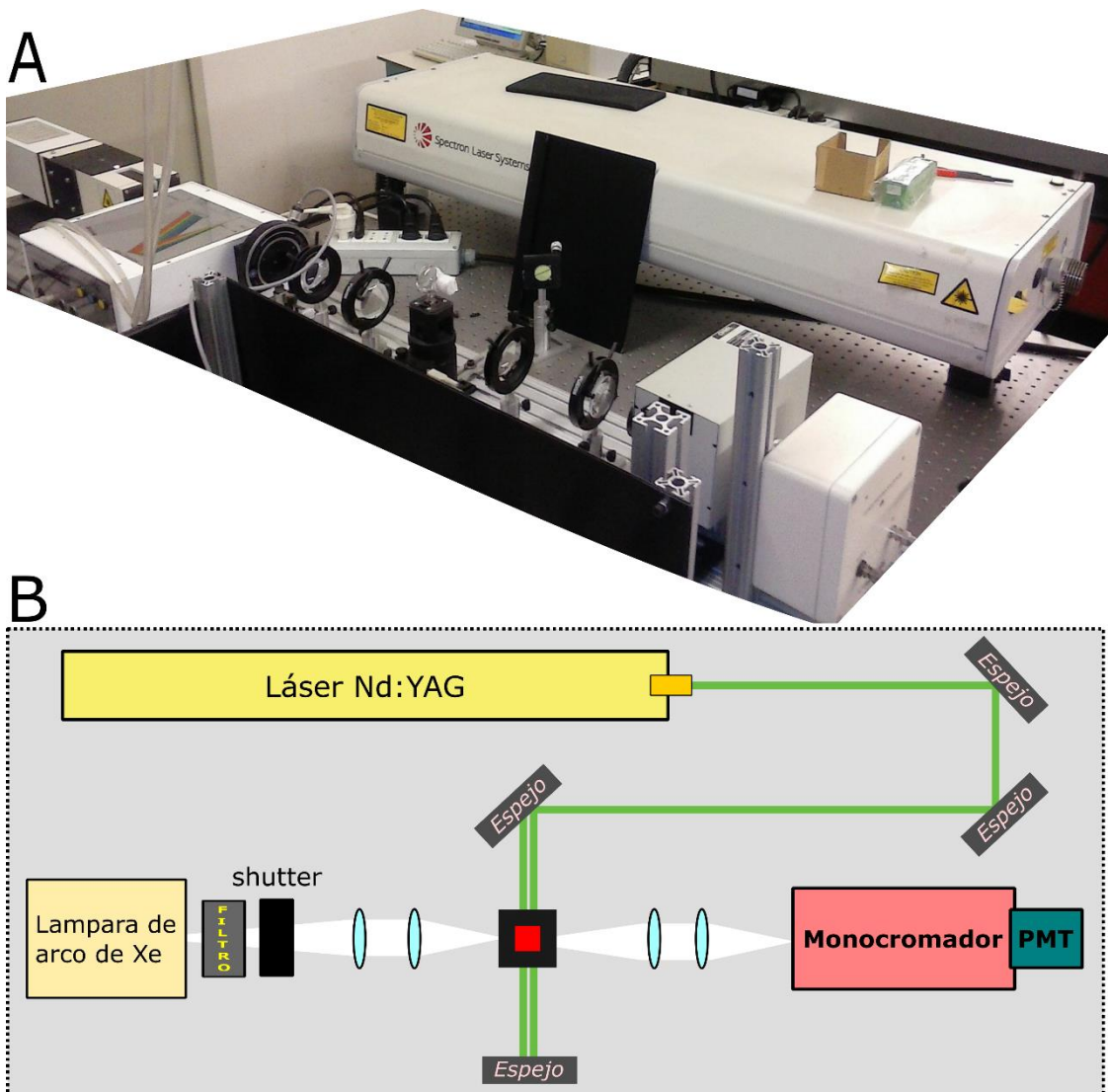


Figura 3: (A) foto del equipo de LFP utilizado. (B) esquema del equipo utilizado.

Una fuente conveniente para la fotólisis de los complejos hemo-CO es la del segundo armónico (532nm) de un láser Nd:YAG. Para los experimentos realizados en esta tesis, se utilizó un láser Surelite II-10 (Continuum).

La polarización lineal del láser es convertida a circular mediante un polarizador de cuarto de onda para minimizar los efectos de fotoselección (en donde las moléculas excitadas y los fotoproductos primarios se forman preferentemente en orientaciones para las cuales el momento de transición del material de partida es paralelo a la polarización del láser)

En general, se obtiene una fotólisis completa con los esquemas descriptos más abajo y utilizando pulsos de láser entre 20-40 mJ, dependiendo de la muestra. La necesidad de obtener condiciones de fotólisis completa es simplemente debido al requerimiento de trabajar con concentraciones conocidas de reactivos (deoxy hemo y CO) para el análisis cuantitativo.

En el esquema mostrado en la Figura 3B los cambios de absorbancia son monitoreados usando un haz de luz sonda de salida continua (CW) monocromática de 75 W de lámpara arco de Xe.

El haz es enfocado hacia el holder de la muestra, luego colimada y finalmente se la hace ingresar en las rendijas de entrada de un monocromador/espectrógrafo MS257 (LOT-Oriel).

En uno de los dos puertos del MS257 es usado para monitorear la cinética a una longitud de onda simple. La luz monocromática es colectada desde la rendija de salida mediante óptica dedicada después de haber sido pasada por un filtro dicróico (Omega optical) y enfocado en un PMT de 5 etapas (Applied Photophysics).

También en este esquema se encuentra posicionado un shutter rápido (Vincent Associates, Uniblitz VS35) entre la salida de la lámpara de Xe y el holder de la muestra.

Usualmente las mediciones consisten en monitorear la diferencia de absorbancia entre el fotoproducto menos el complejo carboxy a varios tiempos de retraso posteriores a la fotodisociación. Dependiendo del espaciado de tiempo, 100 a 120 medidas espaciadas logarítmicamente en el tiempo son registrados para cubrir el curso temporal completo de la cinética de re-unión de CO, desde 10 ns después del destello del láser hasta 0.1 s.

Los valores de absorbancia son obtenidos promediando 100-200 disparos simples en cada delay. La sincronización del experimento en conjunto (disparo del láser, y apertura del shutter) es logrado mediante hardware dedicado.^[61]

También para este esquema, la señal del voltaje es digitalizada por un osciloscopio digital (LeCroy LT374, 500MHZ, 4 GSs-1; LeCroy 9370, 1 Ghz, 1GS s-1)

La velocidad de repetición es un parámetro relevante ya que se desea que la muestra se recupere completamente entre los sucesivos destellos del láser. Por lo tanto, dependiendo de la muestra, suelen ser necesarias velocidades de repetición del orden de 0.1 HZ. Además el shutter también ayuda a evitar que la muestra se dañe si es irradiada continuamente por luz blanca.

La cubeta es de 0.2 cm de paso óptico, y se utiliza un elemento peltier (FLASH 100, Quantum Northwest, Inc.) para termostatizar la muestra (con una precisión de 0.1°C).

Los dispositivos de detección son los que condicionan la resolución del experimento. Muchos detectores se encuentran disponibles hoy en día que permiten alcanzar la resolución

en el nanosegundo. Los PMTs (tubos fotomultiplicadores) de 5 etapas son los tradicionalmente usados en LFP.

Cuando estos detectores son expuestos a monitorear una luz intensa durante unos pocos cientos de ms, estos detectores pueden resistir a corrientes de 2-4 mA sin presentar fatiga^[62]. Estas corrientes resultan en señales de 100-200 mV cuando son alimentados dentro de 50 ohm de carga de un osciloscopio digital, sin la necesidad de una amplificación, y por lo tanto preservando un ancho de banda de señal máxima.

Además, los modernos osciloscopios tienen muy baja capacitancia (típicamente < 20 pF), los cuales resultan en valores RC que permiten la resolución temporal en el nanosegundo. Las longitudes de tiempo más largas que unos pocos cientos de ms pueden ser accedidas limitando la corriente PMT (por ejemplo usando intensidad de luz baja) y usando mayores resistencias de cargas.^[58]

El acceso a una resolución temporal en escalas de tiempo pequeñas también estará limitada por la intensidad de luz dispersada por el láser. Esta dispersión puede ser disminuida con una cuidadosa ubicación de los dispositivos ópticos, y la inserción de uno o más dispositivos que disminuyan las reflexiones de las ópticas incluyendo paredes en la cubeta.

Por otro lado, el uso de un monocromador más un espejo dicróico antes del detector también reduce la luz dispersada a niveles razonables.

Disminuida la dispersión del láser como se mencionó anteriormente, el límite de resolución temporal quedará determinado por los sistemas de detección.

Las señales que son típicamente de interés están en general extendidas en un rango de tiempo de varios órdenes de magnitud. Por lo tanto, el voltaje tiene que ser muestreado a una alta velocidad para los primeros pocos microsegundos después de la excitación del láser para reconstruir la cinética rápida con la resolución apropiada. Para monitorear en un disparo las partes rápidas y lentas de la cinética, es necesario muestrear la señal a una alta velocidad por un largo tiempo, alcanzando la escala de tiempo de 0.1 segundos. Esto requerirá el uso de equipos digitales caros con gran capacidad de memoria de almacenamiento. Actualmente los osciloscopios tienen gran capacidad de memoria de almacenamiento.

La Figura 4 muestra el voltaje de salida del PMT usando un osciloscopio digital y con el dispositivo experimental descrito utilizado en esta tesis.

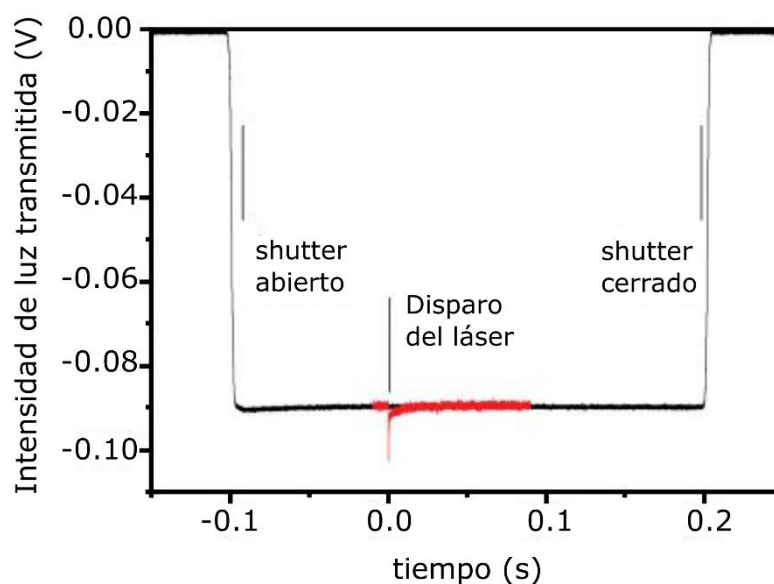


Figura 4: Intensidad de luz transmitida a 436 nm monitoreada por un PMT de 5 etapas.

Notar el signo negativo del voltaje medido, la señal (traza roja en una escala x3) es adquirida en la región plana empezando 100 ms después de la apertura del shutter. La traza roja representará entonces la traza cinética de interés.

II.3.1.2 Determinación de parámetros termodinámicos de activación a partir de medidas cinéticas: modelos de Arrhenius y Eyring

El incremento de la temperatura conduce al incremento de la velocidad de una reacción química, y por ende, de la constante de velocidad de la misma. Svante August Arrhenius fue el primero en señalar empíricamente que la variación de la constante de velocidad con la temperatura se puede representar por una ecuación análoga a la usada para las constantes de equilibrio químico postulada anteriormente por Van't Hoff,^[63] es decir:

$$\frac{\partial \ln k}{\partial T} = \frac{E_a}{RT^2} \quad (II.98)$$

o en su versión integrada:

$$k = Ae^{-\frac{E_a}{RT}} \quad (II.99)$$

donde k representa la constante de velocidad de la reacción, E_a la energía de activación de la misma, A es el factor pre-exponencial de Arrhenius, T la temperatura absoluta y R la constante de los gases.

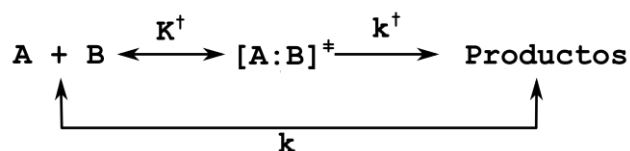
De acuerdo con esta teoría, la cinética de las reacciones no depende de la energía de los reactivos, sino de la energía que los reactivos deben alcanzar para transformarse a productos

(E_a). Es decir, en el camino de mínima energía que conecta los reactivos con productos (coordenada de transformación), existe un máximo llamado “estado de transición” o “complejo activado”, que posee la mínima energía que deben adquirir las moléculas para pasar de reactivos a productos.

A partir de la ecuación (II.99), al gráfico de logaritmo natural de la constante de velocidad en función de la inversa de la temperatura, se lo conoce como gráfico de Arrhenius.

Posteriormente, Henry Eyring desarrolló otra aproximación a este fenómeno basado en la mecánica estadística, conocida como la teoría del estado de transición.^[64] Esta postula que las moléculas para reaccionar deben formar un complejo activado en equilibrio con los reactivos, y este complejo luego se descompone para dar los productos.

Así, una reacción hipotética entre las especies A y B podría esquematizarse como:



El complejo activado se supone que posee ciertas características de una molécula ordinaria, y cierta estabilidad temporal. Con estas ideas básicas se demuestra que la constante de velocidad k de cualquier reacción, está dada por:

$$k = \frac{RT}{Nh} K^\ddagger \quad (\text{II.100})$$

donde N es el número de Avogadro y h la constante de Planck. Teniendo en cuenta la relación termodinámica entre la constante de equilibrio y la energía libre de la reacción, se llega a la siguiente expresión

$$k = \frac{RT}{Nh} e^{-\frac{\Delta G^\ddagger}{RT}} \quad (\text{II.101})$$

y considerando la relación entre $\Delta G^\ddagger$, $\Delta S^\ddagger$ y $\Delta H^\ddagger$ en condiciones de NPT se llega a la siguiente relación

$$k = \frac{RT}{Nh} e^{\frac{\Delta S^\ddagger}{R}} e^{-\frac{\Delta H^\ddagger}{RT}} \quad (\text{II.102})$$

siendo $\Delta G^\ddagger$, $\Delta H^\ddagger$, $\Delta S^\ddagger$ la energía libre de Gibbs, la entalpía y la entropía de activación, respectivamente. En esta ecuación se supuso además que $\Delta H^\ddagger$ y $\Delta S^\ddagger$ son independientes de T .

La relación **(II.102)** es la ecuación fundamental de esta teoría, permitiendo dar interpretaciones de índole termodinámica a procesos de activación.

En definitiva, las últimas dos ecuaciones, permiten el estudio del comportamiento de las constantes cinéticas asociadas a un proceso en función de la temperatura.

Realizando medidas de las constantes de velocidad en función de la temperatura, puede verificarse entonces si el modelo de Eyring se cumple, realizando un gráfico de k vs T y verificando la ecuación **(II.101)**. Si el modelo de Eyring explica el comportamiento entonces se pueden obtener la entropía y entalpía de activación mediante un ajuste de los datos (observando la ecuación **(II.102)**).

Como cálculo control, también se suele estudiar la dependencia de $\Delta G^\ddagger$ (calculado a partir de la ecuación **(II.101)**) en función de la temperatura, si la relación es lineal (es decir cumple que $\Delta G^\ddagger = \Delta H^\ddagger - T \Delta S^\ddagger$) se puede asumir que $\Delta H^\ddagger$ y $\Delta S^\ddagger$ no dependen de la temperatura.

II.4 Referencias

- (1) Gamow, G. *Treinta Años Que Conmovieron La Física: La Historia de La Teoría Cuántica*; Lectores de EUDEBA; Editorial Universitaria de Buenos Aires, **1971**.
- (2) Levine, I. N. *Quantum Chemistry*, 2. ed., 10 print.; The Allyn and Bacon chemistry series; Allyn and Bacon: Boston, **1980**.
- (3) Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; Dover Publications: Mineola, N.Y, **1996**.
- (4) Thomas, L. H. The Calculation of Atomic Fields. *Mathematical Proceedings of the Cambridge Philosophical Society* **1927**, 23 (5), 542.
- (5) Fermi, E. Statistical Method to Determine Some Properties of Atoms. *Rendiconti Accademia Nazionale dei Lincei* **1927**, 6, 602–607.
- (6) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Physical Review* **1964**, 136 (3B), B864–B871.
- (7) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review* **1965**, 140 (4A), A1133–A1138.
- (8) Vosko, S. H.; Wilk, L.; Nusair, M. Accurate Spin-Dependent Electron Liquid Correlation Energies for Local Spin Density Calculations: A Critical Analysis. *Canadian Journal of Physics* **1980**, 58 (8), 1200–1211.
- (9) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Physical Review Letters* **1996**, 77 (18), 3865–3868.
- (10) Perdew, J. P.; Yue, W. Accurate and Simple Density Functional for the Electronic Exchange Energy: Generalized Gradient Approximation. *Physical Review B* **1986**, 33 (12), 8800–8802.
- (11) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Physical Review A* **1988**, 38 (6), 3098–3100.
- (12) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Physical Review B* **1988**, 37 (2), 785–789.
- (13) Perdew, J. P.; Wang, Y. Accurate and Simple Analytic Representation of the Electron-Gas

- Correlation Energy. *Physical Review B* **1992**, 45 (23), 13244–13249.
- (14) Becke, A. D. A New Mixing of Hartree–Fock and Local Density-functional Theories. *The Journal of Chemical Physics* **1993**, 98 (2), 1372–1377.
 - (15) Soler, J. M.; Artacho, E.; Gale, J. D.; García, A.; Junquera, J.; Ordejón, P.; Daniel Sánchez-Portal. The SIESTA Method for Ab Initio Order- N Materials Simulation. *Journal of Physics: Condensed Matter* **2002**, 14 (11), 2745.
 - (16) Sánchez-Portal, D.; Ordejón, P.; Artacho, E.; Soler, J. M. Density-Functional Method for Very Large Systems with LCAO Basis Sets. *International Journal of Quantum Chemistry* **1997**, 65 (5), 453–461.
 - (17) Artacho, E.; Sánchez-Portal, D.; Ordejón, P.; García, A.; Soler, J. M. Linear-Scaling Ab-Initio Calculations for Large and Complex Systems. *physica status solidi (b)* **1999**, 215 (1), 809–817.
 - (18) Reich, S.; Thomsen, C.; Ordejón, P. Electronic Band Structure of Isolated and Bundled Carbon Nanotubes. *Physical Review B* **2002**, 65 (15).
 - (19) Leach, A. R. *Molecular Modelling: Principles and Applications*, 2nd ed.; Prentice Hall: Harlow, England ; New York, 2001.
 - (20) Ponder, J. W.; Case, D. A. Force Fields for Protein Simulations. *Advances in Protein Chemistry*. **2003**, 66, 27–85.
 - (21) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Structure, Function, and Bioinformatics* **2006**, 65 (3), 712–725.
 - (22) Morse, P. M. Diatomic Molecules According to the Wave Mechanics. II. Vibrational Levels. *Physical Review* **1929**, 34 (1), 57–64.
 - (23) Vega, C.; Abascal, J. L. F.; Conde, M. M.; Aragonés, J. L. What Ice Can Teach Us about Water Interactions: A Critical Comparison of the Performance of Different Water Models. *Faraday Discussions* **2009**, 141, 251–276.
 - (24) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *The Journal of Chemical Physics* **1983**, 79 (2), 926–935.
 - (25) Cerutti, D. S.; Freddolino, P. L.; Duke, R. E.; Case, D. A. Simulations of a Protein Crystal with a High Resolution X-Ray Structure: Evaluation of Force Fields and Water Models. *The Journal of Physical Chemistry B* **2010**, 114 (40), 12811–12824.
 - (26) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. Interaction Models for Water in Relation to Protein Hydration. In *Intermolecular Forces*; Pullman, B., Ed.; Springer Netherlands: Dordrecht, **1981**; Vol. 14, pp 331–342.
 - (27) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *The Journal of Physical Chemistry* **1987**, 91 (24), 6269–6271.
 - (28) Crespo, A.; Scherlis, D. A.; Martí, M. A.; Ordejón, P.; Roitberg, A. E.; Estrin, D. A. A DFT-Based QM-MM Approach Designed for the Treatment of Large Molecular Systems: Application to Chorismate Mutase. *The Journal of Physical Chemistry B* **2003**, 107 (49), 13728–13736.
 - (29) Lin, H.; Schöneboom, J. C.; Cohen, S.; Shaik, S.; Thiel, W. QM/MM Study of the Product–Enzyme Complex in P450_{c_{am}} Catalysis. *The Journal of Physical Chemistry B* **2004**, 108 (28), 10083–10088.
 - (30) Schöneboom, J. C.; Thiel, W. The Resting State of P450_{c_{am}} : A QM/MM Study. *The Journal of Physical Chemistry B* **2004**, 108 (22), 7468–7478.
 - (31) Martí, M. A.; Crespo, A.; Bari, S. E.; Doctorovich, F. A.; Estrin, D. A. QM–MM Study of Nitrite Reduction by Nitrite Reductase of *Pseudomonas Aeruginosa*. *The Journal of Physical Chemistry B* **2004**, 108 (46), 18073–18080.
 - (32) Bikiel, D. E.; Boechi, L.; Capece, L.; Crespo, A.; De Biase, P. M.; Di Lella, S.; González Lebrero, M. C.; Martí, M. A.; Nadra, A. D.; Perissinotti, L. L.; Scherlis, D. A.; Estrin, D. A. Modeling Heme Proteins Using Atomistic Simulations. *Physical Chemistry Chemical*

- Physics* **2006**, *8* (48), 5611–5628.
- (33) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angewandte Chemie International Edition* **2009**, *48* (7), 1198–1229.
 - (34) Lin, H.; Truhlar, D. G. QM/MM: What Have We Learned, Where Are We, and Where Do We Go from Here? *Theoretical Chemistry Accounts* **2007**, *117* (2), 185–199.
 - (35) Eichinger, M.; Tavan, P.; Hutter, J.; Parrinello, M. A Hybrid Method for Solutes in Complex Solvents: Density Functional Theory Combined with Empirical Force Fields. *The Journal of Chemical Physics* **1999**, *110* (21), 10452–10467.
 - (36) Rovira, C.; Schulze, B.; Eichinger, M.; Evanseck, J. D.; Parrinello, M. Influence of the Heme Pocket Conformation on the Structure and Vibrations of the Fe-CO Bond in Myoglobin: A QM/MM Density Functional Study. *Biophysical Journal* **2001**, *81* (1), 435–445.
 - (37) Crespo, A.; Martí, M. A.; Estrin, D. A.; Roitberg, A. E. Multiple-Steering QM–MM Calculation of the Free Energy Profile in Chorismate Mutase. *Journal of the American Chemical Society* **2005**, *127* (19), 6940–6941.
 - (38) Verlet, L. Computer “Experiments” on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review* **1967**, *159* (1), 98–103.
 - (39) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. . Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *Journal of Computational Physics* **1977**, *23* (3), 327–341.
 - (40) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *The Journal of Chemical Physics* **1984**, *81* (8), 3684–3690.
 - (41) Di Lella, S.; Martí, M. A.; Álvarez, R. M. S.; Estrin, D. A.; Ricci, J. C. D. Characterization of the Galectin-1 Carbohydrate Recognition Domain in Terms of Solvent Occupancy. *The Journal of Physical Chemistry B* **2007**, *111* (25), 7360–7366.
 - (42) Gauto, D. F.; Di Lella, S.; Guardia, C. M. A.; Estrin, D. A.; Martí, M. A. Carbohydrate-Binding Proteins: Dissecting Ligand Structures through Solvent Environment Occupancy. *The Journal of Physical Chemistry B* **2009**, *113* (25), 8717–8724.
 - (43) López, E. D.; Arcon, J. P.; Gauto, D. F.; Petruk, A. A.; Modenutti, C. P.; Dumas, V. G.; Marti, M. A.; Turjanski, A. G. WATCLUST: A Tool for Improving the Design of Drugs Based on Protein-Water Interactions: Fig. 1. *Bioinformatics* **2015**, *31* (22), 3697–3699.
 - (44) Jarzynski, C. Nonequilibrium Equality for Free Energy Differences. *Physical Review Letters* **1997**, *78* (14), 2690–2693.
 - (45) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, a Package of Computer Programs for Applying Molecular Mechanics, Normal Mode Analysis, Molecular Dynamics and Free Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Computer Physics Communications* **1995**, *91* (1–3), 1–41.
 - (46) Cohen, J.; Arkhipov, A.; Braun, R.; Schulten, K. Imaging the Migration Pathways for O₂, CO, NO, and Xe Inside Myoglobin. *Biophysical Journal* **2006**, *91* (5), 1844–1857.
 - (47) Cohen, J.; Olsen, K. W.; Schulten, K. Finding Gas Migration Pathways in Proteins Using Implicit Ligand Sampling. In *Methods in Enzymology*; Elsevier, 2008; Vol. 437, pp 439–457.
 - (48) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics* **1996**, *14* (1), 33–38.
 - (49) Austin, R. H.; Beeson, K. W.; Eisenstein, L.; Frauenfelder, H.; Gunsalus, I. C. Dynamics of Ligand Binding to Myoglobin. *Biochemistry* **1975**, *14* (24), 5355–5373.
 - (50) Beece, D.; Eisenstein, L.; Frauenfelder, H.; Good, D.; Marden, M. C.; Reinisch, L.; Reynolds, A. H.; Sorensen, L. B.; Yue, K. T. Solvent Viscosity and Protein Dynamics. *Biochemistry* **1980**, *19* (23), 5147–5157.
 - (51) Doster, W.; Beece, D.; Bowne, S. F.; Dilorio, E. E.; Eisenstein, L.; Frauenfelder, H.; Reinisch, L.; Shyamsunder, E.; Winterhalter, K. H.; Yue, K. T. Control and pH Dependence of Ligand Binding to Heme Proteins. *Biochemistry* **1982**, *21* (20), 4831–4839.

- (52) Steinbach, P. J.; Ansari, A.; Berendzen, J.; Braunstein, D.; Chu, K.; Cowen, B. R.; Ehrenstein, D.; Frauenfelder, H.; Johnson, J. B.; Lamb, D. C. Ligand Binding to Heme Proteins: Connection between Dynamics and Function. *Biochemistry* **1991**, *30* (16), 3988–4001.
- (53) Hofrichter, J.; Sommer, J. H.; Henry, E. R.; Eaton, W. A. Nanosecond Absorption Spectroscopy of Hemoglobin: Elementary Processes in Kinetic Cooperativity. *Proceedings of the National Academy of Sciences U.S.A.* **1983**, *80* (8), 2235–2239.
- (54) Murray, L. P.; Hofrichter, J.; Henry, E. R.; Eaton, W. A. Time-Resolved Optical Spectroscopy and Structural Dynamics Following Photodissociation of Carbonmonoxyhemoglobin. *Biophysical Chemistry* **1988**, *29* (1–2), 63–76.
- (55) Esquerra, R. M.; Goldbeck, R. A.; Reaney, S. H.; Batchelder, A. M.; Wen, Y.; Lewis, J. W.; Kliger, D. S. Multiple Geminant Ligand Recombinations in Human Hemoglobin. *Biophysical Journal* **2000**, *78* (6), 3227–3239.
- (56) Dewilde, S.; Kiger, L.; Burmester, T.; Hankeln, T.; Baudin-Creux, V.; Aerts, T.; Marden, M. C.; Caubergs, R.; Moens, L. Biochemical Characterization and Ligand Binding Properties of Neuroglobin, a Novel Member of the Globin Family. *Journal of Biological Chemistry* **2001**, *276* (42), 38949–38955.
- (57) Kriegl, J. M.; Bhattacharyya, A. J.; Nienhaus, K.; Deng, P.; Minkow, O.; Nienhaus, G. U. Ligand Binding and Protein Dynamics in Neuroglobin. *Proceedings of the National Academy of Sciences* **2002**, *99* (12), 7992–7997.
- (58) Bonneau, R.; Wirz, J.; Zuberbühler, A. D. Methods for the Analysis of Transient Absorbance Data (Technical Report). *Pure and Applied Chemistry* **1997**, *69* (5).
- (59) Chen, E.; Goldbeck, R. A.; Kliger, D. S. Nanosecond Time-Resolved Spectroscopy of Biomolecular Processes. *Annual Review of Biophysics and Biomolecular Structure* **1997**, *26* (1), 327–355.
- (60) Tetreau, C.; Lavalette, D. Dominant Features of Protein Reaction Dynamics: Conformational Relaxation and Ligand Migration. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2005**, *1724* (3), 411–424.
- (61) Banderini, A.; Sottini, S.; Viappiani, C. Method for Acquiring Extended Real-Time Kinetic Signals in Nanosecond Laser Flash Photolysis Experiments. *Review of Scientific Instruments* **2004**, *75* (7), 2257–2261.
- (62) Fenster, A.; LeBlanc, J. C.; Taylor, W. B.; Johns, H. E. Linearity and Fatigue in Photomultipliers. *Review of Scientific Instruments* **1973**, *44* (6), 689–694.
- (63) Arrhenius, S. On The Reaction Velocity of the Inversion of Cane Sugar By Acids. In *Selected Readings in Chemical Kinetics*; Elsevier, 1967; pp 31–35.
- (64) Eyring, H. The Activated Complex in Chemical Reactions. *The Journal of Chemical Physics* **1935**, *3* (2), 107–115.

III. ESTUDIO TEÓRICO DE ESTIMADORES DE ENERGÍA LIBRE ASOCIADOS A LA IGUALDAD DE JARZYNSKI

III.1 Energía libre: motivaciones

Como ya se ha mencionado en la introducción, la energía libre es considerada la cantidad más importante en las teorías termodinámicas aplicadas a cualquier proceso físico o químico. Su valor determina la espontaneidad de dicho proceso bajo condiciones de temperatura, presión (e igual a la presión externa) y número de partículas constantes, o bien bajo condiciones de temperatura, volumen y número de partículas constantes (a partir de la energía libre de Gibbs o energía libre de Helmholtz, respectivamente) y cuando no se realizan trabajos distintos del de volumen.

Los experimentos de simulación de sistemas biofísicos son llevados a cabo usualmente bajo dos condiciones de equilibrio: número de partículas, temperatura y volumen constantes (NVT) o número de partículas, temperatura y presión constantes (NPT). Para estas condiciones, como se mencionó la energía libre de Helmholtz (F) y la energía libre de Gibbs (G), respectivamente en cada condición, son las funciones de estado, que en última instancia se desean calcular.

Finalmente, los diversos modelos de la química computacional aplicados durante el estudio del proceso de interés, permitirán interpretar desde un punto de vista químico/físico, los principales factores determinantes de la espontaneidad (o no) del proceso.

Como se mencionó en la introducción de la presente tesis, una correcta evaluación del cambio de energía libre del proceso de interés (o su valor a lo largo de un conjunto de coordenadas que describe el proceso) permitirá aportar información crucial para determinar, por un lado, la espontaneidad de dicho proceso (signo del cambio de energía libre), y por el otro, la velocidad con que dicho proceso sucede (estudio de barreras de activación).

En el presente capítulo se abordará un estudio teórico acerca de varios estimadores de energía libre que se derivan de la ecuación de Jarzynski. Finalmente, se mostrará el impacto que puede tener la elección del estimador, tomando como caso paradigmático la determinación del perfil de energía libre asociada a la migración del anión superóxido en la Manganese superóxido dismutasa (MnSOD) humana. Este trabajo fue realizado en colaboración con la Dra. Mariela Sued, la Dra Daniela Rodriguez y la Dra María Eugenia Zretter Noste, del Instituto de Cálculo (UBA-CONICET) de la Facultad de Ciencias Exactas y Naturales de la Universidad de Buenos Aires

III.2 El “problema del muestreo” para el cálculo de energía libre

Las moléculas o proteínas en solución acuosa presentan, durante su dinámica, muchos grados de libertad (muchas combinaciones de velocidades y posiciones $(\mathbf{p}, \mathbf{r})$ accesibles de todos los átomos constituyentes del sistema), que dan como resultado diferentes mínimos locales de energía (regiones de valores de $(\mathbf{p}, \mathbf{r})$ que evaluados en el Hamiltoniano, dan como resultados mínimos o “pozos” de energía), separados entre sí por barreras de energía.

Los sistemas entonces se moverán más probablemente recorriendo configuraciones cercanas a estos mínimos de energía, y las distintas barreras de energía que separan estas regiones, determinarán la tasa temporal para las cuales el sistema va visitando esas configuraciones.

Por otro lado, la termodinámica estadística muestra que, para un sistema dinámico en equilibrio en condiciones de NVT, la relación entre el valor medio de la energía libre de Helmholtz y los grados de libertad, a través de la función de partición es la siguiente:

$$\langle F \rangle = \frac{1}{\beta} \ln \left(\frac{1}{h^{3N} N!} \iint d\mathbf{p} d\mathbf{r} e^{\beta H(\mathbf{p}, \mathbf{r})} \right) \quad (III.1)$$

Donde $H(\mathbf{p}, \mathbf{r})$ es el hamiltoniano del sistema en función de los momentos y posiciones de las partículas, y determina el valor de la energía total del sistema para cada configuración $(\mathbf{p}, \mathbf{r})$. $\beta = (k_b T)^{-1}$, k_b es la constante de Boltzmann, h es la constante de Planck, N es el número de partículas.

El valor medio de F quedará entonces determinado por la expresión que se encuentra entre paréntesis en la ecuación (III.1), llamada función de partición, y adquiere esa forma funcional para el ensamble NVT.

En particular, el valor de $\langle F \rangle$ quedará determinado por el valor de la integral sobre todos los grados de libertad de todas las partículas intervinientes del sistema a través del integrando $e^{\beta H(\mathbf{p}, \mathbf{r})}$. Este integrando tendrá diferente peso según el valor de $H(\mathbf{p}, \mathbf{r})$: los valores que tendrán un mayor peso en la integral serán entonces los correspondientes valores más altos del integrando, es decir a valores grandes de $H(\mathbf{p}, \mathbf{r})$.

Lo anterior demuestra de manera cualitativa, que la energía libre resulta ser un valor difícil de determinar a partir de una simulación “estándar” de dinámica molecular, ya que la simulación debería muestrear adecuadamente regiones del espacio de $(\mathbf{p}, \mathbf{r})$ que tienen valores altos de $H(\mathbf{p}, \mathbf{r})$, y estos valores son los menos probables de conseguir durante la simulación, debido a que representan configuraciones de alta energía.

Lo anterior constituye lo que se suele denominar en la química computacional como “el problema del muestreo” y obliga a realizar simulaciones mucho más extendidas temporalmente para lograr muestrear varias veces configuraciones de alta energía para poder converger el cálculo de energía libre. En general para los sistemas biofísicos relevantes como proteínas o moléculas en solución acuosa, este proceder resulta impracticable.

Una forma de mejorar el “problema del muestreo” es recurrir a técnicas denominadas de muestreo “sesgado”, en donde uno “obliga” al sistema a evolucionar sobre una coordenada de transformación elegida. En este tipo de esquemas, aun restringiendo el espacio de $(\mathbf{p}, \mathbf{r})$ el problema del muestreo persiste pero en menor medida.

En algunas técnicas, como la de Muestreo sesgado (Umbrella Sampling) o Integración Termodinámica, si bien restringen el espacio de $(\mathbf{p}, \mathbf{r})$ muestreado sobre una coordenada de transformación seleccionada, no realizan un trabajo externo sobre el sistema. Son técnicas de sistemas en equilibrio que redefinen el Hamiltoniano original, agregando un término de restricción del espacio de configuraciones.

En otras técnicas, como la de Perturbación de Energía Libre o la de Dinámica Molecular Guiada (sMD), el sistema es “forzado” a recorrer la coordenada de transformación realizando sobre el sistema un trabajo externo que lleva al sistema fuera del equilibrio. En lo que sigue del capítulo se explicarán los fundamentos teóricos de la técnica de sMD.

III.3 Obtención de energía libre a partir de medidas de trabajo.

La termodinámica clásica establece una relación directa entre el trabajo irreversible realizado sobre el sistema y el cambio en el valor de energía libre asociado (ecuación (III.2), con ΔF o ΔG según el ensamble), entre los estados inicial y final del proceso de interés. Se dice que el sistema es “llevado” del estado inicial al final mediante la aplicación de un trabajo W :

$$W \geq \Delta F \quad (III.2)$$

Donde la igualdad vale *si y solo si* la transformación se realiza de manera reversible.

En una transformación reversible se cumple que:

- La transformación es cuasiestática: el sistema es llevado del estado inicial al final a través de una sucesión de estados de equilibrio intermedios, los cuales pueden ser descritos cada uno completamente a partir de sus variables termodinámicas correspondientes (como P, T, etc).
- No hay efectos disipativos: la transformación desde el estado final hasta el inicial pasa por los mismos estados intermedios que el proceso inverso y no presenta pérdida de energía en el ciclo.

En la mayoría de los sistemas la primera condición se satisface haciendo evolucionar al sistema de manera muy lenta. Esta condición es impracticable desde un punto de vista computacional, ya que se requerirían tiempos de simulación infinitos para lograr que en cada paso intermedio de la transformación el sistema esté equilibrado.

Por otro lado, no todos los caminos que lleven del estado inicial al final serán caminos potencialmente reversibles, por lo tanto surge una nueva complicación en la elección de la coordenada de transformación para el conocimiento del ΔF .

En conclusión, la obtención de ΔF a partir de medidas de trabajos reversibles en sistemas biorelevantes resulta humanamente inviable por métodos computacionales.

Durante el siglo XIX se verificó que la ecuación (III.2) podía ser extendida para el valor medio del conjunto de trabajos realizados sobre el sistema^[1], es decir:

$$\langle W \rangle \geq \Delta F \quad (III.3)$$

Y que la igualdad vale para cuando el proceso es llevado a cabo de manera tal que todos los valores de W son iguales (es decir, sin dispersión para el conjunto de $W^{[2]}$). Se demostró también, que la dispersión de los trabajos decrece cuando la transformación se va realizando más lentamente y alcanza el valor cero en el límite reversible (caso para el cual se recupera la igualdad original, ecuación (III.2)).^[2]

Por otro lado, el teorema de Fluctuación-Disipación^[3,4] establece que para procesos irreversibles pero cercanos al equilibrio vale que:

$$\langle W \rangle - \Delta F = \frac{\beta}{2} \sigma^2 = W_{dis} \quad (III.4)$$

Donde σ^2 es la varianza del conjunto de los trabajos realizados sobre el sistema.

En la ecuación (III.4) se define el trabajo disipado, W_{dis} , como una medida de la energía que se disipa al realizar la transformación de manera irreversible. Dicho de otra manera, el teorema de Fluctuación-Disipación muestra que la energía disipada resulta proporcional a las fluctuaciones del sistema para regímenes cercanos al equilibrio.

Posteriormente, en 1997 Christopher Jarzynski demostró una igualdad muy poderosa que relaciona trabajos realizados, ahora de manera *arbitrariamente* irreversible, con el ΔG entre el estado inicial y final.^[5]

La ecuación, en su primera definición, puede ser escrita como sigue:

$$e^{[-\beta \Delta F(\xi)]} = \lim_{N \rightarrow \infty} \langle e^{[-\beta W_i(\xi, v)]} \rangle_N \quad (III.5)$$

Donde las barras denotan el promedio sobre el conjunto de N trayectorias de trabajo, $W_i(\xi, v)$ denota el trabajo de la i -ésima trayectoria del conjunto sobre la coordenada de transformación ξ . v es la velocidad a la cual se produce el cambio sobre la coordenada ξ del sistema.

La característica más importante de esta ecuación es que (como se verá más adelante) vale independientemente del camino que conecta los estados y la velocidad con que se realiza el cambio. Como corolario, lo anterior implica que no es necesario estar en un régimen cercano al equilibrio durante el proceso de cambio para obtener el ΔF .

III.3.1 ¿Qué significa medir un conjunto de trabajos en un experimento de simulación de Dinámica Molecular?

Sea un sistema mecánico clásico atomístico en contacto con un reservorio a temperatura T . El sistema total está aislado térmicamente y evoluciona de acuerdo con la dinámica Hamiltoniana.

Supongamos que existe (no es un requerimiento explícito) un camino γ que conecta los estados A y B (por ejemplo una coordenada de transformación), y hay un parámetro λ que parametriza ese camino, es decir en $\lambda = \lambda_i$ el sistema se encuentra en el estado A y $\lambda = \lambda_f$ el sistema se encuentra en el estado B. λ puede ser también una variable termodinámica intensiva o extensiva.^[2]

El sistema evoluciona en contacto con el reservorio mientras el valor de λ cambia de λ_i a λ_f sobre un total de tiempo t_s . Se asume (sin pérdida de generalidad) que la velocidad de cambio

$$v = |\lambda_f - \lambda_i| \cdot t_s^{-1} \quad (III.6)$$

es constante.

La evolución del sistema quedará descripta por una trayectoria determinista evolucionando bajo H_λ , mientras λ cambia desde su valor inicial al final sobre el tiempo t_s . Además, el trabajo aplicado durante un tiempo t_s será:

$$W = \int_0^{t_s} dt \, v \frac{\partial H_\lambda(\xi(t))}{\partial \lambda} \quad (III.7)$$

O equivalentemente integrando en el parámetro λ

$$W = \int_{\lambda_0}^{\lambda_f} d\lambda \frac{\partial H_\lambda(\xi(t))}{\partial \lambda} = \int_{\lambda_0}^{\lambda_f} F(\xi(t)) d\lambda \quad (III.8)$$

La ecuación (III.8) permite calcular el trabajo hasta cualquier punto λ como la integral acumulada de la fuerza sobre la coordenada de transformación al cambiar el parámetro λ . Esto permite definir un perfil de trabajo (y en última instancia de energía libre) a lo largo de la coordenada de transformación.

Finalmente, de manera análoga, se puede generar un conjunto de valores de trabajos W_i (con λ y v fijos), partiendo de distintas condiciones iniciales correspondientes a microconfiguraciones obtenidas a partir de algún ensamble termodinámico, fijando el parámetro λ a su valor inicial λ_i , correspondiente al sistema en el estado de partida.

Sin pérdida de generalidad supongamos que para todo el ensamble el camino t_s y v no cambian entre una medida de W y la otra. Bajo estas condiciones, la naturaleza determinista entonces establece que cada una de las trayectorias quedará determinada por su condición inicial, es decir que habrá entonces exactamente una trayectoria que pasa por el punto determinado $H(\mathbf{p}, \mathbf{r})$.

Para una simulación de dinámica molecular típica, esta situación entonces corresponderá computacionalmente a equilibrar al sistema en el estado inicial (es decir con $\lambda = \lambda_i$) y a partir de un conjunto de “fotos” (o microconfiguraciones con distintos valores de $(\mathbf{p}, \mathbf{r})$) decorrelacionadas temporalmente medir los trabajos asociados partiendo de cada condición inicial muestreada, asociado al cambio de la coordenada de transformación elegida (hasta $\lambda = \lambda_f$), usando la ecuación (III.7) o (III.8).

III.4 Demostración de la igualdad de Jarzynski

Para el conjunto total (sistema más reservorio) se puede escribir la siguiente relación para el Hamiltoniano total

$$H_{\lambda}^{SB}(\Gamma, \Theta) = H_{\lambda}^S(\Gamma) + H^B(\Theta) \quad (III.9)$$

Donde Γ y Θ corresponden al punto del espacio de fases (posiciones y momentos) de los grados de libertad del sistema y baño térmico, respectivamente.

Se asume también, que el reservorio es suficientemente grande, entonces no dependerá de λ , (en otras palabras, de lo que le pase al sistema) y además, sus grados de libertad estarán suficientemente desacoplado con los del sistema.

El conjunto total (sistema y reservorio) está térmicamente aislado (entonces, por el primer principio de la termodinámica, $\Delta E = 0$), la distribución de los estados iniciales puede ser representada por un ensamble microcanónico (NVE constantes).

Como comentario adicional, se ha demostrado que, en lo que sigue, resulta equivalente el uso del ensamble canónico o microcanónico.^[2]

Entonces para la función de partición en el ensamble microcanónico se obtiene que:

$$Z_{\lambda}^{SB} = \int d\Gamma d\theta e^{[-\beta H_{\lambda}^{SB}(\Gamma, \theta)]} \quad (III.10)$$

Luego, reemplazando la relación (III.9) se obtiene

$$Z_{\lambda}^{SB} = \int d\Gamma e^{[-\beta H_{\lambda}^S(\Gamma)]} \cdot \int d\theta e^{[-\beta H^B(\theta)]} = Z_{\lambda}^S Z^B \quad (III.11)$$

Además, la probabilidad de encontrar el estado con energía $H_{\lambda_i}^{SB}(\Gamma_0, \theta_0)$ en el ensamble de microconfiguraciones iniciales, se obtiene a partir del factor de Boltzman:

$$P(H_{\lambda_i}^{SB}(\Gamma_0, \theta_0)) = \frac{1}{Z_{\lambda_i}^{SB}} e^{[-\beta H_{\lambda_i}^{SB}(\Gamma_0, \theta_0)]} \quad (III.12)$$

$$\text{con } Z_{\lambda_i}^{SB} = \int d\Gamma_0 e^{[-\beta H_{\lambda_i}^S(\Gamma_0)]} \cdot \int d\theta e^{[-\beta H^B(\theta_0)]} \quad (III.13)$$

Donde Γ_0 y θ_0 corresponden al punto del espacio de fases (posiciones y momentos), en el estado inicial, de los grados de libertad del sistema y baño térmico, respectivamente.

Debido a la conservación de la energía, y como el sistema total está aislado, para cualquier instante t se cumple que:

$$W = H_{\lambda}^{SB}(\Gamma_t, \theta_t) - H_{\lambda_i}^{SB}(\Gamma_0, \theta_0) \quad (III.14)$$

Donde W es el trabajo acumulado para llevar el sistema del estado inicial hasta el estado correspondiente a tiempo t , correspondiente a un valor de parámetro de transformación λ .

Por otro lado, el valor medio de cualquier medida inicial X , puede ser expresado como la integral del producto de la medida a evaluar por el factor de probabilidad en el ensamble usado, es decir:

$$\langle X \rangle = \int P(H_{\lambda_i}^{SB}(\Gamma_0, \theta_0)) \cdot X \cdot d\Gamma_0 d\theta_0 \quad (III.15)$$

Entonces a partir de la ecuación (III.15) y (III.12), el promedio exponencial de los trabajos tiene la siguiente forma:

$$\langle e^{-\beta W} \rangle = \int \frac{1}{Z_{\lambda_i}^{SB}} e^{[-\beta H_{\lambda_i}^{SB}(\Gamma_0, \theta_0)]} \cdot e^{-\beta W} \cdot d\Gamma_0 d\theta_0 \quad (III.16)$$

Luego, insertando la ecuación (III.14) se obtiene que

$$\begin{aligned}
\langle e^{-\beta W} \rangle &= \int \frac{1}{Z_{\lambda_i}^{SB}} e^{[-\beta H_{\lambda_i}^{SB}(\Gamma_0, \theta_0)]} \cdot e^{-\beta [H_{\lambda}^{SB}(\Gamma_t, \theta_t) - H_{\lambda_i}^{SB}(\Gamma_0, \theta_0)]} \cdot d\Gamma_0 d\theta_0 \quad (III.17) \\
&= \int \frac{1}{Z_{\lambda_i}^{SB}} e^{[-\beta H_{\lambda}^{SB}(\Gamma_t, \theta_t)]} \cdot d\Gamma_0 d\theta_0
\end{aligned}$$

Las variables de integración iniciales pueden transformarse a las variables de transformación a tiempo t . De acuerdo con el teorema de Liouville, el Jacobiano de esta transformación es unitario, por lo tanto

$$d\Gamma_0 d\theta_0 = d\Gamma_t d\theta_t \quad (III.18)$$

La evolución temporal a través de la ecuación de Liouville establece una relación importante entre las densidades de espacio de fase a distintos tiempos. En otras palabras, el espacio de fases final puede ser obtenido a partir del espacio inicial haciendo “avanzar” en la evolución temporal del sistema Hamiltoniano. Asimismo, el espacio de fases inicial desde el final haciendo “retroceder” en la evolución temporal. El mapeo es uno a uno.

El teorema de Liouville aplicado al sistema establece que la densidad del espacio de fases se conserva en el tiempo a lo largo de cualquier trayectoria. Insertando la ecuación (III.18) en la ecuación (III.17), se obtiene que

$$\langle e^{-\beta W} \rangle = \int d\Gamma_t d\theta_t \frac{1}{Z_{\lambda_i}^{SB}} e^{[-\beta H_{\lambda}^{SB}(\Gamma_t, \theta_t)]} = \frac{Z_{\lambda}^{SB}}{Z_{\lambda_i}^{SB}} \quad (III.19)$$

Finalmente usando la relación (III.11), y simplificando Z^B se llega a que

$$\langle e^{-\beta W} \rangle = \frac{Z_{\lambda}^S}{Z_{\lambda_i}^S} = e^{(-\beta \Delta F^S)} \quad (III.20)$$

En donde en la igualdad de la derecha se utilizó la relación termodinámica entre la energía libre de Helmholtz para el sistema F^S y la función de partición ($F^S = -(1/\beta) \cdot \ln(Z^S)$)

La ecuación (III.20) representa la igualdad de Jarzynski que relaciona el conjunto de trabajos W y la diferencia de energía libre de Helmholtz ΔF^S para el sistema.

En conclusión la demostración de la igualdad de Jarzynski solo requirió de la conservación de la energía y del teorema de Liouville.

Jarzynski en su primer trabajo, impone como condición durante la realización de los trabajos el acoplamiento entre el baño térmico y el sistema debe ser despreciable, de hecho propone directamente “apagar” el baño térmico durante la realización de los trabajos. En trabajos posteriores, se ha demostrado que la igualdad de Jarzynski sigue valiendo para sistemas con baños térmicos acoplados al sistema. Lo cual generaliza mucho más la aplicación de esta ecuación a muchos sistemas biorelevantes, en donde mantener la temperatura constante resulta crucial.

Jarzynski muestra además que a partir de la relación demostrada (ecuación (III.20)) se demuestran dos relaciones bien conocidas de la termodinámica, las pertenecientes al método de Integración Termodinámica y Perturbación de Energía Libre.^[2,5]

Posteriormente, también se demostró que para una transformación entre dos estados de equilibrio de un sistema en contacto con un reservorio de volumen y calor a temperatura T y presión P , el trabajo W realizado sobre el sistema durante toda la transformación y la diferencia de energía libre ΔG^S entre los dos estados de equilibrio satisfacen una relación análoga a la mostrada anteriormente

$$\langle e^{-\beta W} \rangle = e^{-\beta \Delta G^S} \quad (III.21)$$

Este resultado resulta muy aplicable a los sistemas típicos de simulación computacional donde T y P suelen ser constantes.

Finalmente, las simulaciones computacionales no pueden incluir baños térmicos de tamaño infinito (o mucho más grandes que el sistema de interés) sin embargo es posible simular los efectos de los baños y varios algoritmos fueron desarrollados para esto como se mencionó en la sección de métodos.^[2] La forma más habitual de incorporar los efectos de baños térmicos en los métodos de Dinámica Molecular, es incluir términos adicionales (como fricción o ruido al azar) en las ecuaciones de movimientos de manera tal que las trayectorias resultantes muestreen el ensamble estadístico apropiado. Para la mayoría de estos algoritmos la igualdad de Jarzynski también es operativa.

Como conclusión final, la igualdad de Jarzynski también permite poder calcular un perfil de energía libre asociado a la coordenada de transformación definida. A partir del conjunto de trabajos acumulados para cada valor de λ y aplicando la ecuación (III.17) (o (III.21)) se puede obtener un perfil de energía libre asociado a la coordenada de transformación elegida.

III.5 Aproximación de resorte rígido en las simulaciones de sMD

Como se mencionó en la sección III.2 del presente capítulo, una forma de mejorar el “problema del muestreo” es recurrir a técnicas denominadas de muestreo “sesgado”, en donde se fuerza al sistema a recorrer regiones determinadas de la coordenada de transformación durante la dinámica molecular. Para lograr esto, estas técnicas modifican la descripción matemática del sistema agregando un potencial en el Hamiltoniano (denominado “sesgo”), que agrega una penalidad energética para configuraciones en donde el sistema se encuentra lejos de la coordenada de transformación. El objetivo final de estos métodos consiste en obtener información del sistema sin el sesgo, extrayendo información a partir de la dinámica con el sesgo.

En la presente sección, se explicará una aproximación que es muy utilizada en los cálculos de sMD acoplada a la ecuación de Jarzynski: la aproximación de resorte rígido. Esta aproximación resulta crucial porque establece la relación entre el perfil de energía libre a lo largo de la coordenada de transformación (denominado Potencial de Fuerza Media, PMF) del sistema sesgado con el PMF del sistema sin el sesgo.

Supongamos un sistema mecánico clásico de N partículas, en contacto con un baño de calor a temperatura T . $(\mathbf{r}, \mathbf{p})$, son las variables de posición y momento de cada partícula, respectivamente. Supongamos que se ha identificado una coordenada de transformación $\xi(r)$.

El PMF (Potential of Mean Force, PMF), $\Phi(\xi)$, a lo largo de ξ se define como:

$$e^{[-\beta\Phi(\xi')]} = \int d\mathbf{r}d\mathbf{p} \cdot \delta(\xi(r) - \xi') \cdot e^{[-\beta H(\mathbf{r}, \mathbf{p})]} \quad (III.22)$$

Con esta definición, el PMF es la energía libre de Helmholtz a lo largo de $\xi(r)$, el gráfico de energía libre versus la coordenada de transformación $\xi(r)$, se lo llama perfil de energía libre asociado a la coordenada de reacción $\xi(r)$

Para el ensamble NPT se tiene que

$$e^{[-\beta\Phi(\xi')]} = \int dV \int d\mathbf{p} \int d\mathbf{r} \cdot \delta(\xi(r) - \xi') \cdot e^{[-\beta H(\mathbf{r}, \mathbf{p}, V) - \beta PV]} \quad (III.23)$$

Para este ensamble entonces, el PMF es la energía libre de Gibbs a lo largo de $\xi(r)$.

En lo que sigue, la demostración se basará en la ecuación (III.22), pero para el ensamble NPT el resultado final es equivalente.

Una simulación de tipo SMD, como técnica de muestreo sesgado, es una manera eficiente de explorar el sistema a lo largo de la coordenada de transformación. En estas técnicas se agrega un potencial de guía de forma cuadrática (ecuación (III.24)) al Hamiltoniano original H .

$$h_\lambda = \frac{k}{2} (\xi(r) - \lambda)^2 \quad (III.24)$$

Donde k es la constante del potencial armónico utilizado, y λ es el parámetro de acoplamiento entre los dos Hamiltonianos (como se mencionó en las secciones anteriores).

Al Hamiltoniano original $H(r, p)$ se agrega el potencial antes descrito de manera de tener una nueva descripción para el sistema “modificado” (a partir de $\tilde{H}_\lambda(r, p)$)

$$\tilde{H}_\lambda(r, p) = H(r, p) + h_\lambda(r) \quad (III.25)$$

Como se mencionó, λ es un parámetro que típicamente se cambia a velocidad fija v a lo largo del tiempo

$$\lambda(t) = \lambda(0) + v \cdot t \quad (III.26)$$

Para cubrir la región relevante de $\xi(r)$

En términos generales, para un instante t , el valor del parámetro $\lambda(t)$ estará fijo al valor establecido por la ecuación (III.26). Por su parte, la coordenada de reacción $\xi(r)$ puede tomar cualquier valor, aunque el sesgo lo mantenga cerca de $\lambda(t)$. La fuerza con que lo mantiene cerca estará dada por la constante k .

El objetivo entonces es obtener el PMF, asociado a la coordenada $\xi(r)$, bajo el Hamiltoniano original (H) a partir de calcular la energía libre asociada al sistema con el sesgo $\tilde{H}_\lambda(r, p)$, usando el trabajo acumulado W . La aproximación de resorte-rígido mostrará que la relación entre ambos perfiles es directa. La idea general de esta aproximación, es minimizar la fluctuación de la coordenada de transformación a lo largo de diferentes trayectorias eligiendo una constante k suficientemente grande para el potencial de guía.

Retomando, a partir de la definición (III.22), el PMF asociado al sistema modificado $\tilde{H}_\lambda(r, p)$ en función del parámetro λ se puede escribir de la siguiente manera:

$$\begin{aligned}
e^{[-\beta F_\lambda]} &= \int dr dp \int d\xi' \cdot \delta(\xi(r) - \xi') \cdot e^{\{-\beta H(r,p) - \frac{\beta k}{2} [\xi(r) - \lambda]^2\}} \\
&= \int d\xi \cdot e^{\{-\beta \Phi(\xi) - \frac{\beta k}{2} [\xi - \lambda]^2\}}
\end{aligned} \tag{III.27}$$

Donde en la segunda igualdad se obtiene a partir de la definición del PMF del sistema sin el sesgo.

Cuando k es grande la mayor contribución a la integral vendrá dada por el segundo término en la suma dentro del integrando exponencial. Para este caso, la función resultante es una Gaussiana con valores máximos en la región $\xi \sim \lambda$ (ver Figura 1)

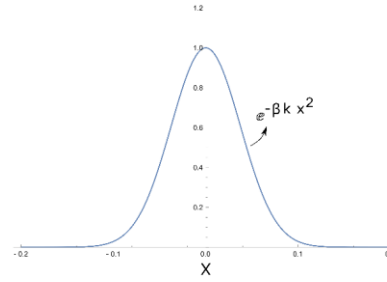


Figura 1: forma de la función Gaussiana con valor $\lambda = 0$ (centro de la campana)

Entonces, se puede realizar una serie de Taylor de la función $e^{-\beta \Phi(\xi)}$ alrededor de $\xi = \lambda$.

$$\begin{aligned}
e^{[-\beta \Phi(\xi)]} &= e^{[-\beta \Phi(\lambda)]} \left\{ 1 - \beta \frac{\partial \Phi(\lambda)}{\partial \lambda} (\xi - \lambda) \right. \\
&\quad \left. - \frac{\beta}{2} \left[\frac{\partial^2 \Phi(\lambda)}{\partial \lambda^2} - \beta \left(\frac{\partial \Phi(\lambda)}{\partial \lambda} \right)^2 \right] (\xi - \lambda)^2 + \dots \right\}
\end{aligned} \tag{III.28}$$

Y luego reemplazando la ecuación (III.28) en la segunda igualdad de la ecuación (III.27), distribuyendo los términos dentro de la exponencial e integrando cada termino, se obtiene finalmente que

$$e^{[-\beta F_\lambda]} = e^{[-\beta \Phi(\lambda)]} \sqrt{\frac{2\pi}{\beta k}} \left\{ 1 - \frac{1}{2k} \left[\frac{\partial^2 \Phi(\lambda)}{\partial \lambda^2} - \beta \left(\frac{\partial \Phi(\lambda)}{\partial \lambda} \right)^2 \right] + O(1/k^2) \right\} \tag{III.29}$$

Luego tomando logaritmo (y aproximando $\ln(a + (1 - bx)) = \log(a) - bx$ y despreciando los términos que son independientes de λ , se obtiene que

$$F_\lambda = \Phi(\lambda) - \frac{1}{2k} \left(\frac{\partial \Phi(\lambda)}{\partial \lambda} \right)^2 + \frac{1}{2\beta k} \frac{\partial^2 \Phi(\lambda)}{\partial \lambda^2} + O(1/k^2) \tag{III.30}$$

que invertida da lugar a:

$$\Phi(\lambda) = F_\lambda + \frac{1}{2k} \left(\frac{\partial F_\lambda}{\partial \lambda} \right)^2 - \frac{1}{2\beta k} \frac{\partial^2 F_\lambda}{\partial \lambda^2} + O\left(1/k^2\right) \quad (III.31)$$

En donde si k es suficientemente grande, a primer orden se obtiene

$$\Phi(\lambda) = F_\lambda \quad (III.32)$$

O equivalentemente (como $\xi \sim \lambda$, cuando k es suficientemente grande)

$$\Phi(\xi) = F_\xi \quad (III.33)$$

Con lo cual el PMF calculado en el procedimiento de sMD acoplado a la ecuación de Jarzynski se aproxima al PMF real del sistema sin el sesgo.

En aplicaciones prácticas uno elige la constante k suficientemente grande como para que la fluctuación en la energía del potencial guía (ecuación (III.24)) a lo largo de diferentes trayectorias sea menor que la energía térmica ($k_B T$).

Bajo esta condición la coordenada de transformación será parecida al centro del potencial cuadrático, por lo tanto satisfaciendo la condición de resorte rígido.

Como se mencionó, la forma de proceder implica generar un número de trayectorias sMD partiendo de distintas condiciones iniciales, y el trabajo es calculado como función del tiempo final t (o de λ) para cada trayectoria.

Luego F_λ es calculado a partir de la ecuación de Jarzynski. Finalmente, el PMF es obtenido de la última ecuación hasta un cierto orden en $1/k$, el siguiente orden puede ser usado para chequear la validez de la aproximación de resorte rígido, evaluando el peso de los términos superiores de potencias de k .

III.6 El problema de la estimación de la energía libre a través de la ecuación de Jarzynski

En este punto, es importante remarcar que en sus primeros trabajos Jarzynski demostró su igualdad de forma similar a la mostrada en este capítulo, y esta igualdad establece la relación entre la energía libre y el *promedio del ensamble* de valores de W , sin hacer mención al problema matemático/estadístico asociado a la *estimación* del valor de energía libre a partir de un conjunto finito de valores de W . Recién en publicaciones posteriores y a partir del creciente uso de su ecuación, empezó a surgir el problema matemático que se desprendía para la

estimación del valor de energía libre^[6-10], y específicamente acerca de la dificultad asociada al uso de un número inevitablemente finito de trayectorias empleadas.^[6]

El resultado general obtenido por Jarzynski (ecuación **(III.21)**) plantea que es posible obtener el ΔF entre dos estados de equilibrio, a partir del promedio exponencialmente pesado del ensamble de trabajos irreversibles (en su forma más general) para llevar el sistema del estado inicial al final, partiendo de microconfiguraciones muestreadas a partir del ensamble completo del estado inicial.

En la práctica, lo anterior es computacionalmente inabordable, y en general se parten de un conjunto finito de microconfiguraciones. El problema ahora resulta de índole más estadística que fisicoquímica, en el cual se busca obtener un valor confiable como estimador del ΔG de interés.

Una manera muy habitual de estimar el ΔF , consiste en la forma tradicional de sumar y dividir por la cantidad de datos (ecuación **(III.34)**) la forma de esta ecuación surge de la propia ecuación de Jarzynski.

$$\widehat{\Delta F}_J = -\beta^{-1} \ln \left[\frac{1}{N} \sum_{i=1}^N e^{-\beta W_i} \right] \quad (\text{III.34})$$

Donde $\widehat{\Delta F}_J$ se llamará de ahora en adelante, el estimador de Jarzynski, para la energía libre. Una forma equivalente de escribir la ecuación **(III.34)** es la siguiente:

$$\overline{e^{-\beta W}} = \sum_{i=1}^N P(W_i) \cdot e^{-\beta W_i} \quad (\text{III.35})$$

Donde $\overline{e^{-\beta W}}$ es el promedio exponencial de la variable W . $P(W_i)$, es la distribución de probabilidad asociada a la variable W y la suma se realiza sobre los N valores W_i . Esta forma de escribir la estimación, permite explicar con más claridad el problema de la estimación.

Supongamos entonces que se realiza un experimento de simulación sMD en donde se miden un conjunto de valores de trabajo W_i , para un valor dado de coordenada de transformación.

A partir de los datos de W_i medidos se puede construir un histograma (o curva de densidad empírica) para obtener $P(W_i)$, en este punto y solo como ejemplo ilustrativo, supongamos que $P(W_i)$ puede aproximarse como una distribución Normal, por ejemplo la curva sólida azul de la Figura 2.

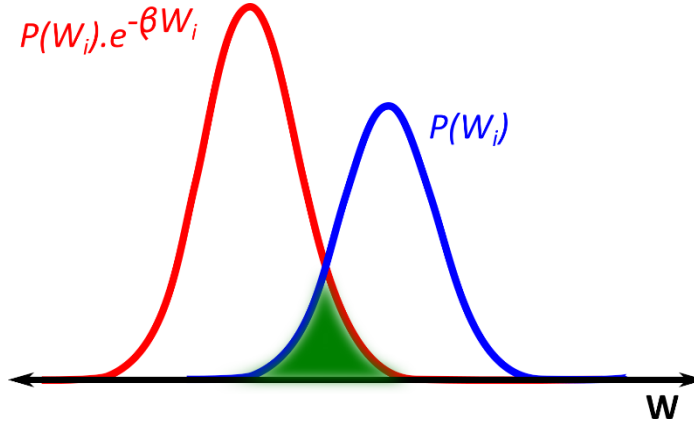


Figura 2: Problema del estimador de Jarzynskii. Curva azul, curva de $P(W_i)$ para los trabajos medidos. Curva roja sumando de la ecuación (III.35). La zona de los valores más pesados en la ecuación (III.35) caen en las zonas menos probables de $P(W_i)$ (zona sombreada en verde)

Los valores que tendrán más peso en la suma de la ecuación (III.35), serán los valores en la zona cercana al centro de la curva de $P(W_i) \cdot e^{-\beta W_i}$ (línea roja discontinua en la Figura 2). Esta curva pierde la normalización y está desplazada de la curva $P(W_i)$, hacia valores menores de W . Si $P(W_i)$ es una curva Gaussiana con una dispersión σ , entonces $P(W_i) \cdot e^{-\beta W_i}$ será otra curva Gaussiana (pero no normalizada) con el mismo ancho pero con el valor medio corrido a la izquierda en (β, σ^2) .

En el caso general, para dispersiones razonables en los experimentos típicos de SMD, los valores que tienen más peso en la ecuación (III.35), “caen” en regiones de baja probabilidad de $P(W_i)$, zona sombreada con verde en la Figura 2. Si el corrimiento es mucho más grande que el ancho, entonces habrá poco solapamiento entre las curvas, lo que hace que la estimación del promedio exponencial sea impracticable. En general, como se verá más adelante, la estimación es práctica cuando σ no es mucho más grande que $k_B \cdot T$.

El análisis anterior, muestra de manera cualitativa el problema principal del estimador de Jarzynski, $\widehat{\Delta F}_J$. Esta forma de estimación, estará dominada por valores de trabajos que son poco probables de obtener en la dinámica guiada, usualmente denominadas trayectorias “raras” [7,10]. Este problema, implica que se deben realizar numerosas trayectorias guiadas para obtener valores pocos probables para mejorar la convergencia del promedio a través del estimador de Jarzynski. Se dice entonces que para un muestreo finito de valores de W , $\widehat{\Delta F}_J$ da un valor sesgado para la estimación del valor real de ΔF .

Debido a la dificultad de estimar el promedio exponencial, otro estimador del ΔF que suele usarse que se define a partir de la expansión en cumulants de la ecuación (III.21). Esta expansión se obtiene a partir de los desarrollos en serie de Taylor del logaritmo y de la exponencial, y adquiere la siguiente forma:

$$\Delta F = \sum_{n=1}^{\infty} \frac{\kappa_n (-\beta)^{-1}}{n!} \quad (III.36)$$

$$\kappa_1 = \langle W \rangle, \kappa_2 = \langle W^2 \rangle - \langle W \rangle^2 = \sigma_w^2, \kappa_3 = \langle W^3 \rangle - \langle W \rangle^3 - 3\sigma_w^2 \langle W \rangle$$

Con $\kappa_1, \kappa_2, \kappa_3$ los primeros tres términos de la suma para ΔF . σ_w^2 y $\langle W \rangle$ son la varianza y el valor medio del ensamble de W , respectivamente.

De manera análoga a lo planteado para $\widehat{\Delta F}_j$, es posible plantear un estimador asociado a la expansión en cumulants (ecuación (III.36)) en donde el valor medio y la varianza de W son expresados en sus formas estimadas asociadas. A partir de la ecuación (III.36), se obtiene entonces el siguiente estimador asociado

$$\widehat{\Delta F}_c = \sum_{n=1}^{\infty} \frac{\widehat{\kappa}_n (-\beta)^{-1}}{n!} \quad (III.37)$$

$$\hat{\kappa}_1 = \bar{W}, \hat{\kappa}_2 = \overline{W^2} - \bar{W}^2 = \widehat{\sigma}_w^2, \hat{\kappa}_3 = \overline{W^3} - \bar{W}^3 - 3\widehat{\sigma}_w^2 \bar{W}$$

Como puede observarse ahora la estimación a ordenes bajos va a depender de promedios asociados a potencias de W ($\bar{W}, \overline{W^2}$, etc).

La Figura 3 muestra la ventaja de utilizar $\widehat{\Delta F}_c$ para la estimación de la energía libre usando los primeros dos cumulants de la expansión.

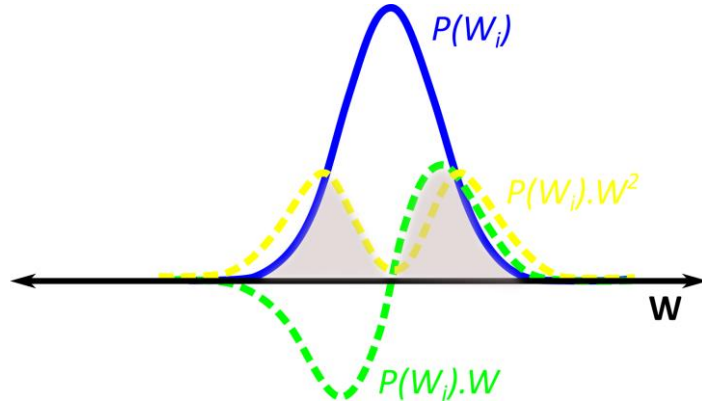


Figura 3: Curva azul, curva de $P(W_i)$ para los trabajos medidos. Curva verde sumando de la ecuación (III.35) para $\bar{W}$. Curva verde sumando de la ecuación (III.35) para $\overline{W^2}$. En ambos casos, la zona de valores más pesados, caen en las zonas más probables de $P(W_i)$ (zona sombreada en gris)

Aplicando un razonamiento análogo a la estimación por la ecuación (III.35) para estimar el valor medio de W y W^2 , puede observarse que las zonas que necesitan ser muestreadas (zona sombreada en gris en la Figura 3), corresponde a la zona cercana al máximo de $P(W_i)$. Evitando la necesidad de muestrear apreciablemente trayectorias “raras” para los valores de W .

Esta forma de estimación tiene ahora dos fuentes de error: un error debido al truncamiento de la serie asociada a $\widehat{\Delta F}_c$ (ecuación (III.37)) y un error debido al muestreo insuficiente de datos (especialmente para órdenes superiores).

En el caso extremo, en donde se evalúe la serie completa, no existirá un error debido al truncamiento de la serie, sin embargo se recuperará el sesgo asociado al muestreo finito mediante el estimador de Jarzynski, explicado anteriormente. Por otro lado, como se mencionó, los términos de la expansión en cumulants (a órdenes bajos) serán más fáciles de estimar.

En secciones siguientes, se mostrará que para algunos casos particulares, la expansión truncada a segundo orden, es la forma óptima para la estimación que la expansión completa.^[7,11]

A partir de la ecuación de Jarzynski y suponiendo una distribución Gaussiana para los valores de W , surge el tercer estimador considerado para la energía libre, llamado estimador de Fluctuación-Disipación (ecuación (III.38)) por su relación con el teorema de Fluctuación-Disipación, explicado en la sección III.3:

$$\widehat{\Delta F}_{FD} = \bar{W} - \frac{\beta \hat{\sigma}_w^2}{2} \quad (III.38)$$

Donde $\bar{W}$ y $\hat{\sigma}_w^2$ son el promedio y el desvío estándar asociada a una muestra finita de valores de W .

Como puede observarse, la ecuación (III.38), también coincide con la expansión a segundo orden de la serie de cumulants (ecuación (III.37) con $n = 2$), mostrando que para este caso en particular todos los cumulants se anulan exceptuando los primeros dos términos y no hay penalidad por truncar la expansión.

Como se mencionó en la sección III.3, el teorema de Fluctuación-Disipación resulta válido para procesos cercanos al equilibrio, en otras palabras para procesos irreversibles llevados a cabo de forma lenta. En estos casos la distribución de trabajos esperada será Gaussiana como sugiere la relación entre el teorema de Fluctuación-Disipación y la expansión en cumulants para distribución de trabajos Gaussianas. Para aplicaciones de sMD en general, la distribución de trabajos podría no ser Gaussiana. En estos casos el comportamiento entre $\widehat{\Delta F}_{FD}$ y $\widehat{\Delta F}_J$ podría ser muy diferente.

III.7 Objetivos y lineamientos

Una de las más notables validaciones de la igualdad de Jarzynski, publicada en 2002 en Science por el grupo del Prof. Bustamante, ha generado gran interés en la comunidad científica, con más de 900 citas hasta el día, y ha sido incorporada como material de enseñanza en muchos programas educativos alrededor del mundo^[1]. En ese artículo, los autores validaron la igualdad de Jarzynski con experimentos muy elegantes de molécula simple, y además concluyeron que $\widehat{\Delta F}_J$ es mejor estimador que $\widehat{\Delta F}_{FD}$. Este descubrimiento importante, apoyado también por otros autores^[12,13], provocó un gran aumento de las aplicaciones de la igualdad de Jarzynski, usando $\widehat{\Delta F}_J$.^[5,14–17]

En lo que sigue del capítulo, se demostrará que el estimador $\widehat{\Delta F}_{FD}$ es el más eficiente de todos los estimadores que pueden proponerse para distribuciones Gaussianas de trabajo, a través de, como se verá, la demostración que este estimador surge del proceso estimación de máxima-verosimilitud para distribuciones de trabajo Gaussianas, y como tal, es por construcción, la estimación de energía libre más eficiente para tales distribuciones.

Luego se extenderá este análisis a dos distribuciones no Gaussianas relevantes, Gamma y skew-normal, y finalmente se ilustrará el impacto de la elección del estimador en un experimento computacional de molécula simple, la migración de un ligando en una metaloproteína, específicamente la superóxido dismutasa de Mn humana (MnSOD) cuyo sustrato es el anión superóxido.

Como última instancia, las siguientes secciones, permiten aclarar que la igualdad de Jarzynski no necesita estar exclusivamente vinculada al estimador $\widehat{\Delta F}_J$, y que, aunque es el más frecuentemente invocado, es quizás uno de los estimadores de la energía libre menos eficiente que se definen a partir la igualdad. La conclusión final, a la que se llegará, es sobre la ventaja de aprovechar el estimador de máxima verosimilitud para una distribución de trabajo particular, y por lo tanto que el conocimiento de la distribución conduce al uso óptimo de la igualdad de Jarzynski.

III.8 Resultados

III.8.1 Estimación de máxima-verosimilitud para distribuciones Gaussianas

Es importante enfatizar la diferencia entre los estimadores de Jarzynski y Fluctuación-Disipación. El estimador de Jarzynski (ecuación (III.34)) es sólo equivalente a la igualdad de Jarzynski en el límite de $N \rightarrow \infty$, este estimador se basa, como se mencionó, en aproximar de la siguiente manera el promedio del ensamble:

$$\langle e^{-\beta W} \rangle \approx \frac{1}{N} \sum_{i=1}^N e^{-\beta W_i} \quad (III.39)$$

Ahora bien, si se conoce la familia estadística a la que pertenece la distribución subyacente de los valores de trabajo, o incluso puede ser anticipada, el promedio del ensamble requerido en la igualdad de Jarzynski puede calcularse usando los estimadores que surgen del procedimiento de máxima-verosimilitud para la distribución particular.

El procedimiento de máxima-verosimilitud lo que provee es la forma de calcular estimadores de cada parámetro que caracteriza a la distribución.

Por ejemplo, para una distribución de trabajo Gaussiana, el promedio del conjunto puede expresarse *exactamente* en función de la media poblacional de la distribución del trabajo μ y la varianza poblacional σ^2 :

$$\langle e^{-\beta W} \rangle = e^{(-\beta\mu + \frac{\beta^2\sigma^2}{2})} \quad (III.40)$$

Luego para este caso, en lugar de calcular directamente el promedio exponencial pesado de una muestra finita como en el estimador de Jarzynski, el procedimiento de máxima-verosimilitud establece en este caso que los estimadores de la media (μ) y la varianza (σ^2) tendrán la forma funcional de la media $\bar{W}$ y la varianza $\hat{\sigma}^2$ de la muestra. Entonces, a partir de la ecuación (III.40) se obtiene que

$$\langle e^{-\beta W} \rangle = e^{(-\beta\bar{W} + \frac{\beta^2\hat{\sigma}^2}{2})} \quad (III.41)$$

Para el caso general en que los trabajos tengan otras distribuciones, la forma de los estimadores asociados a cada parámetro que caracterizan la distribución será diferente.

Al sustituir la ecuación (III.41) en la igualdad de Jarzynski , se obtiene el estimador de máxima-verosimilitud de la energía libre para las distribuciones de trabajo Gaussianas, que coincide con el estimador $\widehat{\Delta F}_{FD}$, resultante del teorema de no-equilibrio de Fluctuación-Disipación (FD)^[3] (ecuación (III.38)).

En el límite de N grande, los dos estimadores convergen con el valor verdadero de ΔF . El estimador $\widehat{\Delta F}_J$ es más robusto, en el sentido de que no requiere conocimiento de la distribución de trabajo subyacente. Sin embargo, si dicha información está disponible, el estimador de máxima verosimilitud ($\widehat{\Delta F}_{FD}$ para una distribución Gaussiana) es por construcción el estimador estadísticamente más eficiente posible.

III.8.2 Simulaciones numéricas de los estimadores de Jarzynski y de máxima verosimilitud

En muchas aplicaciones de la igualdad de Jarzynski, las distribuciones de trabajo son reportadas como aproximadamente Gaussianas graficando una curva Gaussiana usando el valor de la media y del desvío estándar de los datos (como se verá en el capítulo III de la presente tesis). De hecho, en ciertas aplicaciones, como en la técnica sMD, se puede predecir la obtención de una distribución de trabajo Gaussiana si se trabaja en condiciones donde la aproximación de resorte rígido (sección III.5) es válida (como se verá en la discusión, sección III.9).

En el presente capítulo se considerarán distribuciones de trabajo Gaussianas generadas numéricamente usando diferentes dispersiones (varianzas) para los datos, así como también dos distribuciones de trabajo no-Gaussianas: skew-normal y Gamma (Figura 4), que se han sido reportados en otros procesos de no equilibrio^[18,19].

Se puede realizar un procedimiento análogo a lo mostrado en la sección anterior para distribuciones Gaussianas y utilizando la ecuación de Jarzynski para las otras distribuciones consideradas. La resolución de esta ecuación entonces proveerá expresiones *exactas* de los ΔF en términos de los parámetros que caracterizan a cada familia de distribuciones. Los parámetros y las expresiones exactas de ΔF para las familias de distribuciones Gaussianas, skew-normales y Gamma, se resumen en la Tabla 1.

El procedimiento de máxima-verosimilitud permite estimar cada uno de los parámetros para una muestra finita, que finalmente permite obtener la expresión del estimador de máxima-verosimilitud para cada familia, usando las expresiones de la segunda columna de la Tabla 1.

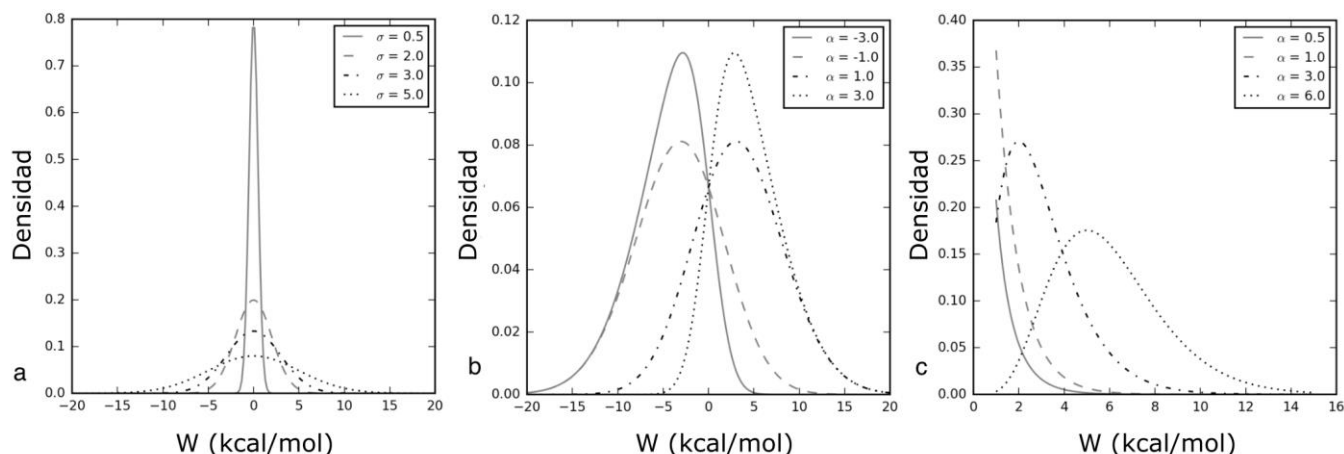


Figura 4: Las funciones de densidad de probabilidad consideradas. (A) Gaussiana ($\mu = 0$), (B) skew-normal ($\mu = 0$ y $\xi = 6$), y (C) las distribuciones Gamma ($\lambda = 1$). Los parámetros que se variaron en cada una de las distribuciones se indican en legendas de diagrama.

Tabla 1: resumen de las familias estadísticas y las expresiones para el ΔF

Distribución	Función de densidad de probabilidad	ΔF
Normal $N(\mu, \sigma^2)$	$\frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$	$\mu - \frac{\beta\sigma^2}{2}$
Skew-Normal $SN(\mu, \xi, \alpha)^{1,2}$	$\frac{1}{\xi\pi} e^{-\frac{(x-\mu)^2}{2\xi^2}} \int_{-\infty}^{\alpha\frac{(x-\mu)}{\xi}} e^{-\frac{t^2}{2}} dt$	$\mu - \frac{\beta\xi^2}{2} - \frac{\log(2\Psi(-\delta\xi\beta))}{\beta}$
Gamma $\Gamma(\alpha, \lambda)^3$	$\frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \quad x > 0$	$\frac{\alpha}{\beta} \ln\left(\frac{\lambda + \beta}{\lambda}\right)$

¹ Ψ es la función de distribución acumulada de $N(0,1)$ y $\delta = \frac{\alpha}{\sqrt{1+\alpha^2}}$

² $\mu = \xi + \omega\delta\sqrt{\frac{2}{\pi}}$

³ $\Gamma(\alpha)$ es la función gama evaluada en el valor alfa

Para la evaluación de los estimadores, se procedió de la siguiente manera. Dada una familia de distribuciones, un conjunto de parámetros que la caracterizan, y un tamaño de muestra N:

- 1) Se generó una muestra de trabajos $W_1, W_2, \dots, W_N$, de manera aleatoria.
- 2) La energía libre es estimada para la muestra de trabajos, de acuerdo con las expresiones de los estimadores.
- 3) Los pasos 1 y 2 se repitieron un total de $M=1000$ veces, de manera de obtener una distribución de M valores estimados para cada estimador, a partir de estos estimadores se construyó la curva estimada para la densidad de cada estimador.

Todas las simulaciones fueron realizadas utilizando el paquete estadístico R^[20]

De cada distribución se generaron 10000 muestras aleatorias de tamaño creciente N , que van desde ($N = 50$ a $N = 500000$), que corresponden medidas de valores de trabajo independientes. De este modo, se generaron las distribuciones de los estimadores y se evaluó el comportamiento de los $\widehat{\Delta F}_J$ y el estimador de máxima-verosimilitud respectivo, con tamaños de muestra cada vez mayores.

III.8.3 Distribuciones de trabajo Gaussianas

Para las distribuciones de trabajo Gaussianas más dispersas ($\sigma > 2 \text{ kcal/mol}$), se observa que $\widehat{\Delta F}_J$ (serie roja, Figura 5) no sólo es menos eficiente que $\widehat{\Delta F}_{FD}$ (serie azul, Figura 5), sino también que está sesgada respecto al valor esperado de ΔF (flecha vertical, Figura 5).

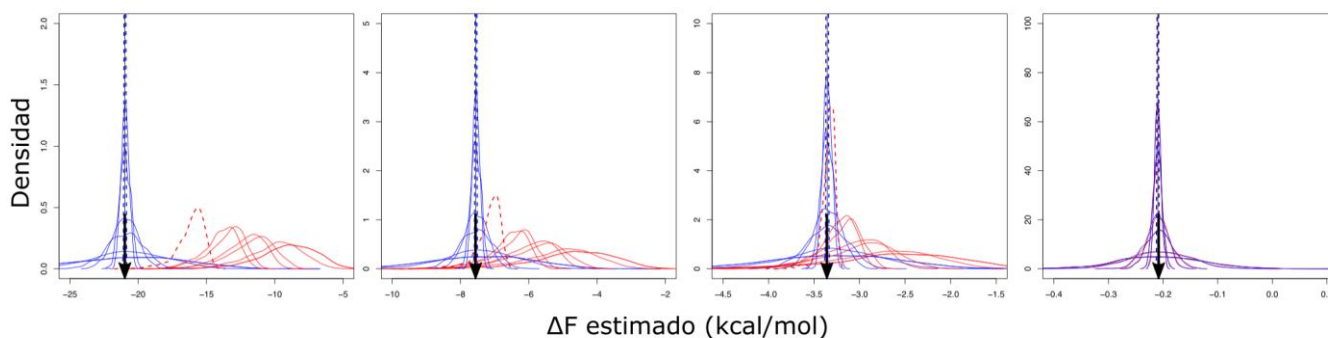


Figura 5: comparación de $\widehat{\Delta F}_J$ (serie roja) y $\widehat{\Delta F}_{FD}$ (serie azul) para distribuciones Gaussianas con $\mu = 0$ y $\sigma = [5, 3, 3, 0.5]$ (de izquierda a derecha). Ambos estimadores fueron evaluados con tamaños de muestra creciente ($N=[50, 100, 500, 1000, 5000, 10000, 500000]$). El valor verdadero de ΔF está indicado como una flecha vertical en cada gráfico.

Aunque este sesgo disminuye con el aumento del tamaño de la muestra, se observa que la distribución de las energías libres obtenidas con $\widehat{\Delta F}_J$ sigue siendo parcial. De hecho, para la distribución con $\sigma = 5 \text{ kcal/mol}$, la distribución del estimador todavía no incluía el valor verdadero de ΔF incluso con un tamaño de muestra de $N = 500000$ (Figura 5, línea roja discontinua).

Notablemente, la dispersión más "grande" considerada aquí ($\sigma = 5 \text{ kcal/mol}$) es realmente bastante común en aplicaciones de la igualdad de Jarzynski, particularmente para aplicaciones que involucran sistemas biomoleculares con muchos grados de libertad.^[21–23]

Para las distribuciones Gaussianas con dispersiones decrecientes, se llegó a la conclusión de que el estimador de Jarzynski finalmente alcanza el valor verdadero de ΔF , aunque con menor rapidez, requiriendo de un muestreo más extenso que $\widehat{\Delta F}_{FD}$. Es interesante notar que esto es cierto incluso para el régimen de N pequeño ($N = 30$), donde vemos que, a pesar de la

distribución de las energías libres obtenidas con el estimador de máxima verosimilitud $\widehat{\Delta F}_{FD}$ es amplio, todavía se centra exactamente en el verdadero valor de ΔF . Sólo para la más estrecha distribución de trabajo Gaussiana considerada aquí ($\sigma = 0.5 \text{ kcal/mol}$) se observó un comportamiento equivalente para ambos estimadores.

III.8.4 Distribuciones de trabajo no-Gaussianas

Consideramos dos distribuciones de trabajo no Gaussianas que también a menudo se presentan en aplicaciones de la igualdad de Jarzynski, las distribuciones skew-normal y Gamma.

Las funciones de densidad de probabilidad consideradas aquí se muestran en la Figura 4, paneles B y C. El valor esperado de ΔF , según la ecuación de Jarzynski, para cada tipo de distribución de trabajo se proporciona en la Tabla 1. Los valores de los estimadores de máxima-verosimilitud para los parámetros que caracterizan cada distribución (datos no mostrados), son calculados por subrutinas propias del programa R.

El $\widehat{\Delta F}_J$ y las correspondientes distribuciones de los estimadores de máxima-verosimilitud para cada una de las distribuciones se muestran en la Figura 6. Como referencia, aquí también comparamos con los valores obtenidos para la expansión en cumulants de segundo orden de la igualdad de Jarzynski, $\widehat{\Delta F}_{FD}$, para las muestras finitas de estas distribuciones no Gaussianas (Figura 6, serie verde).

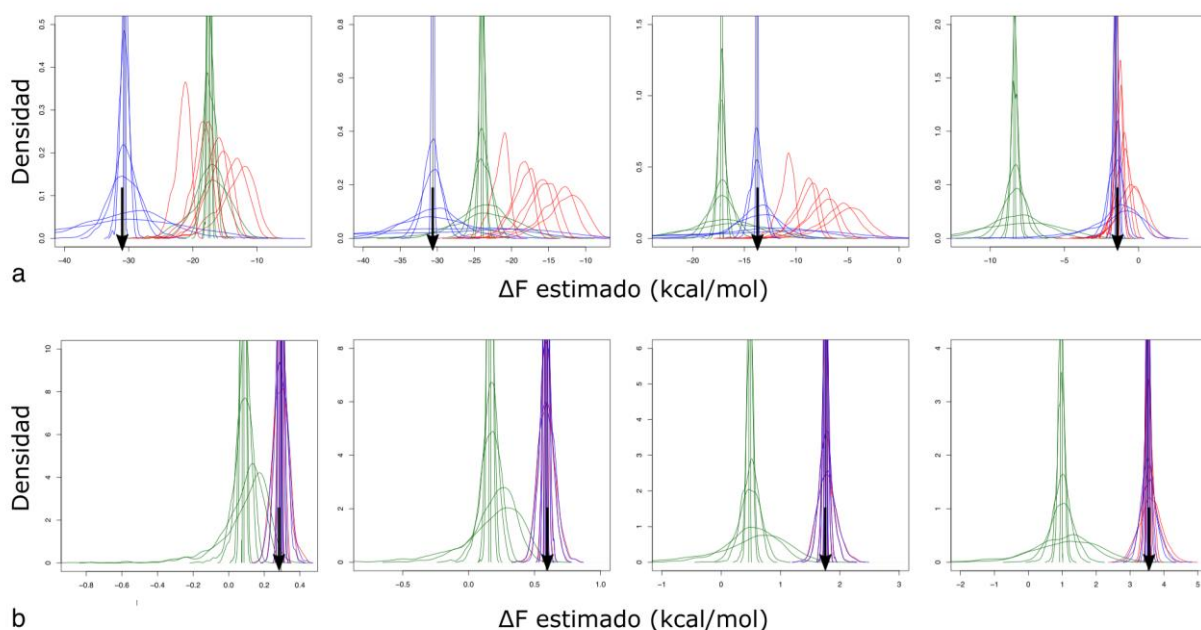


Figura 6: comparación de los estimadores $\widehat{\Delta F}_J$ (serie roja) y máxima-verosimilitud (serie azul) para (A) Distribuciones skew-normal con $\mu = 0$, $\xi = 6$, y $\alpha = [-3, -1, 1, 3]$ (de izquierda a derecha). (B) Distribuciones Gamma con $\lambda = 1$ y $\alpha = [-0.5, 1, 3, 6]$ (de izquierda a derecha). El verdadero valor de ΔF está indicado por la flecha vertical en cada

gráfico. Para cada caso, los estimadores fueron evaluados usando un número creciente de tamaño de muestra (N=[50, 100, 500, 1000, 5000, 10000, 500000])

Para la distribución skew-normal, no hay forma explícita para la estimación de máxima verosimilitud, por lo tanto el parámetro α se calcula numéricamente a partir de la muestra, haciendo la estimación cuestionable para N pequeño. Sin embargo, incluso para tamaños de muestra pequeños (N = 30) de las distribuciones skew-normal, la estimación de máxima verosimilitud fue más precisa que $\widehat{\Delta F}_J$ (Figura 6A). Sorprendentemente, dependiendo de los parámetros de asimetría específicos, $\widehat{\Delta F}_J$ podría incluso ser menos preciso que su expansión en cumulants de segundo orden.

Para la distribución de Gamma (Figura 6B), el estimador de Jarzynski es casi indistinguible de su estimador de máxima verosimilitud correspondiente, y la expansión en cumulants a segundo orden, $\widehat{\Delta F}_{FD}$, generalmente es impreciso, como se esperaba.

III.8.5 Caso ejemplar: migración de superóxido en MnSOD humana.

Como se mencionó en la introducción de la presente tesis, una de las muchas aplicaciones de la igualdad de Jarzynski es el estudio de la migración de ligandos pequeños a proteínas.^[17,21,22,24] Para ilustrar el impacto de la elección del estimador, se considerará una de estas aplicaciones: la migración de superóxido en la Manganese Superóxido Dismutasa (MnSOD) mitocondrial humana.

Esta proteína resulta crítica en mantener la concentración de radicales libre baja en la célula ya que actúa como un eliminador eficaz del anión superóxido (O_2^-). La alta eficiencia enzimática de MnSOD puede explicarse en parte porque la asociación bimolecular entre superóxido y MnSOD es muy rápida. Sin embargo, en un estudio computacional de este proceso en 2010 realizado por nuestro grupo^[25], mostró que el perfil de energía libre asociado al proceso de migración presentaba una barrera sustancial de aproximadamente 6 kcal/mol al acercarse al centro de Mn.

En esta parte del presente capítulo, se revisará este perfil de energía libre, repitiendo los cálculos previamente publicados, con simulaciones de dinámica molecular más extensas, realizadas con una velocidad más lenta y un mayor número de trayectorias de SMD.

Para este fin, se utilizó el campo de fuerzas de Amber 99SBildn^[26] para modelar la proteína (Protein Data Bank, .PDB entrada 1ABM52). El sistema fue solvatado (caja octaédrica de moléculas de agua TIP3P) y los parámetros estructurales y de carga relevantes de los residuos del sitio activo coordinados a los centros Mn y de la molécula superóxido fueron obtenidos a partir de cálculos PBE/6-31G** (fiteo de la SEP y metodología RESP, respectivamente) usando un sistema modelo considerando los aminoácidos que estaban directamente coordinados al centro metálico.

Después de estas varias etapas de equilibración, las simulaciones MD de producción se llevaron a cabo en el ensamble NPT, para un total de 30 ns. Los parámetros más relevantes fueron un time step de 2 fs, algoritmo SHAKE para restringir los hidrógenos^[27], y un valor de corte 12 Å para las interacciones de no-enlace. Las interacciones electrostáticas de largo alcance se modelaron utilizando PME.

Para la utilización de la técnica de sMD se realizaron simulaciones MD de equilibrado y de producción del sistema con la molécula de superóxido restringida a 9 Å del centro de Mn con una constante de fuerza de $k = 400 \text{ kcal/mol.Å}^2$ de acuerdo con el potencial cuadrático $V_\lambda = \frac{k}{2} (\xi(r) - \lambda)^2$.

La misma distancia entre el átomo de oxígeno y el centro de Mn que se restringió durante el equilibrio y la producción MD se utilizó como coordenada de transformación para las simulaciones de sMD, en las que se guió al ligando de 9 a 4 Å usando una velocidad v de 0,0025 Å/ps.

Se simularon 148 trayectorias sMD y para su posterior análisis de los estimadores.

En la Figura 7A se muestra el cálculo del perfil de energía libre, de acuerdo a ambos estimadores $\widehat{\Delta F}_J$ y $\widehat{\Delta F}_{FD}$, resaltando una gran diferencia (aproximadamente 10 kcal/mol) en las dos estimaciones, particularmente cerca del final de las trayectorias sMD. También se muestra, en la Figura 7B, un histograma normalizado del trabajo acumulado al final de las trayectorias sMD que, según la prueba de Kolmogorov-Smirnov^[28], presenta evidencia de una distribución Gaussiana subyacente.

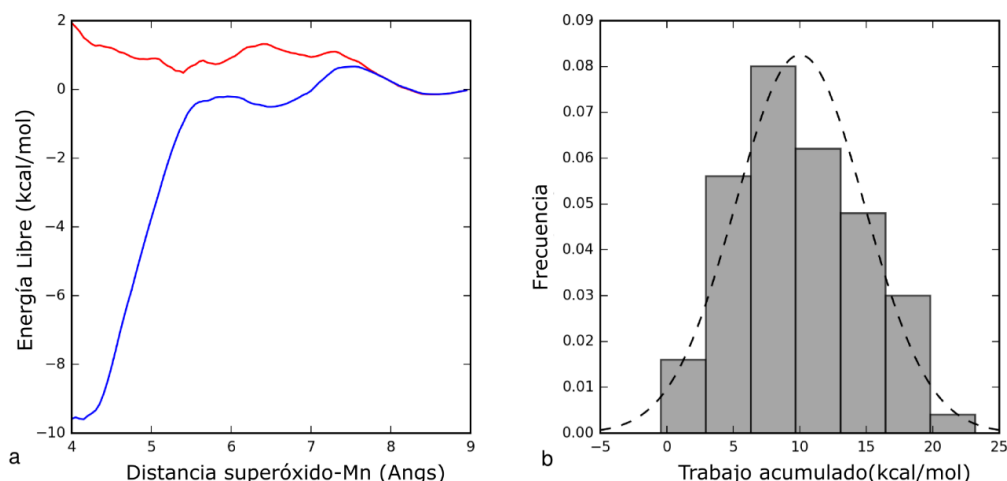


Figura 7: Ilustración de la gran discrepancia entre $\widehat{\Delta F}_J$ y el estimador de máxima verosimilitud correspondiente, $\widehat{\Delta F}_{FD}$. (A) Perfil de energía libre para la migración de superóxido en MnSOD, estimada usando el $\widehat{\Delta F}_J$ (rojo) o $\widehat{\Delta F}_{FD}$ (azul). Las mismas $N = 148$ trayectorias sMD fueron analizadas para los dos casos. (B) Histograma normalizado del trabajo acumulado al final de las trayectorias sMD (coordenada de reacción de 4 Å)

Junto a la evidencia de las simulaciones numéricas presentadas en este capítulo, vemos que esta aplicación se encuentra en el régimen de σ grande para el cual $\widehat{\Delta F}_J$ está sesgada y sobrestima el ΔF .

En el caso presentado, la elección del estimador tiene un impacto enorme en la interpretación química del proceso. El resultado obtenido con el estimador de máxima verosimilitud, $\widehat{\Delta F}_{FD}$, sugiere una migración muy favorable de 9 a 4 Å del centro catalítico, mientras que el obtenido con $\widehat{\Delta F}_J$ sugiere que es casi igual de probable que se mueva hacia o alejándose del centro catalítico.

Los resultados experimentales reportados para este sistema, muestran que la unión de superóxido a MnSOD tenía un valor casi difusional de $k_{on} \sim 10^9 M^{-1} s^{-1}$, implicando que el proceso sucede sin una barrera apreciable para la migración y unión del anión superóxido.

El perfil de energía libre reportado en 2010 por nuestro grupo, era inconsistente con el resultado experimental, mientras que el tratamiento presentado en el presente capítulo permitió obtener resultados más consistentes con los experimentos reportados, mostrando que la elección del estimador es crítica para la obtención de resultados adecuados.

III.9 Discusión

Se ha evaluado la igualdad de Jarzynski ampliamente celebrada y discutida, haciendo especial hincapié en los estimadores necesarios para calcular el promedio que plantea dicha

ecuación. El estimador más convencional, llamado estimador de Jarzynski, es el más usado habitualmente y resulta ser el más robusto estimador ya que no requiere de una caracterización de la distribución de trabajo subyacente, es decir, de una identificación de la familia estadística a la que pertenece esa distribución. Esta robustez sin embargo viene acompañada con una disminución de la eficiencia estadística. Nosotros encontramos que el valor promedio del ensamble o el valor esperado, resulta mejor estimado teniendo en cuenta la distribución subyacente de los trabajos. Esto impulsa una nueva perspectiva sobre los diferentes estudios de la igualdad de Jarzynski que, tal vez injustamente, se restringen al uso del estimador convencional.^[7,23,29]

El estimador de Jarzynski es notorio por su convergencia problemática,^[8,31–35] y se han propuesto expresiones para estimar su sesgo para ambos contextos Gaussianos y no Gaussianos,^[6,36] entre otros enfoques para remediar el sesgo debido a tamaños de muestra finitos.^[37] Algunos autores han utilizado distribuciones de trabajo bidireccionales sumado a argumentos de máxima-verosimilitud para proponer estimadores óptimos,^[34,38–40] pero hasta nuestro conocimiento este es el primer uso de argumentos de máxima verosimilitud para proponer el estimador óptimo para distribuciones de trabajo unidireccionales.

Las simulaciones numéricas mostraron que para una distribución de trabajo Gaussiana con $\sigma = 5 \text{ kcal/mol}$, ni siquiera un tamaño de muestra tan grande como $N=500.000$ fue lo suficientemente grande como para que el estimador de Jarzynski converja al valor esperado. Este resultado es crítico, ya que cambia nuestra comprensión de cuán grande debe ser una muestra con el fin de esperar un comportamiento razonablemente asintótico del estimador. Lo anterior, es coherente con los resultados publicados recientemente por Jarzynski que muestran que el necesario N para que $\widehat{\Delta F}_J$ converja aumenta exponencialmente con la varianza de la distribución de trabajo.^[10,31]

Hemos extendido este análisis a las distribuciones Gamma y skew-normal, también encontrando que el estimador de la energía libre de máxima verosimilitud correspondiente es óptimo para ambas distribuciones. Sin embargo, el estimador Jarzynski convencional es casi indistinguible del estimador de máxima verosimilitud asociado a la distribución Gamma.

Por otra parte, la distribución skew-normal puede ser aún más problemática que una distribución Gaussiana para la convergencia de $\widehat{\Delta F}_J$, ya que el sesgo depende de la combinación de los parámetros de la distribución. Sorprendentemente, aunque el estimador de máxima verosimilitud para la distribución skew-normal no tiene forma explícita, y sus parámetros deben primero ser estimados numéricamente, es aún más preciso que $\widehat{\Delta F}_J$.

La falla clave en la eficiencia del estimador de Jarzynski reside en la presencia de valores de trabajo negativos improbables en la distribución de trabajo subyacente. Esta es precisamente la razón por la que $\widehat{\Delta F}_J$ funciona considerablemente bien para las distribuciones de trabajo de la familia Gamma, en la que los valores de trabajo son estrictamente no-negativos. Hacemos hincapié en que, incluso en estos casos, el estimador de máxima-verosimilitud, por construcción es más eficiente que $\widehat{\Delta F}_J$.

En este contexto, volviendo a las primeras validaciones experimentales de la igualdad de Jarzynski,^[1,12,13] hay varias interpretaciones posibles de la aparente exactitud del estimador de Jarzynski. La primera interpretación es que la dispersión de las distribuciones de trabajo es suficientemente estrecha de manera que los dos estimadores simplemente convergen entre sí y la diferencia entre los números reportados es insignificante. Una segunda interpretación es que la distribución del trabajo puede de hecho no ser Gaussiana, lo que requeriría un análisis más profundo de los datos crudos, pero en este caso la expansión en cumulants a segundo orden $\widehat{\Delta F}_{FD}$ podría ser de hecho un estimador inferior, como se observa por ejemplo en la Figura 3B. En cualquier caso, nuestros resultados confirman que no es posible para $\widehat{\Delta F}_J$, ser más eficiente que el estimador de máxima-verosimilitud correspondiente de la distribución, si de hecho la familia estadística de la distribución está correctamente identificada.

De hecho, para el caso específico de las distribuciones Gaussianas, la expansión de segundo orden de la igualdad de Jarzynski, deja de ser una aproximación y es exacta, ya que todos los términos de orden superior, se espera que sean *exactamente* cero. Si no se tiene en cuenta esta propiedad, la relación entre, $\widehat{\Delta F}_J$ y $\widehat{\Delta F}_{FD}$ (en tanto el segundo coincide con la expansión de segundo orden del primero), parece dar a luz la noción de que el truncamiento de la serie implica cierta pérdida en la precisión^[1]. Esta noción resulta falaz, ya que los términos de orden superior se esperan que sean *exactamente* cero, como se mencionó. Por lo tanto la estimación de los términos superiores sólo empeora la estimación de la energía libre y resulta en un procedimiento incorrecto en la estimación. Este problema aún es mas grave, ya que estos términos son lentos de converger; y además, en última instancia la estimación de todos los términos de orden superior resultarían en el propio estimador de Jarzynski, con su sesgo inherente.

El caso anterior, resalta la ventaja que adquiere la estimación (a través de $\widehat{\Delta F}_{FD}$) utilizando toda la información disponible que le sistema provee. Si la distribución de trabajos subyacente se espera ser Gaussiana, entonces es inconveniente la evaluación de los cumulants superiores ya que por construcción se espera que sean *exactamente* cero, tener en cuenta esta

información en el modelo permite que la evaluación a segundo orden, $\widehat{\Delta F}_{FD}$, sea una estimación no sesgada.

Dos posibles preocupaciones con respecto al uso del estimador de máxima-verosimilitud en lugar del estimador de Jarzynski convencional implican (i) la confianza en la designación de la familia estadística de la distribución del trabajo y (ii) la confianza en las estimaciones de los parámetros necesarios a partir de una muestra finita de valores de trabajo. Con respecto a esta última preocupación, es fundamental enfatizar que los principios de máxima-verosimilitud aseguran que los estimadores de energía libre, son estadísticamente más eficientes que el estimador de Jarzynski. Esto no requiere necesariamente que la estimación de la varianza o del promedio de la muestra esté convergido.

Independientemente de las estimaciones de los parámetros en sí, la expresión exacta de ΔF en función de estos estimadores da como resultado el estimador más eficiente.

En función de lo anterior, encontramos como el “quid” de la cuestión la primera preocupación, con respecto a la designación apropiada de la familia estadística para la distribución de trabajos subyacentes. Dependiendo de la forma en que se conduzca el proceso de no equilibrio, la familia estadística de la distribución de trabajo subyacente puede anticiparse de antemano. Esta anticipación teórica, más que la evidencia directa a partir de la muestra finita de valores de trabajo (en efecto, pequeñas muestras extraídas de verdaderas distribuciones Gaussianas pueden fallar erróneamente las pruebas estadísticas de normalidad), garantiza el uso del estimador de máxima-verosimilitud correspondiente. En ausencia de tal información, uno debe basarse solamente en la evidencia de la propia muestra.

En este sentido, las aplicaciones computacionales de molécula simple pueden ser potencialmente más provechosas que sus análogos experimentos espectroscópicos de fuerza, ya que la evolución de la distribución del trabajo puede reducirse a la distribución asociada a una partícula que experimenta un movimiento Browniano bajo un potencial armónico móvil, que se ha demostrado tener una distribución Gaussiana. Park y Schulten^[7] extendieron rigurosamente este argumento a las trayectorias de sMD, demostrando que con una constante de fuerza rígida para el potencial y suficientes “relajaciones rápidas del sistema” con respecto a la coordenada de transformación cambiante, se espera también que las distribuciones de trabajo obtenidas a partir de las trayectorias de sMD sean Gaussianas. Bajo estas suposiciones, las ecuaciones de movimiento que gobiernan la coordenada de transformación quedarán bien representadas bajo un modelo de movimiento Browniano de una partícula.

La suposición de resorte rígido explicado en este capítulo, tiene entonces dos ventajas: en primer lugar, provee una forma de extraer el PMF como función de una coordenada de transformación (a partir de la energía libre en función de un parámetro externo λ) y en el presente contexto, garantiza la distribución Gaussiana de trabajos, mejorando la estimación de la energía libre. También es instructivo recordar que la naturaleza Gaussiana se mantiene en el tiempo; es decir, las condiciones de arrastre resultantes en una distribución de trabajo Gaussiana seguirán dando lugar a una distribución Gaussiana independientemente de cuánto tiempo arrastre el sistema.^[2]

Todos los resultados presentados, ponen en relieve la ventaja de prever teóricamente la naturaleza de las distribuciones de trabajo subyacentes asociadas a los experimentos de sMD, que dependen del sistema de interés y de la forma en que es llevada a cabo la técnica.

III.10 Referencias

- (1) Liphardt, J. Equilibrium Information from Nonequilibrium Measurements in an Experimental Test of Jarzynski's Equality. *Science* **2002**, 296 (5574), 1832–1835.
- (2) Park, S.; Schulten, K. Calculating Potentials of Mean Force from Steered Molecular Dynamics Simulations. *The Journal of Chemical Physics* **2004**, 120 (13), 5946–5961.
- (3) Callen, H. B.; Welton, T. A. Irreversibility and Generalized Noise. *Physical Review* **1951**, 83 (1), 34–40.
- (4) Chandler, D. *Introduction to Modern Statistical Mechanics*; Oxford University Press: New York, 1987.
- (5) Jarzynski, C. Nonequilibrium Equality for Free Energy Differences. *Physical Review Letters* **1997**, 78 (14), 2690–2693.
- (6) Gore, J.; Ritort, F.; Bustamante, C. Bias and Error in Estimates of Equilibrium Free-Energy Differences from Nonequilibrium Measurements. *Proceedings of the National Academy of Sciences U.S.A.* **2003**, 100 (22), 12564–12569.
- (7) Park, S.; Khalili-Araghi, F.; Tajkhorshid, E.; Schulten, K. Free Energy Calculation from Steered Molecular Dynamics Simulations Using Jarzynski's Equality. *The Journal of Chemical Physics* **2003**, 119 (6), 3559–3566.
- (8) Pohorille, A.; Jarzynski, C.; Chipot, C. Good Practices in Free-Energy Calculations. *The Journal of Physical Chemistry B* **2010**, 114 (32), 10235–10253.
- (9) Jarzynski, C. Equilibrium Free-Energy Differences from Nonequilibrium Measurements: A Master-Equation Approach. *Physical Review E* **1997**, 56 (5), 5018–5035.
- (10) Jarzynski, C. Rare Events and the Convergence of Exponentially Averaged Work Values. *Physical Review E* **2006**, 73 (4).
- (11) Jensen, M. O.; Park, S.; Tajkhorshid, E.; Schulten, K. Energetics of Glycerol Conduction through Aquaglyceroporin GlpF. *Proceedings of the National Academy of Sciences* **2002**, 99 (10), 6731–6736.
- (12) Douarche, F.; Ciliberto, S.; Petrosyan, A.; Rabbiosi, I. An Experimental Test of the Jarzynski Equality in a Mechanical Experiment. *Europhysics Letters (EPL)* **2005**, 70 (5), 593–599.
- (13) Harris, N. C.; Song, Y.; Kiang, C.-H. Experimental Free Energy Surface Reconstruction from Single-Molecule Force Spectroscopy Using Jarzynski's Equality. *Physical Review Letters* **2007**, 99 (6).
- (14) Bizzarri, A. R.; Cannistraro, S. Free Energy Evaluation of the p53-Mdm2 Complex from Unbinding Work Measured by Dynamic Force Spectroscopy. *Physical Chemistry Chemical Physics* **2011**, 13 (7), 2738–2743.
- (15) Orzechowski, M.; Cieplak, P. Application of Steered Molecular Dynamics (SMD) to Study DNA–drug Complexes and Probing Helical Propensity of Amino Acids. *Journal of Physics: Condensed Matter* **2005**, 17 (18), S1627–S1640.
- (16) Oberhofer, H.; Dellago, C.; Boresch, S. Single Molecule Pulling with Large Time Steps. *Physical Review E* **2007**, 75 (6).
- (17) Forti, F.; Boechi, L.; Estrin, D. A.; Marti, M. A. Comparing and Combining Implicit Ligand Sampling with Multiple Steered Molecular Dynamics to Study Ligand Migration Processes in Heme Proteins. *Journal of Computational Chemistry* **2011**, 32 (10), 2219–2231.
- (18) Crooks, G. E.; Jarzynski, C. Work Distribution for the Adiabatic Compression of a Dilute and Interacting Classical Gas. *Physical Review E* **2007**, 75 (2).
- (19) Nategholeslam, M.; Gray, C. G.; Tomberli, B. Stiff Spring Approximation Revisited: Inertial Effects in Nonequilibrium Trajectories. *The Journal of Physical Chemistry B* **2017**, 121 (2), 391–403.
- (20) RStudio Team, RStudio: Integrated Development Environment for R. RStudio, Inc.: Boston, MA, 2015.

- (21) Boubeta, F. M.; Bari, S. E.; Estrin, D. A.; Boechi, L. Access and Binding of H_2S to Hemeproteins: The Case of Hbl of *Lucina Pectinata*. *The Journal of Physical Chemistry B* **2016**, *120* (36), 9642–9653.
- (22) Zhang, D.; Gullingsrud, J.; McCammon, J. A. Potentials of Mean Force for Acetylcholine Unbinding from the Alpha7 Nicotinic Acetylcholine Receptor Ligand-Binding Domain. *Journal of the American Chemical Society* **2006**, *128* (9), 3019–3026.
- (23) Hayashi, K.; Ueno, H.; Iino, R.; Noji, H. Fluctuation Theorem Applied to F₁-ATPase. *Physical Review Letters* **2010**, *104* (21).
- (24) Boechi, L.; Martí, M. A.; Milani, M.; Bolognesi, M.; Luque, F. J.; Estrin, D. A. Structural Determinants of Ligand Migration in Mycobacterium Tuberculosis Truncated Hemoglobin O. *Proteins: Structure, Function, and Bioinformatics* **2008**, *73* (2), 372–379.
- (25) Moreno, D. M.; Martí, M. A.; De Biase, P. M.; Estrin, D. A.; Demicheli, V.; Radi, R.; Boechi, L. Exploring the Molecular Basis of Human Manganese Superoxide Dismutase Inactivation Mediated by Tyrosine 34 Nitration. *Archives of Biochemistry and Biophysics* **2011**, *507* (2), 304–309.
- (26) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber ff99SB Protein Force Field. *Proteins: Structure, Function, and Bioinformatics* **2010**, NA-NA.
- (27) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. . Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *Journal of Computational Physics* **1977**, *23* (3), 327–341.
- (28) Daniel, W. W. *Applied Nonparametric Statistics*, 2. ed.; The Duxbury advanced series in statistics and decision sciences; PWS-KENT: Boston, **1990**.
- (29) Baştuğ, T.; Chen, P.-C.; Patra, S. M.; Kuyucak, S. Potential of Mean Force Calculations of Ligand Binding to Ion Channels from Jarzynski's Equality and Umbrella Sampling. *The Journal of Chemical Physics* **2008**, *128* (15), 155104.
- (30) Minh, D. D. L.; McCammon, J. A. Springs and Speeds in Free Energy Reconstruction from Irreversible Single-Molecule Pulling Experiments [†]. *The Journal of Physical Chemistry B* **2008**, *112* (19), 5892–5897.
- (31) Yunger Halpern, N.; Jarzynski, C. Number of Trials Required to Estimate a Free-Energy Difference, Using Fluctuation Relations. *Physical Review E* **2016**, *93* (5).
- (32) Hummer, G.; Szabo, A. Free Energy Reconstruction from Nonequilibrium Single-Molecule Pulling Experiments. *Proceedings of the National Academy of Sciences* **2001**, *98* (7), 3658–3661.
- (33) Minh, D. D. L.; Chodera, J. D. Optimal Estimators and Asymptotic Variances for Nonequilibrium Path-Ensemble Averages. *The Journal of Chemical Physics* **2009**, *131* (13), 134110.
- (34) Shirts, M. R.; Bair, E.; Hooker, G.; Pande, V. S. Equilibrium Free Energies from Nonequilibrium Measurements Using Maximum-Likelihood Methods. *Physical Review Letters* **2003**, *91* (14).
- (35) *Free Energy Calculations*; Chipot, C., Pohorille, A., Eds.; Castleman, A. W., Toennies, J. P., Yamanouchi, K., Zinth, W., Series Eds.; Springer Series in Chemical Physics; Springer Berlin Heidelberg: Berlin, Heidelberg, 2007; Vol. 86.
- (36) Palassini, M.; Ritort, F. Improving Free-Energy Estimates from Unidirectional Work Measurements: Theory and Experiment. *Physical Review Letters* **2011**, *107* (6).
- (37) Zuckerman, D. M.; Woolf, T. B. Theory of a Systematic Computational Error in Free Energy Differences. *Physical Review Letters* **2002**, *89* (18).
- (38) Crooks, G. E. Path-Ensemble Averages in Systems Driven far from Equilibrium. *Physical Review E* **2000**, *61* (3), 2361–2366.
- (39) Minh, D. D. L.; Adib, A. B. Optimized Free Energies from Bidirectional Single-Molecule Force Spectroscopy. *Physical Review Letters* **2008**, *100* (18).
- (40) Maragakis, P.; Spichty, M.; Karplus, M. Optimal Estimates of Free Energies from Multistate

Nonequilibrium Work Data. *Physical Review Letters* **2006**, 96 (10).

IV. REACTIVIDAD QUÍMICA DE ACIDO SULFHÍDRICO (H₂S) EN HEMOPROTEÍNAS.

IV.1 Introducción

Como se mencionó en la introducción de la presente tesis, se han encontrado especies de sulfuro de hidrógeno endógenas en mamíferos^[1-8] y plantas.^[9-12] Además, debido a que se han descrito algunas funciones biológicas de estas especies, se ha incluido al H₂S dentro de la familia de gasotransmisores.

Uno de los targets moleculares del H₂S son las hemoproteínas.^[13-15,23,24] Por ejemplo, la reacción reversible de H₂S con la citocromo C oxidasa, disminuye la actividad de la enzima reducida^[16,17], induciendo un estado de tipo hibernación en ratones.^[18,19]

Para mioglobina, además de la unión, se destaca la formación de los llamados compuestos de tipo sulfhemos (que se producen debido a la adición de sulfuro sobre el macrociclo del hemo). Este proceso ha sido asignado tanto a un proceso nocivo^[20], como a un proceso de detoxificación.^[21]

En hemoglobina se ha reportado que la unión de sulfuro al complejo férrico es el paso inicial de una ruta oxidativa-catabólica para especies de sulfuro reducido.^[22]

La afinidad en hemoproteínas ha sido descrita, en general, en términos de la estabilización del complejo hexacoordinado. Sin embargo, un trabajo realizado en conjunto con la Dra. Sara Bari utilizando el sistema microperoxidasa-11 (MP11, que no permite mecanismos de estabilización distal, en solución acuosa pudo demostrar que la histidina proximal estabiliza el complejo Fe(III)-(sulfuro)^[25], aun siendo el sulfuro un reductor muy fuerte.

Existen dos características que distinguen a las especies de sulfuro de otros gasotransmisores (por ejemplo CO, NO y O₂). La primera es la unión preferida por las especies de sulfuro a grupos hemo en estado férrico. La segunda es la posibilidad de establecer sus equilibrios ácido-base en solución acuosa con las especies aniónicas desprotonadas, hidrosulfuro (HS⁻) y sulfuro (S²⁻).

La unión de sulfuro a hemo proteínas férricas se encuentra actualmente bajo activo debate desde un punto de vista biológico y físico-químico. La literatura muestra que la especie que forma el complejo hexacoordinado está en discusión. Por un lado, en la citocromo C oxidasa^[17] y en Hbl de *L. pectinata*^[24] se postula que es H₂S la especie preferida en los complejos ferrosos/férricos, mientras que en la hemoglobina de *T. fusca*^[26] es HS⁻. En todos los casos las evidencias experimentales para postular la especie coordinada resultan ser insuficientes.

Finalmente, como se destacó en la introducción, el estudio de la unión de pequeños ligandos en hemoproteínas puede dividirse en dos etapas (una primer etapa de migración y una segunda etapa de unión del ligando). Para el caso particular de las especies de sulfuro, debe considerarse H₂S y HS⁻ por separado, ya que coexisten en proporciones similares en condiciones biorelevantes.

En conclusión, para comprender la unión de especies de sulfuro en hemoproteínas se deben estudiar las dos etapas:

- 1) Migración de la especie a través de la proteína.
- 1) Formación del enlace con Fe y posiblemente con aminoácidos distales. Cabe destacarse que la formación del complejo hexacoordinado podría modular el pKa de las especies de sulfuro, alterando las proporciones entre las especies coordinadas.

Para ello se requiere el uso de técnicas computacionales que permiten una visión atomística de los procesos químicos de interés. En este sentido, la química computacional en hemoproteínas ha sido ampliamente usada^[24,27-33], mostrando ser una buena estrategia para describir detalles moleculares y una poderosa herramienta para la predicción de eventos biológicamente relevantes.

Se decidió abordar el estudio de la interacción de especies de sulfuro con hemoproteínas tomando como caso paradigmático la proteína Hbl de *L. pectinata*. Esta proteína pertenece a un molusco que vive en ambientes lodosos ricos en azufre.

Además, esta proteína ha sido extensamente estudiada desde un punto de vista experimental, y ha mostrado una elevada afinidad por las especies de sulfuro^[23], como se mencionó en la introducción de la presente tesis.

Por otro lado, el alto contenido de residuos de fenilalanina en el sitio activo ha suscitado varias controversias acerca de la especie de sulfuro unida en términos de estabilización de la especie coordinada (se ha postulado que la especie coordinada es H₂S, estabilizada por las fenilalaninas mediante lo que se denominó “caja de Phe”)

IV.2 Metodología específica

IV.2.1 Preparación de los sistemas y parámetros para la simulación de Dinámica Molecular (MD)

La estructura de la proteína en los estados pentacoordinado (5c), Fe(III)-SH₂ y Fe(III)-SH⁻ fueron construidos a partir de la estructura de rayos-X correspondiente al PDBid: 1BOB (Hbl de *L. pectinata*), para los estados hexacoordinados (6c) se agregó el correspondiente ligando unido al centro de hierro (H₂S y HS⁻).

El estado de protonación de los amino-ácidos se asumió al correspondiente al pH-fisiológico (es decir, ASP y GLN cargados negativamente, LYS y ARG cargados positivamente), todas los residuos de histidinas expuestos a solvente fueron protonados en el átomo N-δ, así como también la histidina proximal, la cual está coordinada al hierro del grupo hemo.

Las aguas cristalográficas fueron borradas manualmente y el sistema fue solvatado construyendo una caja octaédrica de 10 Å. En este procedimiento se ubicaron aproximadamente 6840 moléculas de agua TIP3P dentro de la caja a través del criterio estándar implementado en el paquete AmberTools.

Los parámetros de todos los residuos (excepto los pertenecientes al grupo hemo) fueron tomados del campo de fuerzas ff99SB^[34] del paquete AMBER^[35].

Para obtener parámetros para las simulaciones MD del grupo hemo, se utilizó un sistema modelo para representar el grupo hemo de la proteína. Este modelo consistió de un anillo imidazólico unido a través de su N- δ a una protoporfirina férrica. Finalmente, en la posición *trans* al anillo imidazólico, una molécula de H₂S o HS⁻ fue coordinado al hierro (ver Figura 1).

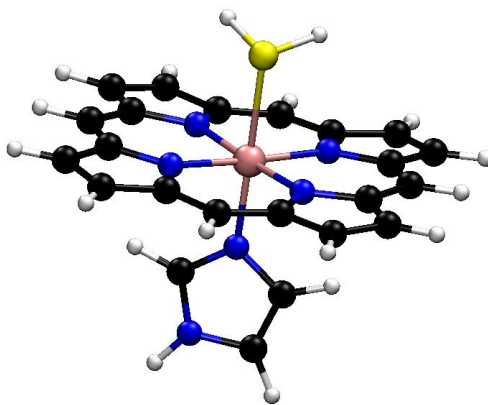


Figura 1: modelo QM utilizado para la obtención de los parámetros para la dinámica molecular clásica. La figura muestra como ejemplo el complejo con H₂S como sexto ligando.

Se realizaron optimizaciones QM sobre estas estructuras modelo para los tres complejos utilizando la teoría de DFT implementada en el programa Gaussian03.^[36] Los cálculos DFT han sido extensamente aplicados a metaloproteínas en general y en particular a porfirinas de hierro^[37–42]. Los cálculos DFT usados en esta parte de la tesis fueron realizados empleando PBE^[43] como funcional de intercambio y correlación, y 6-31G** como conjunto de funciones de base.

En cada procedimiento de optimización, se realizaron cálculos de frecuencia usando la aproximación de modos normales para chequear si la estructura optimizada corresponde a un mínimo local en la superficie de energía potencial.

Para estos sistemas en particular, el estado de spin se impuso en función de resultados experimentales conocidos. Es decir, la carga total y el estado de spin de los complejos 5c, H₂S y HS⁻ fueron impuestos para obtener complejos de tipo férricos de bajo spin ($S=1/2$) para los complejos hexacoordinados y férrico de alto spin ($S=5/2$) para el complejo 5c.

Los parámetros de enlace, ángulo y torsionales que incluían el átomo de Fe y las especies de sulfuro fueron obtenidos mediante un escaneo de la superficie de energía potencial alrededor de los mínimos asociados a la coordenada interna para la cual se deseaba asignar el parámetro. Los perfiles que resultaron de este procedimiento se ajustaron según la forma funcional de la energía potencial dada por el campo de fuerzas de AMBER. Se usaron también los parámetros obtenidos en nuestro grupo para el resto de los parámetros de la porfirina^[40,44].

Por otro lado, las cargas parciales atómicas fueron obtenidas utilizando el método de Potencial Electrostatico Restringido (RESP) sobre las estructuras optimizadas QM, a partir de cálculos de Single Point a nivel PBE/6-31G** y el método CHELPG^[45], imponiendo también simetría para los tipos de átomo equivalentes.

El procedimiento descrito anteriormente fue validado y ampliamente usado en muchos estudios de hemoproteínas en nuestro grupo^[46–48]. En particular, los parámetros

obtenidos para los complejos con las especies de sulfuro usados en las simulaciones de dinámica molecular en la presente tesis, mostraron dar resultados consistentes con los experimentos de Raman Resonante para la interacción de especies de sulfuro con la hemoglobina truncada de *T. fusca*^[26], las mediciones fueron realizadas en el laboratorio de Giulietta Smulevich de la Universidad de Napoles.

Todas las simulaciones MD fueron realizadas utilizando condiciones periódicas de contorno con 9 Å de distancia de corte y sumas de Particle Mesh Ewald (PME) para el tratamiento de las interacciones electrostáticas de largo alcance.

Los enlaces covalentes que involucran átomos de hidrógeno fueron restringidos a su posición de equilibrio utilizando el algoritmo de SHAKE, mientras que la temperatura y presión fueron mantenidas constantes usando el termostato y barostato de Berendsen respectivamente, como está implementado en el paquete AMBER^[35].

Para las tres estructuras (5c, Fe(III)-SH₂, Fe(III)-SH⁻) primeramente se realizó el siguiente protocolo de equilibración de los sistemas:

- I) 2 ns de calentamiento lento de todo el sistema desde 0 hasta 300 K a volumen constante, aplicando restricciones de tipo armónicas con 80 kcal/mol.Å² de constante asociada, para todos los Cα
- II) 2 ns de calentamiento lento de 0 a 300 K de todo el sistema a presión constante
- III) Luego de estos dos primeros pasos se realizó una simulación sin restricción de 300 ns de dinámica molecular a temperatura (300 K) y presión constantes, para obtener así un sistema descrito en el ensamble NPT.

En todos los casos el time step elegido fue de 2 fs.

Para mejorar la convergencia de las estructuras obtenidas luego de este protocolo se usaron dos estrategias:

a) Para cada estado, iniciando desde la misma estructura inicial descrita anteriormente, tres protocolos distintos de equilibración fueron aplicados, variando los parámetros de equilibración (tiempo total de simulación, fuerza de las restricciones aplicadas, etc). Este protocolo permitió obtener tres replicas decorrelacionadas entre sí para cada estado, obteniendo así, un tiempo total de simulación de 600 ns para cada estado.

b) Tomando tres diferentes estructuras decorrelacionadas entre sí tomadas de la primera trayectoria y reiniciando las velocidades atómicas, se extendió la dinámica molecular por 300 ns para cada estado.

Finalmente, luego de realizados todos los protocolos mencionados anteriormente se obtuvieron en total aproximadamente 1,3 μ s de trayectorias de dinámica molecular para cada estado.

IV.2.2 Cálculos de Muestreo de Ligando Implícito (ILS).

Como se mencionó en el capítulo II, el método ILS^[49,50] evalúa el PMF de ubicar un ligando en cualquier posición deseada en el sistema de interés (teniendo en cuenta también diferentes orientaciones del ligando), al ligando se lo suele llamar ligando “sonda”. Esta técnica ha mostrado ser una buena estrategia para la predicción de MT para pequeños ligandos neutros en hemoproteínas^[51–54].

En esta parte de la tesis, los cálculos de ILS fueron realizados en una grilla cuadrada regularmente espaciada de 0.5 Å de resolución que incluye la proteína. La sonda usada fue H₂S y 20 diferentes orientaciones (los vértices de un dodecaedro), fueron tenidas en cuenta en cada punto evaluado.

La interacción fue considerada del tipo Lennard-Jones, truncada a 9 Å, los parámetros de la sonda utilizados para las interacciones pertenecen al campo de fuerzas de AMBER y los parámetros geométricos de la sonda fueron tomados de una optimización QM (a nivel PBE/6-31G**)

Es importante destacar que como HS⁻ es un ligando cargado, la metodología ILS resulta inapropiada. Esto es debido al costo computacional al computar las interacciones de tipo Coulómbicas en las muchas posiciones y orientaciones. Además, la suposición mencionada de una interacción débil entre el ligando y la proteína podría no verse satisfecha, generando posiblemente cambios conformacionales que no serían tenidos en cuenta en las trayectorias analizadas^[50].

Los cálculos de ILS fueron realizados sobre 4000 fotos tomadas de 1200 ns de simulación para el estado 5c de la proteína. Finalmente, se consideró un MT a una zona embebida en la matriz proteica y que conectaba el sitio activo con el solvente para valores de corte de energía libre menores a 1,5 kcal/mol.

IV.2.3 Cálculos de Dinámica Molecular Guiada (sMD).

Se evaluó el perfil de energía libre de H_2S y HS^- a lo largo del MT de *L. pectinata*, utilizando la técnica de sMD^[55,56] acoplada a la ecuación de Jarzynski^[57], eligiendo la distancia Fe-S como la coordenada de transformación.

Como se destacó en el capítulo II, para la aplicación de teoría de Jarzynski se requiere que el sistema modificado esté correctamente equilibrado bajo un ensamble determinado (elegimos ensamble NPT), con el potencial de sesgo fijo al valor inicial de la coordenada de transformación. Por lo tanto, se eligió una estructura a partir de la trayectoria de la proteína en el estado 5c y se ubicó H_2S (o HS^-) a 9 Å para la distancia entre los átomos de Fe y S, la posición elegida fue asignada como la “entrada” del MT a partir de los resultados obtenidos en los cálculos de ILS.

Se realizaron entonces simulaciones MD restringiendo la distancia Fe-S (5x100 ns con H_2S y HS^- respectivamente) al valor inicial de la coordenada de transformación (9 Å) usando un valor de 200 kcal/(mol.Å²) para la constante del resorte, y usando los mismos parámetros para que la dinámica molecular descriptos en las secciones anteriores.

A partir de diferentes estructuras tomadas cada 1 ns de estas simulaciones, se obtuvieron 500 estructuras iniciales decorrelacionadas con la coordenada de transformación fija a 9 Å aproximadamente. Las estructuras que mostraban al ligando alejado al menos 1 Å de la región designada como la entrada del canal, fueron descartadas, el resto de ellas fue usada para realizar la dinámica guiada variando la coordenada de transformación desde 9 Å hasta 4 Å usando una constante de resorte de 200 kcal/(mol.Å²) y usando una velocidad de 0,0025 Å/ps.

Los parámetros de corrida k y v fueron elegidos siguiendo un esquema computacional utilizado en varios estudios anteriores.^[54,58,59] En este caso, la velocidad v fue elegida diez veces menor a la de los trabajos previos para mejorar la convergencia del método.

Finalmente, luego de realizadas las dinámicas guiadas se descartaron las trayectorias para las cuales el ligando migró por debajo del anillo del hemo por no representar el proceso de interés.

Al final del procedimiento descripto anteriormente un conjunto de 378 perfiles de trabajo para H_2S y 193 para HS^- fueron obtenidos para su posterior análisis.

Finalmente, se evaluó el estimador $\widehat{\Delta G}_{FD}$ (capítulo III, ecuación (III.38)), llamado estimador de Fluctuación-Disipación para calcular los valores de energía libre asociados a cada valor de la coordenada de transformación. Este estimador (como se mostró en el capítulo III de la presente tesis) muestra ser mucho más eficiente estadísticamente para la convergencia del valor estimado de energía libre para el caso de distribuciones Gaussianas para los valores de trabajo.

Por otro lado, con fines comparativos también se evaluó el estimador de Jarzynski $\widehat{\Delta G}_J$ (capítulo III, ecuación (III.34))

Finalmente los perfiles de energía libre fueron construidos, correspondientes a los perfiles de migración de H_2S y HS^- para el MT de Hbl de *L. pectinata*.

IV.2.3.1 Criterio de convergencia de los cálculos de Dinámica Molecular Guiada.

El análisis del error de los perfiles de energía libre, fue hecho considerando el sesgo inherente a la estimación de diferencias de energía libre usando el estimador de Fluctuación-Disipación y de Jarzynski

Se asumió que el número de trayectorias de dinámica molecular guiada N fue pequeño. Como fue propuesto por Gore et al, este esquema considera que el error expresado como la raíz del error cuadrático medio (RMSE) asociado con el estimador de Fluctuación-Disipación puede ser expresado en términos de un sesgo y una varianza en el régimen de N chico.

El estimador de Jarzynski, tiene también dos contribuciones: un término de varianza para el conjunto de N trabajos evaluados y un sesgo inherente asociado con la evaluación de la ecuación de Jarzynski usando un numero finito de valores de trabajo.^[60]

Además, asumiendo una distribución gaussiana para los valores de trabajo obtenidos en cada valor de la coordenada de transformación y un régimen de N pequeño, las dos contribuciones de cada estimador, tienen una forma funcional definida proporcional al trabajo disipado $\overline{W}_{dis}$.^[60]

$\overline{W}_{dis}$ toma la siguiente forma funcional bajo las mismas suposiciones:

$$\overline{W}_{dis} = \frac{1}{2} \beta \hat{\sigma}_w^2 \quad (IV.1)$$

Donde $\hat{\sigma}_w^2$ es el cuadrado de la desviación estándar del conjunto de N trabajos.

Finalmente, la naturaleza cuadrática del potencial de sesgo y mediante la suposición de resorte rígido (ver capítulo II) utilizado en las dinámicas moleculares guiadas, se espera obtener

una distribución Gaussiana para los valores de trabajo en cada valor de coordenada de reacción; y bajo estas condiciones, dependiendo de la dispersión de los valores de trabajo, los dos estimadores $\widehat{\Delta G}_J$ y $\widehat{\Delta G}_{FD}$ deberían coincidir.

IV.2.4 Caracterización de los sitios de hidratación (Water Sites, WS).

Los WS corresponden a regiones específicas dentro de la proteína definidas como aquellas que pueden alojar una molécula de agua con una probabilidad mayor que un agua rodeada por el entorno del seno del solvente. En trabajos previos se mostró que esas regiones pueden ser identificadas calculando la probabilidad de encontrar una molécula de agua dentro de una región definida durante una simulación MD realizada con moléculas de agua descritas explícitamente.^[61–63]

La región usada para identificar los WS fue arbitrariamente elegida de $1 \text{ \AA}^3$, y las coordenadas del centro del WS corresponden a la posición promedio de todos los átomos de oxígeno que visitaron el WS a lo largo de la simulación. Una molécula de agua es considerada como ocupante del WS cuando la distancia entre la posición de su átomo de oxígeno y el centro del WS es menor que $0,6 \text{ \AA}$.

Una vez identificado las zonas, para todos los potenciales WS, se computaron las probabilidades de encontrar agua (WFP), correspondiente a la probabilidad de encontrar moléculas de agua en una región definida por el WS y normalizada respecto de la probabilidad del seno del solvente de alojar una molécula de agua en una esfera del mismo volumen a la temperatura y presión correspondientes. Por lo tanto, WFP es de hecho usada como un valor de corte para decidir cuales potenciales WS son considerados relevantes. En este trabajo, solo los WS con valores de WFP mayores que 2 unidades fueron retenidos.

Todos los cálculos de WS fueron realizados usando el módulo de VMD llamado WATCLUST^[64], con los parámetros por defecto y considerando que un WS está formado cuando las moléculas de agua son alojadas más del 10% de la escala de tiempo de las simulaciones MD.

Para determinar los WS, se usó aproximadamente 3000 fotos de toda la simulación MD en cada estado. La convergencia fue chequeada usando diferentes segmentos de todas las dinámicas moleculares analizadas usadas para definir los WS, asegurando que la definición de WS resultó robusta.

IV.2.5 Cálculos QM/MM: optimizaciones restringidas

Los cálculos QM/MM fueron realizados sobre los complejos de H₂S y HS⁻ coordinados al átomo de Fe(III) de Hbl de *L. pectinata*. Las estructuras iniciales para los cálculos QM/MM fueron obtenidas a partir de las simulaciones MD descritas anteriormente. Se eligieron como estructuras iniciales a las que representaban conformaciones relevantes basadas en la estructura y el patrón de puentes de hidrógeno (ver resultados). Estas estructuras fueron enfriadas hasta 0 K, con los mismos parámetros generales que la simulación MD.

A partir de estas estructuras enfriadas se realizaron cálculos de optimización de geometría multiescala QM/MM a nivel DFT usando el código llamado Hybrid^[65], implementado en el programa SIESTA descrito en la sección de métodos generales.

Para todos los átomos QM, se usó un conjunto de base doble zeta con funciones de polarización. Todos los cálculos fueron realizados bajo la aproximación de gradiente generalizado utilizando el funcional de intercambio-correlación llamado PBE.^[43] Este funcional ha sido usado en muchos trabajos anteriores para describir la reactividad de pequeñas moléculas en hemoproteínas.^[47,66] Además, ha dado resultados satisfactorios y confiables para describir reacciones que incluyan transferencia de protón^[67–69].

Para todos los sistemas estudiados, la aproximación de spin no restringido fue utilizada, y usamos el término “bajo spin” para el estado de tipo doblete (spin total, $S=1/2$), mientras que “alto spin” para denominar un estado de tipo sexteto (spin total, $S=5/2$)

Solo los residuos ubicados a menos de 10 Å de la región QM fueron optimizados libremente en los cálculos QM/MM. La porción QM incluyó el grupo hemo (sin los grupos periféricos), el ligando H₂S o HS⁻, el residuo Glutamina (o Alanina) y el anillo imidazólico del sitio proximal.

El resto de la proteína y las moléculas de agua, fueron tratados clásicamente. La frontera entre las porciones QM y MM del sistema fueron tratadas usando el método de Link-Atom. Esta estrategia ha sido usada satisfactoriamente para varios sistemas que incluyen biomoléculas^[47,70–76].

Como obtener perfiles de energía libre precisos requiere un muestreo extenso, lo cual es computacionalmente muy caro y difícil de conseguir a un nivel DFT QM/MM, se calcularon

los perfiles de energía QM/MM usando optimizaciones restringidas a lo largo de una coordenada de transformación que describa el proceso de desprotonación.

Para lograr lo anterior, un término adicional, $V_R = k(\xi - \xi_0)^2$ es agregado a la energía potencial, donde k es una constante de fuerza ajustable (se eligió 200 kcal/mol.Å²) en este caso), ξ es la coordenada de transformación seleccionada y ξ_0 es el valor de inicial de referencia, el cual posteriormente es variado con intervalos sucesivos de 0.5 Å. Variando ξ_0 , el sistema es forzado a recorrer el camino de la coordenada de transformación seleccionada ξ .

La forma cuadrática impone una restricción en energía para configuraciones donde ξ y ξ_0 difieran mucho, generando una fuerza que mantiene el sistema en las cercanías del valor de la coordenada de transformación deseada. Dejando relajar el resto del sistema, el camino recorrido sobre esa coordenada será el de mínima energía en la superficie de energía potencial.

IV.2.6 Cálculos QM/MM: dinámica molecular.

Las simulaciones QM/MM-MD fueron llevadas a cabo usando el módulo Lio-QM/MM^[77,78] implementado en el paquete AMBER. Esta implementación ha recientemente incorporado la posibilidad de ser ejecutada empleando placas de video, lo que la hace especialmente eficiente.^[79]

Los parámetros del campo de fuerza y el protocolo de termalización utilizado puede encontrarse en la literatura^[78].

La porción QM y MM fue similar al usado para las optimizaciones estáticas QM/MM descriptas en la sección anterior. La diferencia en la porción QM es que se decidió incluir una molécula de agua en esta región, esta elección resulta ser menos costosa computacionalmente sin pérdida de generalidad en el resultado (ya que en los cálculos de optimización QMMM estática obtuvimos similares resultados utilizando una o varias moléculas de agua).

La interface entre las porciones QM y MM fueron descriptas utilizando el método de Link Atom^[65]. A diferencia del código Hybrid, en este caso se emplean funciones de base de tipo Gaussiano, y todo el cálculo se realiza en el espacio real.

Para la región QM, los cálculos fueron realizados a nivel de la aproximación de gradiente generalizado, usando PBE^[43] como funcional de intercambio-correlación con un conjunto de bases DZVP para la expansión de los orbitales mono-electrónicos bajo una aproximación de spin

no restringido, utilizando como valores de carga y spin los que determinen un complejo férrico de bajo spin (multiplicidad 2) para el complejo con H_2S .

Las densidades electrónicas fueron también expandidas en un conjunto de bases auxiliares y los coeficientes para el ajuste fueron computados minimizando el error en la energía de repulsión Coulómbica.

Las configuraciones iniciales fueron generadas a partir de las simulaciones MD para el estado Fe(III)-SH_2 , seguidamente se realizó 100 ps de enfriamiento MM desde 300 K a 0 K y una optimización de geometría QM/MM. Finalmente, un procedimiento de termalización fue realizado de 0 K a 300 K usando un esquema de dinámica de Langevin. Se utilizó un algoritmo de Verlet para integrar las ecuaciones de Newton con un time step de 0.1 fs.

La temperatura fue mantenida constante a 300K utilizando el termostato de Langevin. Todas las estructuras resultantes de los cálculos QM/MM (estáticos y dinámicos) fueron visualizados con VMD 1.9.1.^[80]

IV.3 Resultados

IV.3.1 Estabilidad de las dinámicas moleculares

Se calcularon los valores de RMSD en cada dinámica molecular utilizando las trayectorias completas en cada estado ($5c$, Fe(III)-SH_2 y Fe(III)-SH^-). Se encontró que todas las estructuras fueron estables durante toda la escala de tiempo de la simulación, como evidencian los siguientes gráficos de RMSD en función del tiempo (Figura 2)

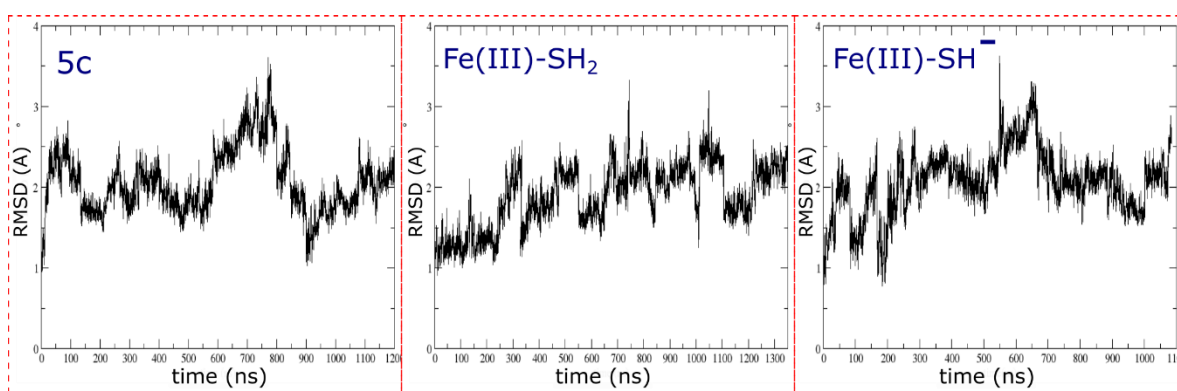


Figura 2: Análisis RMSD para las trayectorias MD para la Hbl de *L. pectinata* en los estados $5c$, Fe(III)-SH_2 y Fe(III)-SH^- (de izquierda a derecha, respectivamente)

IV.3.2 Caracterización de canales internos de migración.

Un canal interno de migración (MT) es un área inmersa en la matriz proteica que conecta zonas de baja energía libre (conocidas como “cavidades transientes”) que pueden ser formadas por las fluctuaciones térmicas en la estructura de una proteína.

Se ha demostrado que algunas proteínas de la familia de las globinas presentan varios MT conectando, el sitio activo con el solvente. Se ha propuesto que tanto los aminoácidos que rodean al MT como las moléculas de agua en el sitio activo pueden tener un rol fundamental en la modulación de la migración de ligandos a través de la matriz proteica y la unión al centro metálico^[81–84].

Se realizaron entonces cálculos de ILS en la proteína en el estado férrico pentacoordinado (5c). Se usó H₂S como molécula sonda. Como se mencionó, el método ILS no puede ser usado para evaluar diferencias entre H₂S y HS⁻, la sonda utilizada entonces provee una evidencia cualitativa acerca de los potenciales canales de migración para pequeños ligandos.

Se encontró un solo MT que conecta el solvente con el sitio activo (Figura 3A). Este canal se corresponde con el llamado “puerta E7” encontrado en mioglobina y hemoglobina^[85,86]. La Figura 3B representa una ampliación de la vecindad de este MT, mostrando los grupos propionatos solvatados y el aminoácido distal Glutamina en la posición 65 (Gln65)

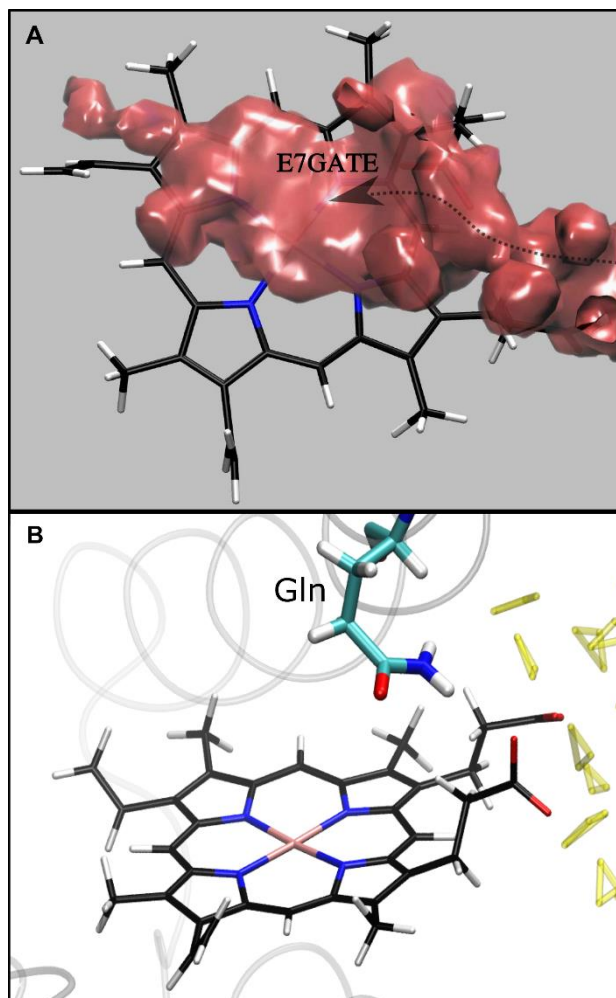


Figura 3: A) Isosuperficie de energía libre (1.5 kcal/mol de valor de corte) calculado con ILS, usando H₂S como ligando sonda, en Hbl de *L. pectinata*, en el estado férrico 5c. B) Detalle de la “puerta E7” del MT, donde se muestran los propionatos solvatados y el residuo polar Gln. Los átomos en rojo y azul representan grupos cargados, las moléculas de agua se muestran en amarillo.

Para chequear la convergencia del cálculo, se dividió el número total de fotos en tres bloques de 1000 fotos que fueron analizados separadamente. Se obtuvieron resultados similares para los tres bloques como se muestra en la Figura 4.

Todos los cálculos ILS fueron realizados utilizando el programa modular VMD 1.9.1^[80].

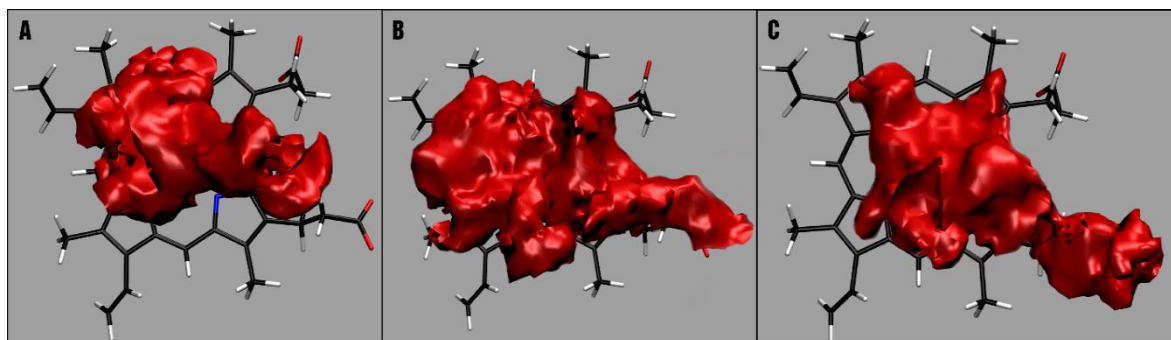


Figura 4: análisis de ILS para tres bloques diferentes (A, B, C, de 1000 fotos cada una) tomadas de una trayectoria MD para el estado 5c de la proteína.

IV.3.3 Perfil de energía libre para la migración de especies de sulfuro (H_2S vs HS^-)

Para evaluar las diferencias entre los ligandos H_2S y HS^- , se calcularon los perfiles de energía libre de migración a lo largo del único MT encontrado, mediante simulaciones de dinámica molecular guiada (sMD) combinadas con la igualdad de Jarzynski (Figura 5). Esta metodología ha sido aplicada satisfactoriamente en muchos estudios de proteínas.^[54,59,87-89] La distancia Fe-S fue elegida como la coordenada de transformación.

En función de los resultados presentados en el capítulo II de la presente tesis, se evaluaron los perfiles de migración de H_2S y HS^- utilizando el estimador de máxima-verosimilitud asociado a una distribución Gaussiana de los trabajos, $\widehat{\Delta G}_{FD}$. Para comparar y ejemplificar la diferencia entre los estimadores, se graficó también el perfil de energía libre de migración estimado a través del estimador de Jarzynski, $\widehat{\Delta G}_J$.

Los resultados muestran H_2S es la especie favorecida para migrar hacia el sitio activo (Figura 5, paneles superiores), con una baja barrera de energía libre (~ 2 kcal/mol) y una baja diferencia de energía libre entre H_2S en solvente y H_2S en sitio activo (~ -1 kcal/mol). Por otro lado, HS^- muestra una alta barrera y diferencia de energía libre entre HS^- en el solvente y HS^- en el sitio activo de aproximadamente ~ 10 kcal/mol, con una ausencia aparente de un mínimo de energía libre cuando el ligando alcanza el sitio activo.

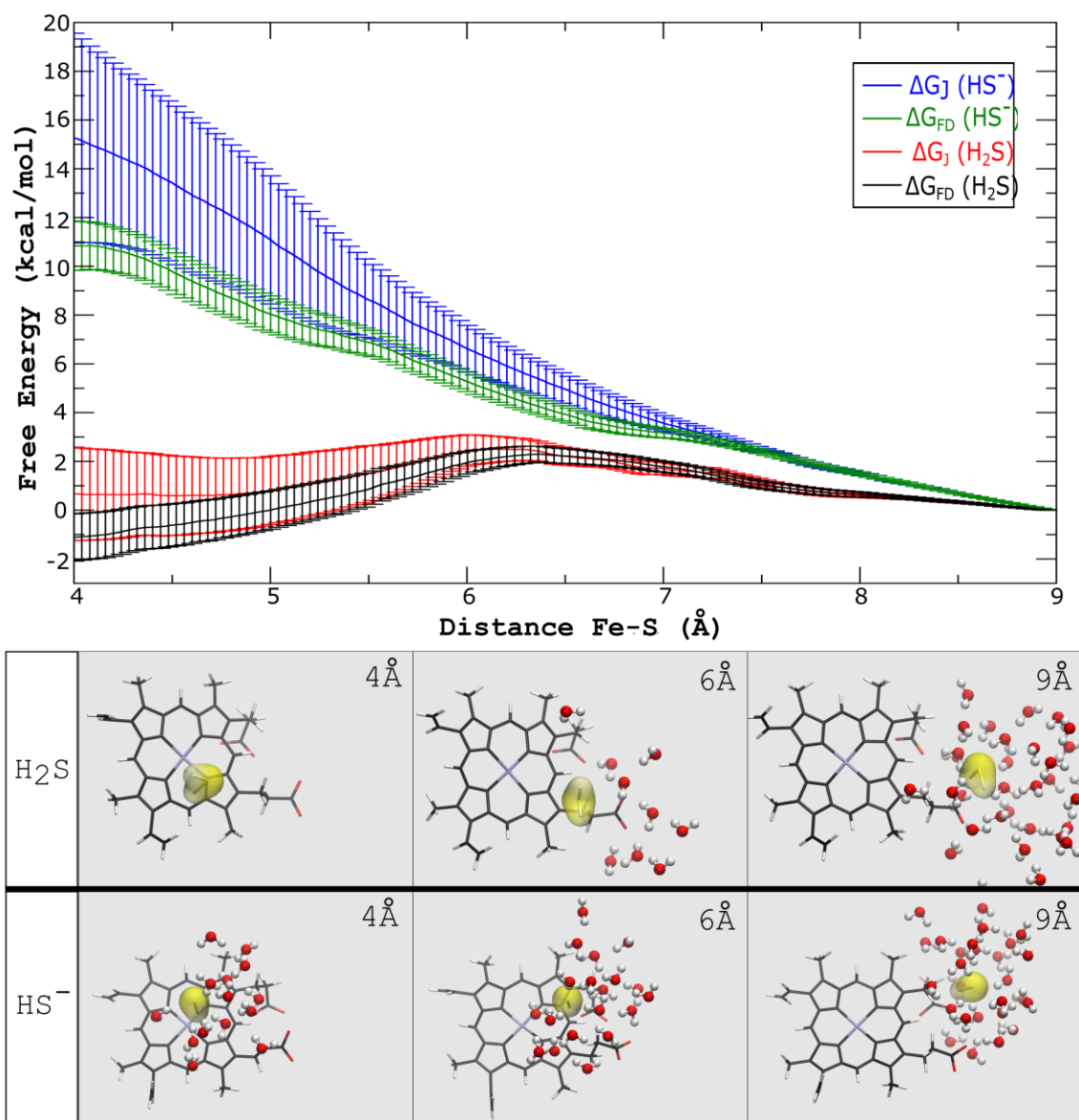


Figura 5: **Paneles superiores:** Perfil de energía libre de migración de H₂S a través del MT de HbI de *L. pectinata*, utilizando el estimador de Jarzynski y de Fluctuación-Disipación (rojo y negro, respectivamente). Perfil de energía libre de migración de HS⁻ a través del MT de HbI de *L. pectinata*, utilizando el estimador de Jarzynski y de Fluctuación-Disipación (azul y verde, respectivamente). Los valores de RMSE asociado a cada estimador se muestran como barras de error en cada perfil. **Paneles inferiores:** figura representativa tomada de las trayectorias sMD para H₂S y HS⁻ a diferentes distancias Fe-S (4, 6 y 9 Å de izquierda a derecha, respectivamente). En todos los casos, se muestran las moléculas de agua ubicadas a menos de 5 Å del átomo de azufre.

Cabe mencionar que los perfiles están finalizados a 4 Å de distancia Fe-S debido al campo de fuerzas clásico que se utiliza que no permite describir la interacción electrónica durante la formación del enlace Fe-S.

La diferencia observada en los perfiles de migración de ambos ligandos puede ser racionalizada inspeccionando las trayectorias de sMD de ambos ligandos. Cuando H₂S migra desde el solvente al sitio activo, su estructura de solvatación cambia dramáticamente, expeliendo moléculas de agua de su alrededor mientras ingresa al sitio activo (Figura 5, paneles

inferiores), mientras que por el contrario, HS^- migra al sitio activo parcialmente solvatado, principalmente debido a su carga neta.

La Figura 6 muestra un análisis de la convergencia de la estimación de la energía libre analizando dos estimadores ($\widehat{\Delta G}_{FD}$ y $\widehat{\Delta G}_J$) para la comparación entre sus valores estimados y sus incertezas asociadas.

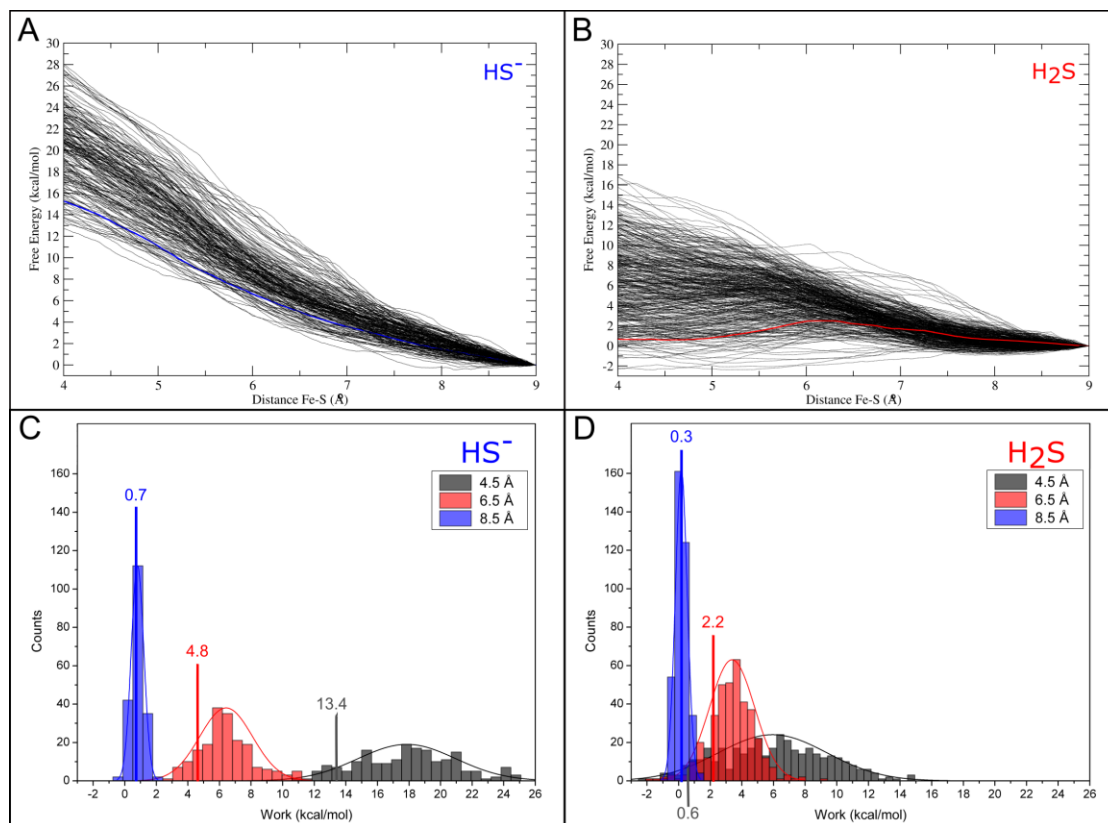


Figura 6: Análisis de las simulaciones de sMD. (A, B) Conjunto de trabajos acumulados para cada valor de la coordenada de transformación, obtenida del protocolo sMD para HS^- y H_2S . Los valores de energía libre obtenidas a partir del estimador de Jarzynski (ecuación (IV.3)) se muestran en azul y rojo. (C, D) Distribuciones de trabajo para los valores de coordenada de 4.5 Å, 6.5 Å y 8.5 Å para HS^- y H_2S . Se graficaron también las distribuciones Normales asociadas al promedio y desvío estándar de cada conjunto de valores de trabajo (líneas sólidas en azul, rojo y gris).

Los números en azul, rojo y gris son los valores de energía libre usando el estimador de Jarzynski para cada distribución de trabajos en 8.5 Å, 6.5 Å and 4.5 Å, respectivamente.

Los paneles A y B muestran conjuntos de trabajos acumulados bien diferenciados durante todas las trayectorias de sMD y por lo tanto ambos perfiles de energía libre son significativamente diferentes.

Por otro lado, los paneles C y D, muestran las correspondientes distribuciones para los trabajos para tres valores distintos de la coordenada de transformación y la curva Gaussiana correspondiente al promedio y la desviación estándar de los datos medidos. Se puede observar que las distribuciones de los trabajos se pueden aproximar a distribuciones Gaussianas, y estas distribuciones son más dispersas hacia valores finales de la coordenada de transformación con una creciente pérdida de evidencia de distribución Gaussiana.

La naturaleza Gaussiana de las distribuciones (en función de lo explicado en el capítulo III de la presente tesis) permite concluir que la estimación del perfil de energía libre usando $\widehat{G}_{FD}$, resulta ser la más eficiente y precisa ya que es el estimador de máxima-verosimilitud asociado.

Con fines comparativos se calculó también el perfil de energía libre usando $\widehat{G}_J$ (ver Figura 5, panel superior)

Los paneles C y D de la Figura 6, muestran además que se espera que los valores estimados con $\widehat{G}_J$ estén satisfactoriamente convergidos, al observar varios valores medidos que caen en la “cola” de la distribución, con una cantidad significativa de datos de trabajo muestreados cerca de esa zona. Además, si bien la dispersión de los valores de trabajo en cada valor de coordenada es mayor para HS^- que para H_2S , ambas dispersiones son menores a $\sigma \sim 3 \text{ kcal/mol}$, y como se vio, para estos valores pequeños de dispersión, los valores estimados serán similares.

Como comentario final, se obtuvieron valores de $\overline{W}_{dis}$ en el orden de aproximadamente 5 a 10 unidades de $k_b.T$ al final de los valores de coordenada de transformación, es esperable entonces que más trayectorias de sMD sean necesarias para obtener una mayor evidencia de distribución Gaussiana para los trabajos medidos.^[60]

La presente comparación muestra un caso particular en donde las conclusiones químicas que se derivan del cálculo sMD no son tan diferentes al usar uno u otro estimador, a diferencia de lo mencionado para MnSOD y el anión superóxido.

IV.3.4 Formación de sitios de hidratación (WS)

Como se mencionó en la introducción de esta tesis las moléculas de agua fuertemente estabilizadas en el sitio activo, pueden hacer más lento el proceso de unión de ligandos, principalmente debido a que un proceso de desplazamiento de aguas debe ocurrir primeramente^[81–83].

Para empezar, definiremos un sitio de hidratación (WS) como la región espacial dentro del sistema con una alta probabilidad de encontrar moléculas de agua en comparación con la probabilidad de encontrar moléculas de agua en una región del mismo tamaño en el seno del solvente.

Se realizaron cálculos de obtención de WS usando las trayectorias de simulaciones MD para los estados 5c, Fe(III)-SH₂ y Fe(III)-SH⁻ considerando un WS como la zona donde las

moléculas de agua están presentes en al menos 10% de la escala de tiempo de la dinámica analizada.

No se observó la presencia de ningún WS en la cara distal del hemo para el estado 5c, como se muestra en la Figura 7A. Se encontró, sin embargo, dos WS distales en la cara distal del hemo en el estado Fe(III)-SH₂ (Figura 7B). La ausencia de WS en la cara distal en el estado Fe(III)-SH⁻ evidencia una disminución en la polaridad del sitio activo, comparado con el estado Fe(III)-SH₂ (Figura 7C).

Es importante destacar, que en los tres casos un WS fue identificado en la cara proximal del grupo hemo. Este WS está conservado y no participa en la formación del enlace Fe-S.

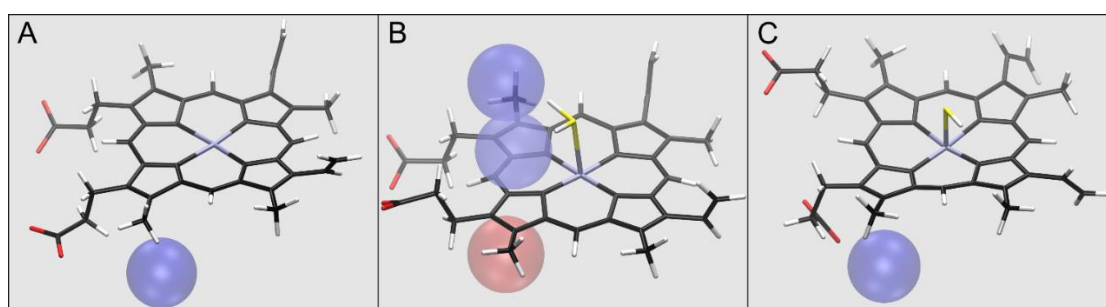


Figura 7: WS observados en los estados férricos 5c (A), 6c H₂S (B) y 6c HS⁻ (C) ferric de Hbl de *L. pectinata*. Las regiones rojas y azules representan una probabilidad baja y alta de encontrar moléculas de agua con respecto al solvente, respectivamente. Los tres casos presentan WS en la cara proximal del grupo hemo (abajo del hemo en la figura), el cual no está involucrado en la unión de sulfuro. La histidina proximal fue omitida para mayor claridad.

Una inspección detallada del sitio activo revela que el N del grupo amino de la Gln65 es capaz de formar un enlace de hidrógeno con el átomo de S en el complejo Fe(III)-SH⁻, evitando así, la formación de WS (Figura 8B). En el complejo Fe(III)-SH₂, sin embargo, la Gln65, asiste en la formación de dos WS en la vecindad del sitio activo, que interacciona con el H₂S coordinado (Figura 8A)

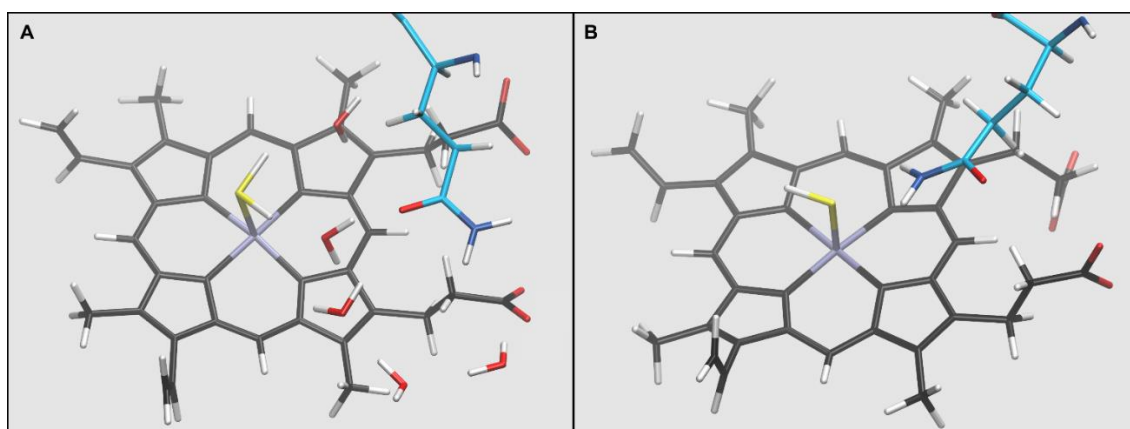


Figura 6: Estabilización del H₂S y HS⁻ coordinado durante la simulación MD. A) Estructuras representativas de la interacción de puentes de hidrógeno entre Gln y H₂S en el complejo Fe(III)-SH₂, se muestran también las moléculas de agua alrededor de Gln65 mostrando la interacción entre el H₂S coordinado y el residuo de Gln. B) Estructura representativa de la interacción de puente de hidrógeno entre N(Gln) y el átomo de S en el complejo Fe(III)-SH⁻.

IV.3.5 Primer paso de desprotonación: de Fe(III)-SH₂ a Fe(III)-SH⁻

En la sección anterior, se mostró que principalmente H₂S y no HS⁻ es la especie que preferencialmente migra a través del canal polar E7 encontrado para Hbl de *L. pectinata*. En esta sección, se mostrarán los resultados de los cálculos QM/MM que permitirán evaluar la especie que permanece finalmente coordinada al hierro. Estos cálculos realizados mediante optimizaciones restringidas usando el código Hybrid.

Encontramos que la desprotonación del H₂S ligado es termodinámicamente favorable (con un valor negativo de ΔE para el proceso global), y que el proceso ocurre sin barrera (Figura 9A). También se observó que el protón es transferido del Fe(III)-SH₂ a una molécula de agua perteneciente al WS, y subsecuentemente al grupo carbonilo de la Gln65 (Figuras 9B y 9C). No se observaron diferencias significativas para los estados de alto y bajo spin (Figura 9A).

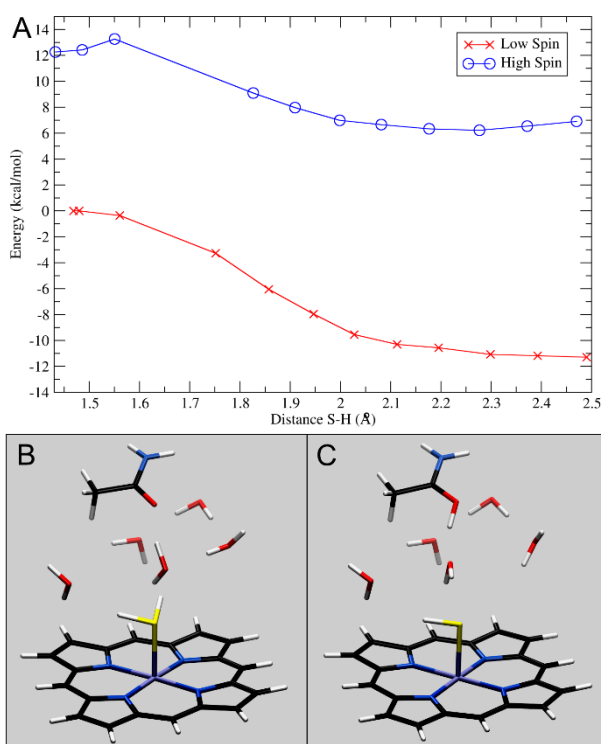


Figura 9: Perfil de energía de desprotonación de Fe(III)-SH₂ en Hbl de *L. pectinata* (A), en rojo y en azul se muestran los perfiles correspondientes a los estados de bajo y alto spin, respectivamente. (B) y (C), estructuras inicial y final para el perfil de energía. La coordenada de transformación fue la distancia S-H.

Como experimento control, se realizaron cálculos QM/MM para el complejo Fe(III)-OH₂, en vez de Fe(III)-SH₂ (Figura 10). En este caso, el ΔE global fue positivo, lo cual indica que la molécula de agua en estas condiciones es menos proclive a liberar el protón.

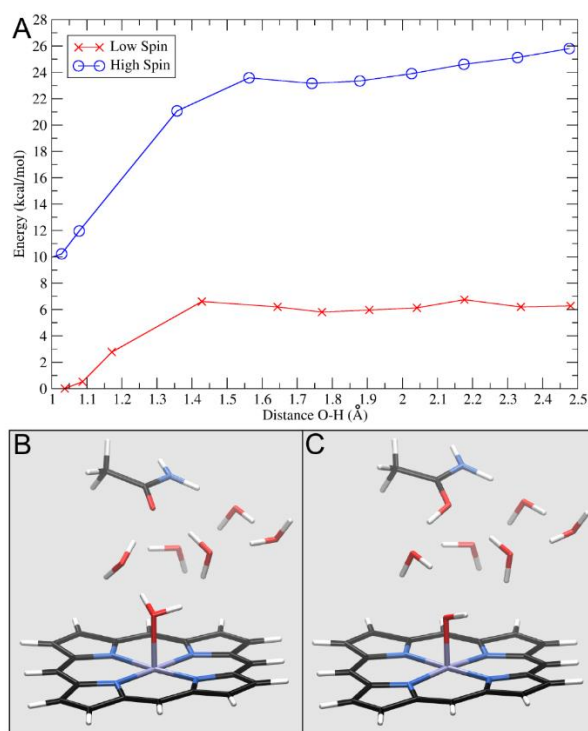


Figura 10: Perfil de energía de desprotonación de Fe(III)-OH₂ en Hbl de *L. pectinata* (A), en rojo y en azul se muestran los perfiles correspondientes a los estados de bajo y alto spin, respectivamente. (B) y (C), estructuras inicial y final para el perfil de energía. La coordenada de transformación fue la distancia O-H.

Resultados similares fueron obtenidos usando diferentes cantidades de moléculas de agua, pertenecientes a los WS en la porción QM para Fe(III)-OH₂ y Fe(III)-SH₂.

Para evaluar la relevancia de la Gln65 en el proceso de desprotonación se realizó el mismo tipo de cálculo mutando Gln65 por un residuo de Alanina (Figura 11).

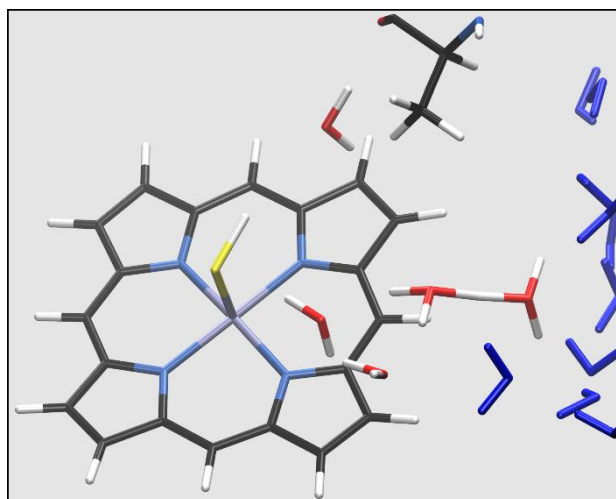


Figura 11: Estructura optimizada QM/MM para el complejo Fe(III)-SH₂ en Hbl de *L. pectinata*, mutando Gln por Ala en el estado férrico de bajo. La figura muestra en azul varias moléculas de agua MM pertenecientes al seno del solvente (nuevamente el anillo imidazólico de la histidina proximal QM fue omitido para mayor claridad). Resultados similares fueron obtenidos para el estado férrico de alto spin.

En ambos estados de spin, la desprotonación ocurre durante el paso de optimización. Este resultado evidencia que la Gln65 no es requerida para la desprotonación, pero podría participar del proceso.

Para agregar más evidencia acerca de la desprotonación del H_2S coordinado, y el rol de la Gln65 durante el proceso, se realizó una simulación de dinámica molecular QM/MM (QM/MM-MD) utilizando el código Lio implementado en el paquete AMBER. Esta metodología permite un mayor muestreo del espacio configuracional que las optimizaciones QM/MM descripta en secciones anteriores ya que incluye fluctuaciones debido a la temperatura.

Se observó que ya en la etapa de termalización el protón es espontáneamente transferido del Fe(III)-SH_2 a la molécula de agua del WS, y luego hacia la Gln65 (Figura 12), evidenciando que la desprotonación es termodinámicamente favorable a T ambiente, consistente con los resultados arrojados por las optimizaciones QM/MM.

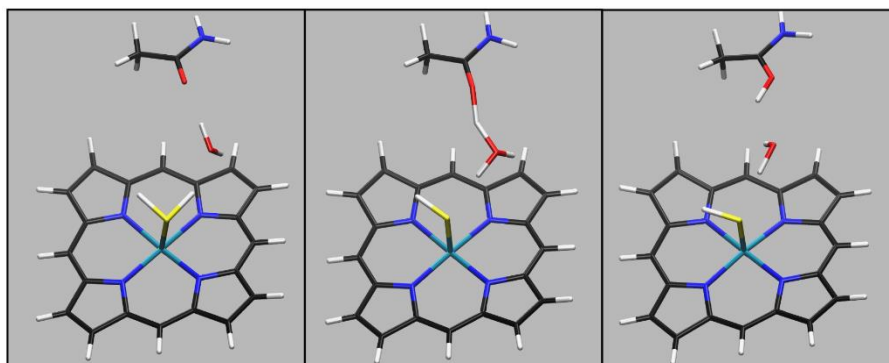


Figura 12: estructuras consecutivas de la porción QM de Hbl de *L. pectinata* para el proceso de desprotonación de Fe(III)-SH_2 durante la etapa de termalización del cálculo QM/MM-MD (el anillo imidazólico de la histidina proximal QM fue omitido para mayor claridad)

IV.3.6 Segundo paso de desprotonación: de Fe(III)-SH^- al Fe(III)-S^{2-}

Se evaluó la posibilidad de una segunda desprotonación del H_2S coordinado (a S^{2-}) realizando cálculos QM/MM para el aducto Fe(III)-SH^- .

Considerando que el aducto Fe(III)-SH^- no posee WS, la barrera cinética para la desprotonación sería alta (Figura 7C), se decidió evaluar la posibilidad de una segunda eventual y rápida desprotonación considerando situaciones antes de que los WS exhibidos en el complejo Fe(III)-SH_2 se desarmen. Para esto, se evaluó la desprotonación en presencia de los WS encontrados para el complejo Fe(III)-SH_2 .

Bajo estas condiciones, la desprotonación resultó ser favorable, aunque con un valor de ΔE cercano a 0 (~ -4 kcal/mol y ~ 0.1 kcal/mol para los estados de bajo y alto spin respectivamente). El proceso no mostró barrera y se caracterizó por una transferencia del protón deslocalizándose sobre la cadena de moléculas de agua que conectan el sitio activo con el solvente (Figuras 13A y 13B).

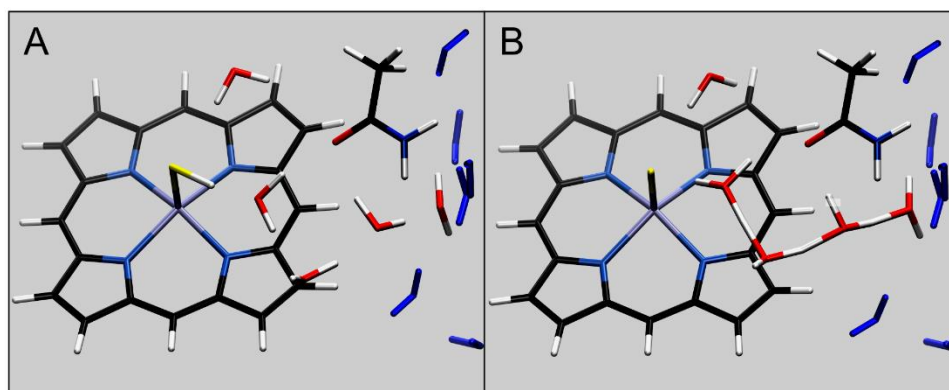


Figura 13: desprotonación del complejo Fe(III)-SH⁻ en Hbl of *L. pectinata* en el estado férrico de bajo spin. Panel A, estructura inicial usada para la optimización. Panel B estructura final obtenida después de la optimización QM/MM. Varias moléculas de agua de la región MM pertenecientes al seno de la solución se muestran en azul (el anillo imidazólico de la histidina proximal QM fue omitido para mayor claridad). Resultados similares fueron obtenidos para el estado de alto spin.

Bajo estas condiciones, el proceso de desprotonación ocurrió sin intervención de la Gln65, dando un protón deslocalizado sobre las moléculas de agua que direccionan hacia el solvente. Considerando el valor pequeño de ΔE obtenido, no se pudo ser concluyentes acerca de la evidencia para el segundo paso de desprotonación, ya que estos resultados pueden ser dependientes de detalles del modelo usado.

IV.4 Discusión

A partir de los resultados de ILS, sMD y QM/MM obtenidos, se mostró nueva información acerca del impacto del equilibrio iónico de H₂S para su migración y unión al Fe(III) de Hbl de *L. pectinata*.

Se encontró un solo canal de migración en Hbl de *L. pectinata*. Los perfiles de energía libre para la migración de ligandos a través de este canal de migración muestran que en general el perfil de energía libre a lo largo de la coordenada de transformación muestreada presenta barreras más altas para HS⁻ que para H₂S. Esta diferencia puede ser atribuida a la mayor costo energético para desolvatar el HS⁻ debido a una mayor polaridad. Esta desolvatación parece ser un requerimiento para el ingreso de ligandos ya que el sitio activo es fundamentalmente hidrofóbico debido al alto contenido de residuos de fenilalaninas. Existe entonces un balance

entre la desolvatación del HS^- y el costo energético del mismo al entrar aún solvatado por varias moléculas de agua. Ambos escenarios generan una alta barrera energética para la migración del HS^- .

La pequeña barrera obtenida para H_2S corresponde al proceso de desolvatación que tiene lugar aproximadamente a 6 Å. Más allá de este valor de coordenada, la energía libre baja debido a la estabilización del ligando en un entorno hidrofóbico.

Es interesante notar que a pesar de que el H_2S migra principalmente sin moléculas de agua, los WS son formados luego de la coordinación, y HS^- migra rodeado por moléculas de agua y los WS no se forman después de la coordinación. La única manera de reconciliar esta discrepancia aparente en la solvatación, es concluir que la especie que migra efectivamente es de hecho H_2S , y la formación del aducto Fe(III)-SH^- es el resultado de una subsecuente desprotonación del ligando H_2S coordinado.

La diferencia en el comportamiento diferencial entre los ligandos coordinados, puede ser explicada a partir de la distribución de cargas de los ligandos sulfuros.

La carga del H_2S pasa a ser de neutra a más positiva cuando se coordina, lo cual atrae a los WS y el grupo carbonilo de la Gln65. Sin embargo, cuando se coordina HS^- , su carga negativa se sigue lo que atrae el grupo amino de la Gln65 que evita la formación de los WS.

Finalmente, podemos enriquecer el análisis considerando los valores de pKa de H_2O y de H_2S , 14 y 7 unidades respectivamente. Se ha demostrado que luego de la coordinación a Fe(III) el valor de pKa del agua en hemoproteínas baja entre 4 y 7 unidades^[23,90–94].

Por lo tanto, si asumimos el mismo comportamiento para los valores de pKa para el H_2S , el equilibrio que se debería tener en cuenta para las especies coordinadas en ambientes neutros biorrelevantes debería ser $\text{Fe(III)-SH}^-/\text{Fe(III)-S}^{2-}$. Los resultados QM/MM muestran que el primer paso de desprotonación del Fe(III)-SH_2 es termodinámicamente favorable. Un valor negativo de ΔE (~ -10 kcal/mol) y una transferencia de protón espontánea en los cálculos QMMM-MD, aportan evidencia sobre este proceso. Debido a la presencia de moléculas de agua en el sitio activo, no hay una barrera cinética apreciable para la transferencia de protón.

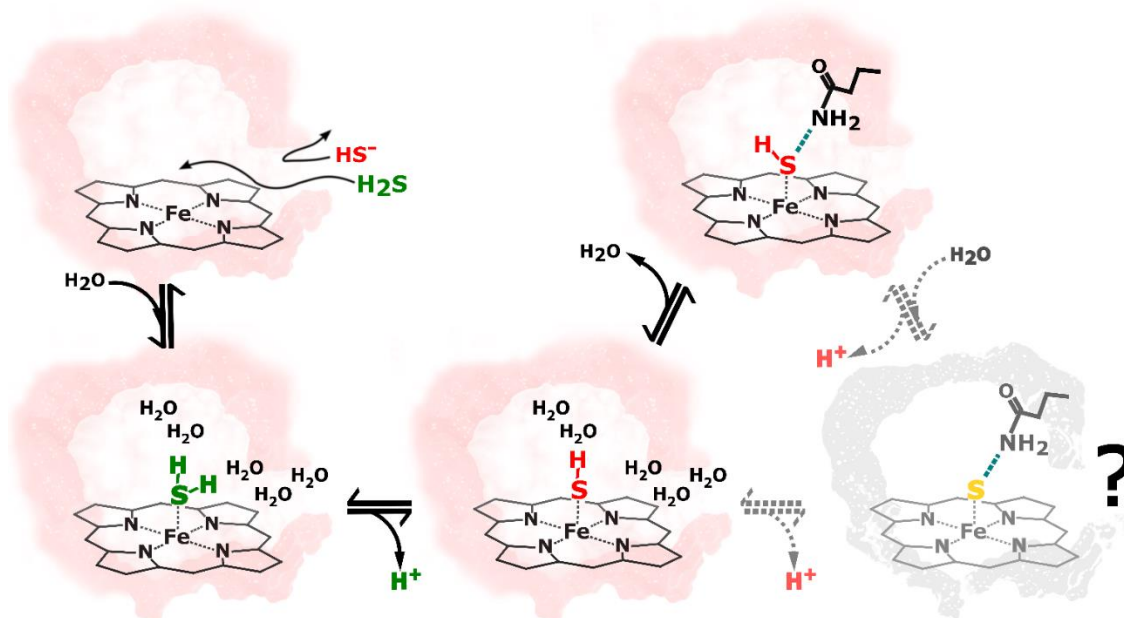
Los resultados para la desprotonación de Fe(III)-SH^- muestran, sin embargo, un valor menos negativo para el ΔE (aproximadamente de -2 kcal/mol). Considerando la pequeña magnitud de este valor, no se puede ser concluyente acerca de que la desprotonación de Fe(III)-SH^- ocurra de manera espontánea o no, en base a nuestros cálculos.

Basado en las simulaciones clásicas y QM/MM, se propone el siguiente mecanismo (Esquema 1) para el proceso global de migración y unión de las especies de sulfuro en HbI de *L. pectinata*: H_2S migra desde el solvente hasta el sitio activo y se coordina al hierro. Inmediatamente después, dos WS son formados en el sitio activo, generando una estructura tipo puente que incluye al ligando coordinado y el grupo carbonilo de la Gln65. Uno de los WS acepta un protón del H_2S coordinado y lo transfiere o bien a la Gln65 (que eventualmente debido a fluctuaciones térmicas podría liberar el protón al seno del solvente) o a otra molécula de agua orientada hacia el seno del solvente.

Esta primera etapa de desprotonación, permite la formación del complejo Fe(III)-SH^- , el cual subsecuentemente o bien rápidamente libera un segundo protón hacia las moléculas de agua, resultando en el complejo Fe(III)-S^{2-} , o se mantiene como Fe(III)-SH^- estabilizado por la Gln65.

Nuestros resultados muestran claramente que la especie Fe(III)-SH_2 no es la especie termodinámica ni cinéticamente más favorecida. En su lugar, Fe(III)-SH^- emerge como la especie coordinada, aunque no se pudo definir mediante las técnicas de simulación usadas acerca de la estabilidad de la especie Fe(III)-S^{2-} .

Esquema 1: mecanismo propuesto para la especiación de las especies de sulfuro en HbI de *L. pectinata*



IV.5 Referencias

- (1) Warenycia, M. W.; Goodwin, L. R.; Benishin, C. G.; Reiffenstein, R. J.; Francom, D. M.; Taylor, J. D.; Dieken, F. P. Acute Hydrogen Sulfide Poisoning. Demonstration of Selective Uptake of Sulfide by the Brainstem by Measurement of Brain Sulfide Levels. *Biochemical Pharmacology* **1989**, 38 (6), 973–981.
- (2) Goodwin, L. R.; Francom, D.; Dieken, F. P.; Taylor, J. D.; Warenycia, M. W.; Reiffenstein, R. J.; Dowling, G. Determination of Sulfide in Brain Tissue by Gas Dialysis/Ion Chromatography: Postmortem Studies and Two Case Reports. *Journal of Analytical Toxicology* **1989**, 13 (2), 105–109.
- (3) Furne, J.; Saeed, A.; Levitt, M. D. Whole Tissue Hydrogen Sulfide Concentrations Are Orders of Magnitude Lower than Presently Accepted Values. *AJP: Regulatory, Integrative and Comparative Physiology* **2008**, 295 (5), R1479–R1485.
- (4) Ishigami, M.; Hiraki, K.; Umemura, K.; Ogasawara, Y.; Ishii, K.; Kimura, H. A Source of Hydrogen Sulfide and a Mechanism of Its Release in the Brain. *Antioxidants & Redox Signaling* **2009**, 11 (2), 205–214.
- (5) Wintner, E. A.; Deckwerth, T. L.; Langston, W.; Bengtsson, A.; Leviten, D.; Hill, P.; Insko, M. A.; Dumpit, R.; VandenEckart, E.; Toombs, C. F.; Szabo, C. A Monobromobimane-Based Assay to Measure the Pharmacokinetic Profile of Reactive Sulphide Species in Blood: Monobromobimane Assay for Sulphide Detection. *British Journal of Pharmacology* **2010**, 160 (4), 941–957.
- (6) Abe, K.; Kimura, H. The Possible Role of Hydrogen Sulfide as an Endogenous Neuromodulator. *J. Neurosci.* **1996**, 16 (3), 1066–1071.
- (7) Hosoki, R.; Matsuki, N.; Kimura, H. The Possible Role of Hydrogen Sulfide as an Endogenous Smooth Muscle Relaxant in Synergy with Nitric Oxide. *Biochemical and Biophysical Research Communications* **1997**, 237 (3), 527–531.
- (8) Moore, P. K.; Whiteman, M. *Chemistry, Biochemistry and Pharmacology of Hydrogen Sulfide*; 2015.
- (9) Zhang, H.; Tang, J.; Liu, X.-P.; Wang, Y.; Yu, W.; Peng, W.-Y.; Fang, F.; Ma, D.-F.; Wei, Z.-J.; Hu, L.-Y. Hydrogen Sulfide Promotes Root Organogenesis in *Ipomoea Batatas*, *Salix Matsudana* and *Glycine Max*. *Journal of Integrative Plant Biology* **2009**, 51 (12), 1086–1094.
- (10) Zhang, H.; Hu, L.-Y.; Hu, K.-D.; He, Y.-D.; Wang, S.-H.; Luo, J.-P. Hydrogen Sulfide Promotes Wheat Seed Germination and Alleviates Oxidative Damage against Copper Stress. *Journal of Integrative Plant Biology* **2008**, 50 (12), 1518–1529.
- (11) Jin, Z.; Pei, Y. Physiological Implications of Hydrogen Sulfide in Plants: Pleasant Exploration behind Its Unpleasant Odour. *Oxidative Medicine and Cellular Longevity* **2015**, 2015, 1–6.
- (12) Papenbrock, J.; Riemenschneider, A.; Kamp, A.; Schulz-Vogt, H. N.; Schmidt, A. Characterization of Cysteine-Degrading and H₂S-Releasing Enzymes of Higher Plants - From the Field to the Test Tube and Back. *Plant Biology* **2007**, 9 (5), 582–588.
- (13) Kabil, O.; Banerjee, R. Redox Biochemistry of Hydrogen Sulfide. *Journal of Biological Chemistry* **2010**, 285 (29), 21903–21907.
- (14) Kajimura, M.; Fukuda, R.; Bateman, R. M.; Yamamoto, T.; Suematsu, M. Interactions of Multiple Gas-Transducing Systems: Hallmarks and Uncertainties of CO, NO, and H₂S Gas Biology. *Antioxidants & Redox Signaling* **2010**, 13 (2), 157–192.
- (15) Li, Q.; Lancaster, J. R. Chemical Foundations of Hydrogen Sulfide Biology. *Nitric Oxide* **2013**, 35, 21–34.
- (16) Wever, R.; Van Gelder, B. F.; Dervartanian, D. V. Biochemical and Biophysical Studies on Cytochrome c Oxidase. XX. Reaction with Sulphide. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1975**, 387 (2), 189–193.

- (17) Collman, J. P.; Ghosh, S.; Dey, A.; Decreau, R. A. Using a Functional Enzyme Model to Understand the Chemistry behind Hydrogen Sulfide Induced Hibernation. *Proceedings of the National Academy of Sciences* **2009**, *106* (52), 22090–22095.
- (18) Blackstone, E. H₂S Induces a Suspended Animation-Like State in Mice. *Science* **2005**, *308* (5721), 518–518.
- (19) Leslie, M. Medicine. Nothing Rotten about Hydrogen Sulfide's Medical Promise. *Science* **2008**, *320* (5880), 1155–1157.
- (20) Chatfield, M. J.; La Mar, G. N.; Balch, A. L.; Lecomte, J. T. J. Multiple Forms of Sulfmyoglobin as Detected by ¹H Nuclear Magnetic Resonance Spectroscopy. *Biochemical and Biophysical Research Communications* **1986**, *135* (1), 309–315.
- (21) Szabó, C. Hydrogen Sulphide and Its Therapeutic Potential. *Nat Rev Drug Discov* **2007**, *6* (11), 917–935.
- (22) Vitvitsky, V.; Yadav, P. K.; Kurthen, A.; Banerjee, R. Sulfide Oxidation by a Noncanonical Pathway in Red Blood Cells Generates Thiosulfate and Polysulfides. *Journal of Biological Chemistry* **2015**, *290* (13), 8310–8320.
- (23) Kraus, D. W.; Wittenberg, J. B. Hemoglobins of the *Lucina Pectinata*/bacteria Symbiosis. I. Molecular Properties, Kinetics and Equilibria of Reactions with Ligands. *J. Biol. Chem.* **1990**, *265* (27), 16043–16053.
- (24) Pietri, R.; Lewis, A.; León, R. G.; Casabona, G.; Kiger, L.; Yeh, S.-R.; Fernandez-Alberti, S.; Marden, M. C.; Cadilla, C. L.; López-Garriga, J. Factors Controlling the Reactivity of Hydrogen Sulfide with Hemeproteins. *Biochemistry* **2009**, *48* (22), 4881–4894.
- (25) Bieza, S. A.; Boubeta, F.; Feis, A.; Smulevich, G.; Estrin, D. A.; Boechi, L.; Bari, S. E. Reactivity of Inorganic Sulfide Species toward a Heme Protein Model. *Inorganic Chemistry* **2015**, *54* (2), 527–533.
- (26) Nicoletti, F. P.; Comandini, A.; Bonamore, A.; Boechi, L.; Boubeta, F. M.; Feis, A.; Smulevich, G.; Boffi, A. Sulfide Binding Properties of Truncated Hemoglobins. *Biochemistry* **2010**, *49* (10), 2269–2278.
- (27) Bakan, A.; Kapralov, A. A.; Bayir, H.; Hu, F.; Kagan, V. E.; Bahar, I. Inhibition of Peroxidase Activity of Cytochrome c: De Novo Compound Discovery and Validation. *Molecular Pharmacology* **2015**, *88* (3), 421–427.
- (28) Vidossich, P.; Alfonso-Prieto, M.; Carpena, X.; Fita, I.; Loewen, P. C.; Rovira, C. The Dynamic Role of Distal Side Residues in Heme Hydroperoxidase Catalysis. Interplay between X-Ray Crystallography and Ab Initio MD Simulations. *Archives of Biochemistry and Biophysics* **2010**, *500* (1), 37–44.
- (29) Fernández-Fueyo, E.; Acebes, S.; Ruiz-Dueñas, F. J.; Martínez, M. J.; Romero, A.; Medrano, F. J.; Guallar, V.; Martínez, A. T. Structural Implications of the C-Terminal Tail in the Catalytic and Stability Properties of Manganese Peroxidases from Ligninolytic Fungi. *Acta Crystallographica Section D Biological Crystallography* **2014**, *70* (12), 3253–3265.
- (30) Ramirez, E.; Cruz, A.; Rodriguez, D.; Uchima, L.; Pietri, R.; Santana, A.; López-Garriga, J.; López, G. E. Effects of Active Site Mutations in Haemoglobin I from *Lucina Pectinata* : A Molecular Dynamic Study. *Molecular Simulation* **2008**, *34* (7), 715–725.
- (31) Campomanes, P.; Rothlisberger, U.; Alfonso-Prieto, M.; Rovira, C. The Molecular Mechanism of the Catalase-like Activity in Horseradish Peroxidase. *Journal of the American Chemical Society* **2015**, *137* (34), 11170–11178.
- (32) Fields, J. B.; Hollingsworth, S. A.; Chreifi, G.; Heyden, M.; Arce, A. P.; Magaña-Garcia, H. I.; Poulos, T. L.; Tobias, D. J. “Bind and Crawl” Association Mechanism of *Leishmania Major* Peroxidase and Cytochrome c Revealed by Brownian and Molecular Dynamics Simulations. *Biochemistry* **2015**, *54* (49), 7272–7282.
- (33) Navapour, L.; Mogharrab, N.; Amininasab, M. How Modification of Accessible Lysines to Phenylalanine Modulates the Structural and Functional Properties of Horseradish Peroxidase: A Simulation Study. *PLoS ONE* **2014**, *9* (10), e109062.
- (34) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E.

- Improved Side-Chain Torsion Potentials for the Amber ff99SB Protein Force Field. *Proteins: Structure, Function, and Bioinformatics* **2010**, NA-NA.
- (35) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, a Package of Computer Programs for Applying Molecular Mechanics, Normal Mode Analysis, Molecular Dynamics and Free Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Computer Physics Communications* **1995**, *91* (1–3), 1–41.
 - (36) Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Montgomery Jr, J.; Vreven, T.; Kudin, K.; Burant, J.; others. Gaussian 03, Revision A. 1. *Gaussian, Inc.: Pittsburgh, PA* **2003**.
 - (37) Soldatova, A. V.; Ibrahim, M.; Olson, J. S.; Czernuszewicz, R. S.; Spiro, T. G. New Light on NO Bonding in Fe(III) Heme Proteins from Resonance Raman Spectroscopy and DFT Modeling. *Journal of the American Chemical Society* **2010**, *132* (13), 4614–4625.
 - (38) Marti, M. A.; Crespo, A.; Capece, L.; Boechi, L.; Bikiel, D. E.; Scherlis, D. A.; Estrin, D. A. Dioxygen Affinity in Heme Proteins Investigated by Computer Simulation. *Journal of Inorganic Biochemistry* **2006**, *100* (4), 761–770.
 - (39) Capece, L.; Marti, M. A.; Crespo, A.; Doctorovich, F.; Estrin, D. A. Heme Protein Oxygen Affinity Regulation Exerted by Proximal Effects. *Journal of the American Chemical Society* **2006**, *128* (38), 12455–12461.
 - (40) Marti, M. A.; Capece, L.; Bidon-Chanal, A.; Crespo, A.; Guallar, V.; Luque, F. J.; Estrin, D. A. Nitric Oxide Reactivity with Globins as Investigated Through Computer Simulation. In *Methods in Enzymology*; Elsevier, 2008; Vol. 437, pp 477–498.
 - (41) Scherlis, D. A.; Estrin, D. A. Structure and Spin-State Energetics of an Iron Porphyrin Model: An Assessment of Theoretical Methods. *International Journal of Quantum Chemistry* **2002**, *87* (3), 158–166.
 - (42) Martí, M. A.; Capece, L.; Bikiel, D. E.; Falcone, B.; Estrin, D. A. Oxygen Affinity Controlled by Dynamical Distal Conformations: The Soybean Leghemoglobin and the Paramecium Caudatum Hemoglobin Cases. *Proteins: Structure, Function, and Bioinformatics* **2007**, *68* (2), 480–487.
 - (43) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Physical Review Letters* **1996**, *77* (18), 3865–3868.
 - (44) Nadra, A. D.; Martí, M. A.; Pesce, A.; Bolognesi, M.; Estrin, D. A. Exploring the Molecular Basis of Heme Coordination in Human Neuroglobin. *Proteins: Structure, Function, and Bioinformatics* **2008**, *71* (2), 695–705.
 - (45) Breneman, C. M.; Wiberg, K. B. Determining Atom-Centered Monopoles from Molecular Electrostatic Potentials. The Need for High Sampling Density in Formamide Conformational Analysis. *Journal of Computational Chemistry* **1990**, *11* (3), 361–373.
 - (46) Howes, B. D.; Giordano, D.; Boechi, L.; Russo, R.; Mucciacciaro, S.; Ciaccio, C.; Sinibaldi, F.; Fittipaldi, M.; Martí, M. A.; Estrin, D. A.; di Prisco, G.; Coletta, M.; Verde, C.; Smulevich, G. The Peculiar Heme Pocket of the 2/2 Hemoglobin of Cold-Adapted *Pseudoalteromonas Haloplanktis* TAC125. *JBIC Journal of Biological Inorganic Chemistry* **2011**, *16* (2), 299–311.
 - (47) Capece, L.; Lewis-Ballester, A.; Yeh, S.-R.; Estrin, D. A.; Marti, M. A. Complete Reaction Mechanism of Indoleamine 2,3-Dioxygenase as Revealed by QM/MM Simulations. *The Journal of Physical Chemistry B* **2012**, *116* (4), 1401–1413.
 - (48) Arroyo Mañez, P.; Lu, C.; Boechi, L.; Martí, M. A.; Shepherd, M.; Wilson, J. L.; Poole, R. K.; Luque, F. J.; Yeh, S.-R.; Estrin, D. A. Role of the Distal Hydrogen-Bonding Network in Regulating Oxygen Affinity in the Truncated Hemoglobin III from *Campylobacter Jejuni*. *Biochemistry* **2011**, *50* (19), 3946–3956.
 - (49) Cohen, J.; Arkhipov, A.; Braun, R.; Schulten, K. Imaging the Migration Pathways for O₂, CO, NO, and Xe Inside Myoglobin. *Biophysical Journal* **2006**, *91* (5), 1844–1857.
 - (50) Cohen, J.; Olsen, K. W.; Schulten, K. Finding Gas Migration Pathways in Proteins Using

- Implicit Ligand Sampling. In *Methods in Enzymology*; Elsevier, 2008; Vol. 437, pp 439–457.
- (51) Forti, F.; Boechi, L.; Bikiel, D.; Martí, M. A.; Nardini, M.; Bolognesi, M.; Viappiani, C.; Estrin, D.; Luque, F. J. Ligand Migration in Methanosarcina Acetivorans Protoglobin: Effects of Ligand Binding and Dimeric Assembly. *The Journal of Physical Chemistry B* **2011**, *115* (46), 13771–13780.
 - (52) Forti, F.; Boechi, L.; Estrin, D. A.; Martí, M. A. Comparing and Combining Implicit Ligand Sampling with Multiple Steered Molecular Dynamics to Study Ligand Migration Processes in Heme Proteins. *Journal of Computational Chemistry* **2011**, *32* (10), 2219–2231.
 - (53) Gabba, M.; Abbruzzetti, S.; Spyraakis, F.; Forti, F.; Bruno, S.; Mozzarelli, A.; Luque, F. J.; Viappiani, C.; Cozzini, P.; Nardini, M.; Germani, F.; Bolognesi, M.; Moens, L.; Dewilde, S. CO Rebinding Kinetics and Molecular Dynamics Simulations Highlight Dynamic Regulation of Internal Cavities in Human Cytochrome. *PLoS ONE* **2013**, *8* (1), e49770.
 - (54) Oliveira, A.; Singh, S.; Bidon-Chanal, A.; Forti, F.; Martí, M. A.; Boechi, L.; Estrin, D. A.; Dikshit, K. L.; Luque, F. J. Role of PheE15 Gate in Ligand Entry and Nitric Oxide Detoxification Function of Mycobacterium Tuberculosis Truncated Hemoglobin N. *PLoS ONE* **2012**, *7* (11), e49291.
 - (55) Isralewitz, B.; Gao, M.; Schulten, K. Steered Molecular Dynamics and Mechanical Functions of Proteins. *Current Opinion in Structural Biology* **2001**, *11* (2), 224–230.
 - (56) Jensen, M. O.; Park, S.; Tajkhorshid, E.; Schulten, K. Energetics of Glycerol Conduction through Aquaglyceroporin GlpF. *Proceedings of the National Academy of Sciences* **2002**, *99* (10), 6731–6736.
 - (57) Jarzynski, C. Nonequilibrium Equality for Free Energy Differences. *Physical Review Letters* **1997**, *78* (14), 2690–2693.
 - (58) Boechi, L.; Martí, M. A.; Milani, M.; Bolognesi, M.; Luque, F. J.; Estrin, D. A. Structural Determinants of Ligand Migration in Mycobacterium Tuberculosis Truncated Hemoglobin O. *Proteins: Structure, Function, and Bioinformatics* **2008**, *73* (2), 372–379.
 - (59) Bidon-Chanal, A.; Martí, M. A.; Crespo, A.; Milani, M.; Orozco, M.; Bolognesi, M.; Luque, F. J.; Estrin, D. A. Ligand-Induced Dynamical Regulation of NO Conversion in Mycobacterium Tuberculosis Truncated Hemoglobin-N. *Proteins: Structure, Function, and Bioinformatics* **2006**, *64* (2), 457–464.
 - (60) Gore, J.; Ritort, F.; Bustamante, C. Bias and Error in Estimates of Equilibrium Free-Energy Differences from Nonequilibrium Measurements. *Proceedings of the National Academy of Science U.S.A.* **2003**, *100* (22), 12564–12569.
 - (61) Di Lella, S.; Martí, M. A.; Álvarez, R. M. S.; Estrin, D. A.; Ricci, J. C. D. Characterization of the Galectin-1 Carbohydrate Recognition Domain in Terms of Solvent Occupancy. *The Journal of Physical Chemistry B* **2007**, *111* (25), 7360–7366.
 - (62) Gauto, D. F.; Di Lella, S.; Guardia, C. M. A.; Estrin, D. A.; Martí, M. A. Carbohydrate-Binding Proteins: Dissecting Ligand Structures through Solvent Environment Occupancy. *The Journal of Physical Chemistry B* **2009**, *113* (25), 8717–8724.
 - (63) Gauto, D. F.; Di Lella, S.; Estrin, D. A.; Monaco, H. L.; Martí, M. A. Structural Basis for Ligand Recognition in a Mushroom Lectin: Solvent Structure as Specificity Predictor. *Carbohydrate Research* **2011**, *346* (7), 939–948.
 - (64) López, E. D.; Arcon, J. P.; Gauto, D. F.; Petruk, A. A.; Modenutti, C. P.; Dumas, V. G.; Martí, M. A.; Turjanski, A. G. WATCLUST: A Tool for Improving the Design of Drugs Based on Protein-Water Interactions: Fig. 1. *Bioinformatics* **2015**, *31* (22), 3697–3699.
 - (65) Crespo, A.; Scherlis, D. A.; Martí, M. A.; Ordejón, P.; Roitberg, A. E.; Estrin, D. A. A DFT-Based QM-MM Approach Designed for the Treatment of Large Molecular Systems: Application to Chorismate Mutase. *The Journal of Physical Chemistry B* **2003**, *107* (49), 13728–13736.
 - (66) Arcon, J. P.; Rosi, P.; Petruk, A. A.; Martí, M. A.; Estrin, D. A. Molecular Mechanism of Myoglobin Autoxidation: Insights from Computer Simulations. *The Journal of Physical Chemistry B* **2015**, *119* (5), 1802–1813.

- (67) Choe, Y.-K.; Tsuchida, E.; Ikeshoji, T. First-Principles Molecular Dynamics Study on Aqueous Sulfuric Acid Solutions. *The Journal of Chemical Physics* **2007**, *126* (15), 154510.
- (68) Hu, X. L.; Klimeš, J.; Michaelides, A. Proton Transfer in Adsorbed Water Dimers. *Physical Chemistry Chemical Physics* **2010**, *12* (16), 3953.
- (69) McDonnell, M. T.; Xu, H.; Keffer, D. J. Ab Initio Molecular Dynamics Simulations of an Excess Proton in a Triethylene Glycol-Water Solution: Solvation Structure, Mechanism, and Kinetics. *Journal Physical Chemistry B* **2016**, *120* (23), 5223–5242.
- (70) Zhang, H.; Ye, Y.-K.; Wang, S.-H.; Luo, J.-P.; Tang, J.; Ma, D.-F. Hydrogen Sulfide Counteracts Chlorophyll Loss in Sweetpotato Seedling Leaves and Alleviates Oxidative Damage against Osmotic Stress. *Plant Growth Regulation* **2009**, *58* (3), 243–250.
- (71) Shaik, S.; Kumar, D.; de Visser, S. P.; Altun, A.; Thiel, W. Theoretical Perspective on the Structure and Mechanism of Cytochrome P450 Enzymes. *Chemical Reviews* **2005**, *105* (6), 2279–2328.
- (72) Senn, H. M.; Thiel, W. QM/MM Studies of Enzymes. *Current Opinion in Chemical Biology* **2007**, *11* (2), 182–187.
- (73) Ranaghan, K. E.; Mulholland, A. J. Investigations of Enzyme-Catalysed Reactions with Combined Quantum Mechanics/molecular Mechanics (QM/MM) Methods. *International Reviews in Physical Chemistry* **2010**, *29* (1), 65–133.
- (74) Martí, M. A.; Crespo, A.; Bari, S. E.; Doctorovich, F. A.; Estrin, D. A. QM–MM Study of Nitrite Reduction by Nitrite Reductase of *Pseudomonas Aeruginosa*. *The Journal of Physical Chemistry B* **2004**, *108* (46), 18073–18080.
- (75) Martí, M. A.; Capece, L.; Crespo, A.; Doctorovich, F.; Estrin, D. A. Nitric Oxide Interaction with Cytochrome C' and Its Relevance to Guanylate Cyclase. Why Does the Iron Histidine Bond Break? *Journal of the American Chemical Society* **2005**, *127* (21), 7721–7728.
- (76) Crespo, A.; Martí, M. A.; Kalko, S. G.; Morreale, A.; Orozco, M.; Gelpi, J. L.; Luque, F. J.; Estrin, D. A. Theoretical Study of the Truncated Hemoglobin HbN: Exploring the Molecular Basis of the NO Detoxification Mechanism. *Journal of the American Chemical Society* **2005**, *127* (12), 4433–4444.
- (77) González Lebrero, M. C.; Estrin, D. A. QM-MM Investigation of the Reaction of Peroxynitrite with Carbon Dioxide in Water. *Journal of Chemical Theory and Computation* **2007**, *3* (4), 1405–1411.
- (78) Zeida, A.; Babbush, R.; González Lebrero, M. C.; Trujillo, M.; Radi, R.; Estrin, D. A. Molecular Basis of the Mechanism of Thiol Oxidation by Hydrogen Peroxide in Aqueous Solution: Challenging the S_N2 Paradigm. *Chemical Research in Toxicology* **2012**, *25* (3), 741–746.
- (79) Nitsche, M. A.; Ferreria, M.; Mocskos, E. E.; González Lebrero, M. C. GPU Accelerated Implementation of Density Functional Theory for Hybrid QM/MM Simulations. *Journal of Chemical Theory and Computation* **2014**, *10* (3), 959–967.
- (80) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics* **1996**, *14* (1), 33–38.
- (81) Bustamante, J. P.; Abbruzzetti, S.; Marcelli, A.; Gauto, D.; Boechi, L.; Bonamore, A.; Boffi, A.; Bruno, S.; Feis, A.; Foggi, P.; Estrin, D. A.; Viappiani, C. Ligand Uptake Modulation by Internal Water Molecules and Hydrophobic Cavities in Hemoglobins. *The Journal of Physical Chemistry B* **2014**, *118* (5), 1234–1245.
- (82) Goldbeck, R. A.; Bhaskaran, S.; Ortega, C.; Mendoza, J. L.; Olson, J. S.; Soman, J.; Kliger, D. S.; Esquerra, R. M. Water and Ligand Entry in Myoglobin: Assessing the Speed and Extent of Heme Pocket Hydration after CO Photodissociation. *Proceedings of the National Academy of Sciences* **2006**, *103* (5), 1254–1259.
- (83) Ouellet, Y. H.; Daigle, R.; Lague, P.; Dantsker, D.; Milani, M.; Bolognesi, M.; Friedman, J. M.; Guertin, M. Ligand Binding to Truncated Hemoglobin N from Mycobacterium Tuberculosis Is Strongly Modulated by the Interplay between the Distal Heme Pocket Residues and Internal Water. *Journal of Biological Chemistry* **2008**, *283* (40), 27270–

- 27278.
- (84) Bustamante, J. P.; Szretter, M. E.; Sued, M.; Martí, M. A.; Estrin, D. A.; Boechi, L. A Quantitative Model for Oxygen Uptake and Release in a Family of Hemeproteins. *Bioinformatics* **2016**, btw083.
 - (85) Boechi, L.; Arrar, M.; Martí, M. A.; Olson, J. S.; Roitberg, A. E.; Estrin, D. A. Hydrophobic Effect Drives Oxygen Uptake in Myoglobin via Histidine E7. *Journal of Biological Chemistry* **2013**, 288 (9), 6754–6762.
 - (86) Scott, E. E.; Gibson, Q. H.; Olson, J. S. Mapping the Pathways for O₂ Entry Into and Exit from Myoglobin. *Journal of Biological Chemistry* **2001**, 276 (7), 5177–5188.
 - (87) Boechi, L.; Mañez, P. A.; Luque, F. J.; Martí, M. A.; Estrin, D. A. Unraveling the Molecular Basis for Ligand Binding in Truncated Hemoglobins: The trHbO *Bacillus Subtilis* Case. *Proteins: Structure, Function, and Bioinformatics* **2010**, 78 (4), 962–970.
 - (88) Capece, L.; Boechi, L.; Perissinotti, L. L.; Arroyo-Mañez, P.; Bikiel, D. E.; Smulevich, G.; Martí, M. A.; Estrin, D. A. Small Ligand–globin Interactions: Reviewing Lessons Derived from Computer Simulation. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2013**, 1834 (9), 1722–1738.
 - (89) Liphardt, J. Equilibrium Information from Nonequilibrium Measurements in an Experimental Test of Jarzynski's Equality. *Science* **2002**, 296 (5574), 1832–1835.
 - (90) Marques, H. M.; Perry, C. B. Hemepeptide Models for Hemoproteins: The Behavior of N-Acetylmicroperoxidase-11 in Aqueous Solution. *Journal of Inorganic Biochemistry* **1999**, 75 (4), 281–291.
 - (91) Brunori, M.; Amiconi, G.; Antonini, E.; Wyman, J.; Zito, R.; Rossi Fanelli, A. The Transition between “acid” and “alkaline” Ferric Heme Proteins. *Biochimica et Biophysica Acta (BBA) - Protein Structure* **1968**, 154 (2), 315–322.
 - (92) Sun, J.; Wilks, A.; Ortiz de Montellano, P. R.; Loehr, T. M. Resonance Raman and EPR Spectroscopic Studies on Heme-Heme Oxygenase Complexes. *Biochemistry* **1993**, 32 (51), 14151–14157.
 - (93) Takahashi, S.; Wang, J.; Rousseau, D. L.; Ishikawa, K.; Yoshida, T.; Host, J. R.; Ikeda-Saito, M. Heme-Heme Oxygenase Complex. Structure of the Catalytic Site and Its Implication for Oxygen Activation. *Journal of Biological Chemistry* **1994**, 269 (2), 1010–1014.
 - (94) Song, Y.; Michonova-Alexova, E.; Gunner, M. R. Calculated Proton Uptake on Anaerobic Reduction of Cytochrome C Oxidase: Is the Reaction Electroneutral? *Biochemistry* **2006**, 45 (26), 7959–7975.

V. EFECTO DE LA TEMPERATURA EN LA MIGRACIÓN Y UNIÓN DE LIGANDOS A HEMOPROTEÍNAS

V.1 Introducción

En la naturaleza existen organismos adaptados a diferentes temperaturas. Estos organismos pueden ser clasificados en “psicrófilos” de crecimiento óptimo entre 268 K y 285 K, “termófilos”, con temperaturas de crecimiento óptimo entre 333 K y 353 K; “hipertermófilos”, con crecimientos óptimos entre 353 K y 383 K, finalmente se encuentran los llamados “mesófilos”, también denominados “no-adaptados”, cuyos crecimientos óptimos se encuentran entre 298 K y 323 K.^[1]

Como se mencionó en la introducción de la presente tesis, la temperatura es un factor crucial que determina el éxito de todos los organismos. La capacidad de supervivencia estará determinada por la susceptibilidad de los procesos bioquímicos de los organismos ante cambios en la temperatura. En particular, un conjunto de procesos que se verá afectado al cambio de temperatura es la función de las proteínas.

Una manera de abordar estudios temperatura dependientes de diversos procesos físico-químicos en proteínas es observando el comportamiento de algún parámetro medido en función de modelos termodinámicos (como Arrhenius o Eyring, entre otros) en un amplio rango de temperaturas.^[2–9]

En el caso particular del modelo de Eyring (o su versión linealizada), las desviaciones se evidencian como disrupciones marcadas en los gráficos de constante de velocidad en función de la temperatura.^[10–15]

En el presente capítulo se estudiará cómo afecta la temperatura a la migración y posterior unión de CO a hemoproteínas adaptadas y no adaptadas a la temperatura, a través de mediciones de Flash Fotólisis usando como ligando CO, que es una sonda ampliamente usada por sus características experimentales.

Se estudiarán tres hemoglobinas truncadas (trHb): una trHb perteneciente al microorganismo termófilo *Thermobifida fusca* (Tf-trHbO), una trHb perteneciente al organismo mesófilo *Basilus subtilis* (Bs-trHbO) y una trHb perteneciente a un organismo psicrófilo, *Pseudoalteromonas haloplanktis* (Ph-trHbO).

Todas las mediciones realizadas en este capítulo fueron realizadas en el laboratorio del Prof. Cristiano Viappiani de la Universidad de los Estudios de Parma, Italia

V.2 Metodología específica

Para medir experimentalmente las constantes cinéticas de unión del ligando CO (k_{on}) de las hemoproteínas mencionadas (Ph-TrHbO, Bs-TrHbO y Tf-TrHbO), se utilizó la técnica de Láser Flash Fotólisis, LFP.

Como se mencionó en la introducción, esta técnica se basa en partir de soluciones en donde se encuentra formado complejo hexacoordinado y mediante una fotólisis del complejo, generar el complejo pentacoordinado, y observar la evolución temporal de la absorbancia del complejo (o la diferencia entre el complejo pentacoordinado y hexacoordinado, o equivalentemente la diferencia en absorbancia entre tiempos después y antes del destello del láser).

De esta manera, a lo largo del tiempo se re-unirá el ligando hasta restablecer la cantidad inicial del complejo hexacoordinado.

Finalmente, la cinética de interés se obtiene observando las distintas escalas de tiempo que el sistema presenta para volver al estado hexacoordinado inicial.

Las escalas temporales asociadas a los eventos de re-unión que se pueden observar mediante esta técnica rondan los 100ns a 0.1s.

Las mediciones de LFP se realizaron utilizando soluciones stock original de las proteínas diluidas con buffer fosfato sódico de concentración 100 mM (pH=7.00), 1 mM de EDTA, y 2 mM de ditionito de sodio para evitar procesos oxidativos. Todas las soluciones utilizadas fueron previamente desgasadas con corriente de argón. Con este procedimiento, el valor de concentraciones finales de las proteínas fue dentro del rango de 20 a 60 μ M.

Las soluciones stock originales de la proteínas fueron provistas por el grupo de investigación a cargo del Dr Stefano Bruno del Departamento de Ciencia de los Alimentos y de

Farmacia de la Universidad del Estudio de Parma (muestra de Bs) y la Dra Cinzia Verde del Instituto de Biociencias y Biorecursos de Nápoles (muestras de Tf y Ph)

Antes de realizados los experimentos, las soluciones fueron equilibradas con CO (o mezcla de nitrógeno-CO) de presión parcial conocida, durante aproximadamente 30 minutos.

La fotólisis fue lograda con un láser de Nd:YAG y los cambios de absorbancia fueron monitoreados a 436 nm, esta longitud de onda seleccionada maximiza el valor de la diferencia de absorbancia entre el complejo hexacoordinado y pentacoordinado durante la cinética.

La muestra fue termostatzada utilizando un elemento de Peltier (Flash1000, Quantum Northwest, Inc.), permitiendo fijar la temperatura durante todo el experimento, con una precisión de 0,1 °C.

Para observar más detalladamente la cinética de los distintos eventos temporales asociados a la cinética de re-unión, se varió la escala temporal para la observación de la señal de salida del osciloscopio, se tomaron escalas de 1 μ s, 100 μ s y 10 ms por división, permitiendo así poder observar valores de señal cada 25 ps, 20 ns y 2 μ s respectivamente (ver más abajo).

En general, para cada uno de los tres distintos intervalos de tiempo de observación del experimento (determinados por la escala temporal utilizada en el osciloscopio) se tomaron 200 trazas a una repetición de 0.5 Hz, que se utilizaron para obtener una traza promedio.

Por otro lado, se realizó el mismo procedimiento pero bloqueando la incidencia del láser sobre la muestra (shutter apagado, no fotólisis), estas condiciones permiten determinar una línea de base que tenga en cuenta medidas de absorbancia debido a la luz espuria (proveniente de la lámpara, del láser, etc). Posteriormente, como se verá, a la señal del experimento de fotólisis se corrigen teniendo en cuenta la línea de base.

Los valores de voltaje medidos en el detector, registran entonces valores de voltaje debido a cambios en la transmitancia del haz de luz sonda que atraviesa la muestra antes ($V(t < t_0)$) y después ($V(t)$) de la excitación del láser.

Teniendo en cuenta la relación entre la absorbancia y la transmitancia se obtiene entonces:

$$\Delta A(t) = A(5c) - A(6c) = -\log \frac{T(t)}{T(t < t_0)} = -\log \frac{I(t)/I(0)}{I(t < t_0)/I(0)} \quad (V.I)$$

Donde t_0 es el tiempo en el cual se produce la excitación con el láser. $I(0)$ es la intensidad de luz antes de la muestra, $I(t < t_0)$ y $I(t)$ son las intensidades de luz al pasar por la muestra antes y después del destello del láser, respectivamente.

Finalmente, suponiendo una relación de proporcionalidad entre la intensidad de luz y el voltaje a la salida del detector (PMT) resulta equivalente entonces evaluar $\Delta A(t)$ en función de los voltajes antes y después del destello del láser.

$$A(5c) - A(6c) = \Delta A(t) = -\log \frac{V(t)}{V(t < t_0)} \quad (V.2)$$

Para obtener las trazas finales asociadas a variaciones temporales de ΔA se procedió de la siguiente manera:

En primer lugar se corrigió la medida temporal de voltaje (con incidencia pulsada del láser) restando la línea de base (sin incidencia del láser) punto a punto.

La Figura 1 muestra las salidas típicas que se obtienen al medir en este tipo de experimentos, tómese por ejemplo la cinética de re-unión de CO en Bs-TrHbO a 30 °C.

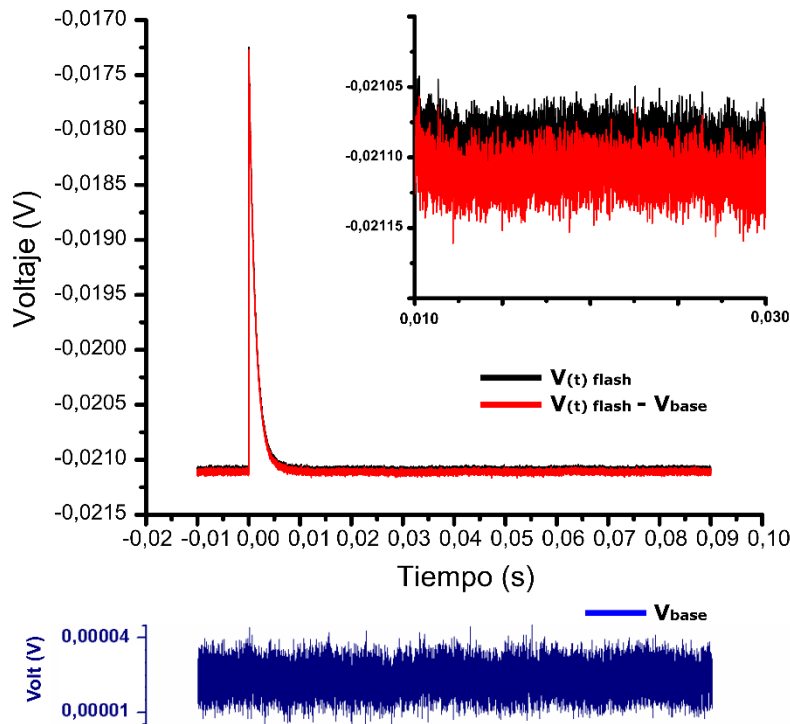


Figura 1: voltajes promedio medidos en el osciloscopio (200 trazas). Panel superior: línea negra, evolución temporal del voltaje sin corrección por la línea de base ($V(t)_{flash}$, negro) y corrigiendo por línea de base ($V(t)_{flash} - V_{base}$, rojo). Se muestra también una ampliación de una región de este gráfico en el borde superior-derecho. Panel inferior: voltaje de línea de base

Esta figura muestra el voltaje de salida en función del tiempo para una escala de 10 ms. La línea negra ($V(t)_{flash}$) y azul (V_{base}) muestran el voltaje promedio obtenido a partir de

aproximadamente 200 trazas medidas bajo condiciones con y sin incidencia del láser, respectivamente.

La línea roja muestra el voltaje corregido ($V(t)_{flash}-V_{base}$).

Como puede observarse en la Figura 1, el destello es realizado 0.01 segundos después del tiempo inicial de observación. Por lo tanto, los valores de voltaje correspondiente a tiempos entre -0.01 segundos y 0.00 corresponderán a valores de voltaje asociados al complejo en el estado hexacoordinado ($V(t < t_0)$). Asimismo, los valores de voltaje correspondientes a tiempos mayores a 0 corresponderán asociados al experimento en condiciones de fotólisis ($V(t)$), en donde el destello produce un súbito aumento de proteína en estado pentacoordinado, aumentando súbitamente el valor de $\Delta A(t)$.

Finalmente, el proceso de re-unión de CO con la proteína, reestablece a tiempos largos (aproximadamente a 0.01 segundos en este ejemplo) el voltaje observado al valor obtenido antes del destello.

Se corrigen todos los valores de absorbancia por la línea de base y a partir de un promedio de los valores de voltaje sobre la parte inicial se obtiene $V(t < t_0)$.

Finalmente con los valores de voltaje después del destello del láser $V(t)$, se calcula $\Delta A(t)$ a partir de la ecuación (V.2). Todo el procedimiento antes descrito, se realiza para cada escala temporal de medición. Obteniendo finalmente los valores de $\Delta A(t)$ como se muestra en la Figura 2.

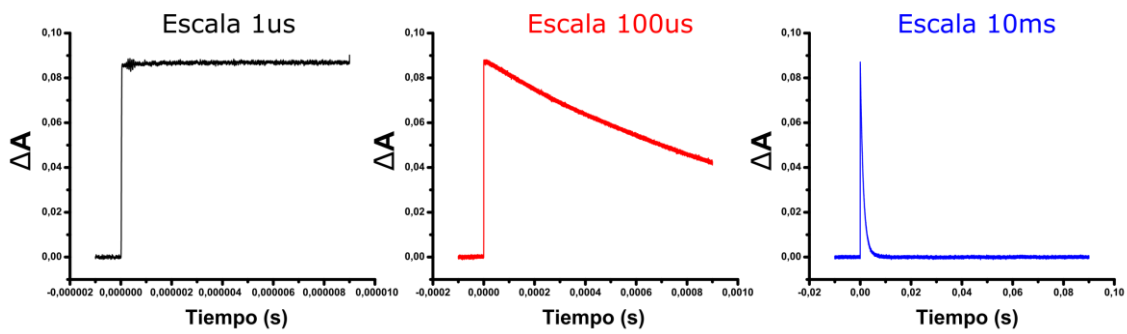


Figura 2: valores de ΔA en función del tiempo para tres escalas de medición temporal (1 μs , 100 μs , 10 ms de izquierda a derecha respectivamente)

Como puede observarse en la Figura 2, el tiempo total de medición queda determinado por la escala utilizada en el osciloscopio: el instrumento presenta en su display 10 segmentos para la medición temporal y 8 segmentos para la medición de voltaje (ver Figura 3).

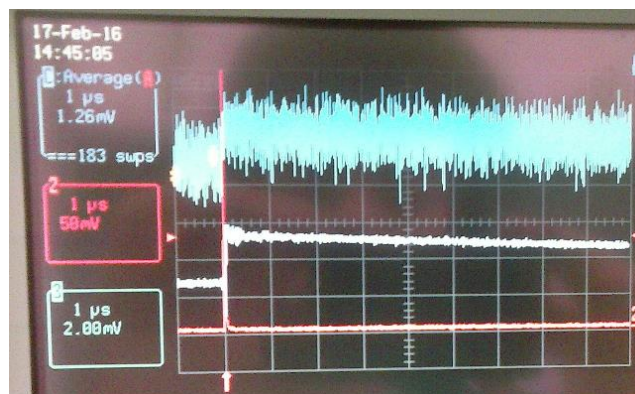


Figura 3: display del osciloscopio utilizado en las medidas LFP. Presenta 10 segmentos para el eje temporal (horizontal) y 8 para el eje de voltaje (vertical). En este caso se muestran tres trazas temporales: traza superior: la medida de voltaje instantáneo antes y después del destello del láser. Traza del medio: el voltaje temporal promedio (en este caso de 183 disparos, $V(t)_{flash}$). Traza inferior: el voltaje asociado al láser

Por ejemplo, una escala de 1 μ s por segmento (como el de la Figura 3), implicará que el tiempo total de observación será 10 μ s.

Por otro lado, como se mencionó en la sección de métodos, la resolución del equipo (a través del detector y osciloscopio) restringe la resolución temporal en cada bloque, en otras palabras, cada cuanto tiempo se mide un valor de voltaje: una menor escala tendrá menor tiempo total medido y una mayor resolución temporal entre dos mediciones sucesivas de voltaje. Mientras que para escalas temporales mayores, el tiempo total de medición será mayor pero la resolución temporal será menor.

Asimismo, resulta necesario cambiar la escala de medición ya que mayores tiempos totales de medición permitirán una observación más completa de la cinética (como puede observarse en la Figura 2)

Para lograr la mejor resolución temporal y tener información temporal en el mayor rango posible, se suele entonces realizar la unión de todos los datos de las escalas medidas, conservando los datos tengan mayor resolución, sin solaparlos con los de otra escala menos resolutive.

De esta forma, a las mediciones de la primera escala se le unen las mediciones de la segunda (descartando los valores temporales de voltaje de la segunda escala que se repetían con los de la primera). El mismo procedimiento se realizó con la tercera escala: se unieron a los datos de la segunda escala, descartando los datos de la tercera escala que se repetían con los de la segunda.

Este procedimiento dará como resultado una traza temporal en todo el rango de medición observado (hasta ~ 0.1 ms). La Figura 4A muestra la traza (suavizada) para la medida

cinética de LFP para la re-unión de CO en Bs-TrHbO a $T=30\text{ }^{\circ}\text{C}$. No se muestran los valores de voltaje correspondientes a tiempos anteriores al destello, también se muestra la traza en escala temporal logarítmica (Figura 4B) que permite una mejor visualización, ya que permite aumentar las diferencias entre valores de absorbancia.

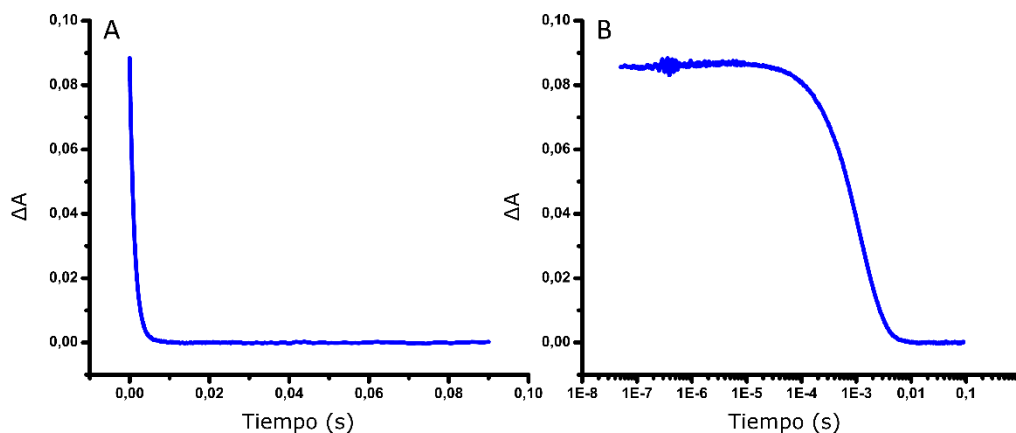


Figura 4: (A) Traza temporal suavizada para la re-unión de CO en Bs-TrHbO a $T=30\text{ }^{\circ}\text{C}$. (B) La misma traza temporal en escala temporal logarítmica. En ambos casos, no se muestran los instantes antes del destello del láser.

V.3 Resultados

V.3.1 Evolución temporal de las trazas de re-unión de ligando CO

A continuación se muestran los resultados para las cinéticas de migración de CO para las tres hemoproteínas estudiadas Ph-TrHbO, Bs-TrHbO y Tf-TrHbO utilizando la técnica de LFP. Las mediciones fueron obtenidas a distintas temperaturas, equilibrando el sistema inicialmente con una presión parcial de CO de 1 atm.

Todas las medidas fueron tratadas como se describió en la sección anterior. La Figura 5 muestra los tres conjuntos de trazas para las tres proteínas estudiadas.

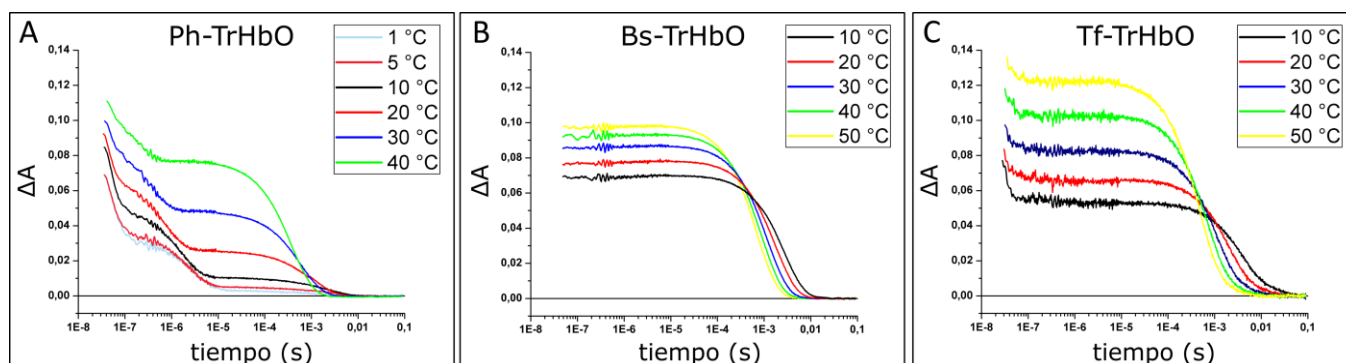


Figura 5: Trazas temporales para las cinéticas de unión de CO (1 atm) para las tres proteínas (Ph-TrHbO, Bs-TrHbO y Tf-TrHbO; paneles A, B y C, respectivamente), obtenidas a diferente temperatura.

El rango total de tiempo de medición de la técnica permite la observación detallada de eventos de re-unión con tiempos compatibles con moléculas de CO que se re-unen desde cavidades internas de la proteína (~ 10 - 100 ns hasta ~ 0.1 ms), así como también a procesos de re-unión de ligando con tiempos compatibles a poblaciones de CO para las cuales se encuentra fuera de la matriz proteica, es decir en el solvente (tiempos del orden ~ 0.1 ms hasta ~ 0.1 s).^[16-19]

Por otra parte, los resultados de la Figura 5 muestran que, para todas las proteínas, las trazas se comportan de manera similar frente a cambios de temperatura: cuando la temperatura es incrementada, se observa, en primer lugar que el valor inicial de ΔA (interpretada como la diferencia entre la cantidad de complejo pentacoordinado y hexacoordinado) aumenta, debido a que la población complejo pentacoordinado al tiempo inicial de medida del experimento se incrementa.

Resulta importante remarcar es que en estos experimentos y dadas las condiciones experimentales (potencia del láser, concentraciones de la muestra, etc) se considera que el total de moléculas de complejo se fotolizan. Además, la mayor fracción de moléculas de CO que se fotolizaron se re-únen inmediatamente después del destello del láser en tiempos que rondan las decenas de femtosegundos (se dice entonces que las moléculas de CO se recombinan geminalmente) por lo cual estos eventos de recombinación no pueden ser observados mediante la técnica de LFP. Se estima que este porcentaje de recombinación geminal ronda alrededor del 95%.^[20]

Teniendo en cuenta lo anterior y considerando los tiempos iniciales de observación (del orden de los ns después de la fotólisis), se puede concluir que al aumentar la temperatura hay una mayor cantidad de moléculas de ligando que pudieron superar las barreras energéticas para la recombinación geminal rápida. Generando así, una mayor población inicial de ligando no unido, y por lo tanto una mayor cantidad de proteína en estado pentacoordinado, dando como

resultado valores iniciales mayores de ΔA . Evidenciando entonces que las constantes asociadas a la salida del ligando se afectan más con la temperatura que las constantes de entrada.

Por otro lado, el sistema también presentará una población asociada a estados en donde existen moléculas de CO en el seno del solvente con la proteína en estado pentacoordinado, cuyos tiempos asociados para sus procesos de re-unión serán mayores (tiempos del orden de ~ 0.1 ms hasta $\sim 0,1$ s, como se mencionó).

En estos casos, al aumentar la temperatura se observa una disminución en el tiempo necesario para obtener valores de $\Delta A = 0$ (cuando la población de complejo hexacoordinado es igual a la del pentacoordinado).

Es importante remarcar, que estos tiempos de medición son compatibles también con que exista una proporción del complejo hexacoordinado (producto de la recombinación geminal descrita anteriormente).

El equilibrio dinámico del sistema permite que sucedan simultáneamente los dos procesos asociados a la unión/disociación del ligando: por un lado, la formación del estado hexacoordinado (con velocidad asociada a la constante de unión de ligando k_{on}) a partir de las moléculas de CO y el complejo pentacoordinado formados y por el otro lado, la disociación de las moléculas que se re-unieron formando el complejo hexacoordinado hacia el solvente (con constante de velocidad asociada a la disociación del ligando, k_{off}). Ambos procesos se producirán con mayor velocidad al aumentar la temperatura como lo predice el modelo de Eyring.

Sin embargo, los resultados muestran que al aumentar la temperatura hay una mayor capacidad de las molécula de ligando en migrar hacia el sitio activo que viceversa. El aumento de temperatura permite que las moléculas de CO migren con mayor velocidad y puedan superar más fácilmente las barreras asociadas para la unión que para la disociación, generando netamente que se alcancen valores de $\Delta A = 0$ a menores tiempos de medición.

En términos termodinámicos, lo anterior implicaría que la afinidad del ligando CO con la temperatura aumentaría, evidenciada con un aumento en K_d .

V.3.2 Obtención de las constantes cinéticas de asociación (k_{on}) de CO.

Las velocidades de unión de ligando, como se mencionó en la introducción, se caracterizan por ser procesos bimoleculares con una ley de velocidad de segundo orden global (orden 1 en proteína y orden 1 en ligando).

A partir de lo anterior, un cambio en la concentración de CO, puede evidenciar en la traza total medida la zona en donde la reacción procede dependiente de CO, y por lo tanto es posible asignar esa región como la traza correspondiente al proceso de unión bimolecular de ligando CO desde el solvente.

La Figura 6 muestra los experimentos de LFP equilibrando el sistema a una presión parcial de CO 1 atm y los realizados equilibrando el sistema a una presión parcial de CO de 0.1 atm.

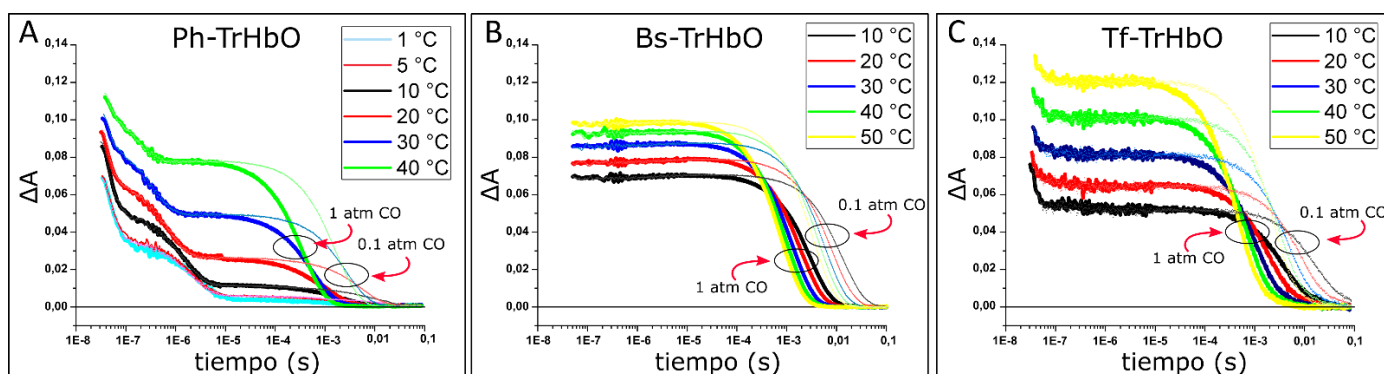


Figura 6: Superposición de las trazas temporales para las cinéticas de unión de CO equilibrado a 1 atm (línea gruesa) y a 0.1 atm (línea fina), para las tres proteínas (Ph-TrHbO, Bs-TrHbO y Tf-TrHbO; paneles A, B y C, respectivamente), obtenidas a diferente temperatura.

Se observa en todos los casos que para el sistema equilibrado a una presión de 0.1 atm de CO, la parte inicial (hasta 10^{-5} segundos) de la traza es similar a las medidas a 1 atm de CO, mientras que los tiempos de re-unión se incrementan en la parte final de las trazas, evidenciando entonces que la reacción se enlentece.

Este comportamiento se presenta para todas las proteínas y para todas las temperaturas. Se puede entonces identificar de esta manera la zona de la traza en donde la reacción ocurre dependiente de la concentración de CO (10^{-5} segundos en adelante). Esta región entonces caracteriza tiempos de medición en donde el proceso de unión de CO es desde el solvente y de forma bimolecular con la proteína.

Se observa también en la Figura 6 el intervalo de tiempo seleccionado para la reacción bimolecular se corresponde al proceso bimolecular para todas las proteínas y las temperaturas.

Para obtener los valores de las constantes de asociación k_{on} para la unión de CO, se utilizaron entonces las trazas obtenidas utilizando una presión parcial de 1 atm de CO, y se seleccionó la zona donde la reacción procede dependiente de la concentración de CO.

La Figura 7 muestra un ejemplo de lo explicado anteriormente. Como se puede observar, existe una zona en donde la reacción procede temporalmente con valores de ΔA mayores a los obtenidos con presión parcial de 1 atm (curvas negra y roja respectivamente).

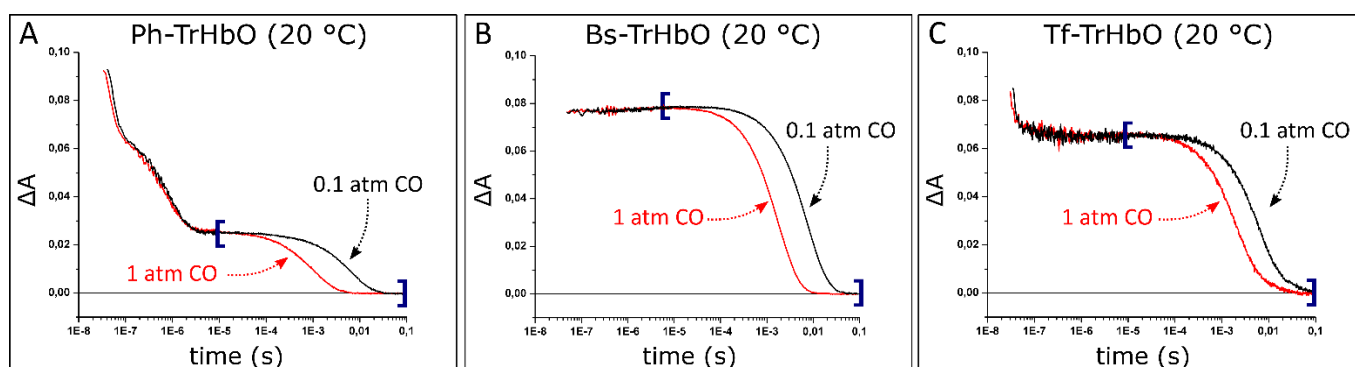


Figura 7: cinéticas de re-unión de CO para las tres hemoglobinas truncadas (Ph-TrHbO, Bs-TrHbO y Tf-TrHbO paneles A, B y C, respectivamente), a $T=20\text{ }^{\circ}\text{C}$ obtenidas a diferentes presiones parciales de CO (1 atm CO y 0.1 atm CO, curvas rojas y negras, respectivamente)

A partir de la zona en donde la reacción procede de manera bimolecular (zona entre corchetes en los paneles A, B y C de la Figura 7) y se puede obtener a partir de un ajuste exponencial las constantes de asociación bimolecular de CO para las tres proteínas y para cada temperatura.

En todos los casos considerados, las constantes determinadas por el ajuste corresponden a constantes aparentes de pseudo-orden de ligando CO ($k_{app} = k_{on} \cdot [CO]$ con $[CO]$ la concentración de CO disuelto en la solución dada una temperatura). La condición de pseudo-orden se cumple cuando la concentración de un reactivo es 100 veces mayor al otro.

Lo anterior se puede explicar de la siguiente manera: como la concentración de la proteína es del orden de μM y la concentración de CO es del orden de mM (considerando la solubilidad de CO en solución acuosa a una presión parcial de 1 atm en función de la temperatura), la concentración de CO libre puede considerarse constante a lo largo de todo el proceso. Además como se mencionó en la introducción, la relación entre $k_{on} \cdot [CO]$ y k_{off} permite asumir que durante el proceso de re-unión se pueden despreciar el aporte en concentración del proceso de disociación.

Lo anterior permite concluir que obtenidas las constantes del ajuste exponencial de las trazas temporales que corresponden al proceso bimolecular, se puede dividir el valor por la concentración de CO a cada temperatura (dada por su solubilidad) y obtener entonces el valor de k_{on} .

Es importante remarcar, que cuando el sistema es equilibrado con 0.1 atm CO la solubilidad de este gas a esa presión parcial determina que el sistema no cumpla con la condición de pseudo-orden. Es por esto, que el experimento realizado con 0.1 atm de CO solo permitió identificar la región temporal a partir de la cual la reacción procede de forma bimolecular.

V.3.3 Mediciones de k_{on} en función de la temperatura y el tratamiento de Eyring.

Se ajustó la zona de las trazas temporales asignadas como las pertenecientes a la reacción bimolecular de unión de CO. Para el ajuste se utilizó, para todas las temperaturas medidas, una ecuación cuya forma funcional corresponde a la suma de dos exponenciales, como muestra la ecuación (V.3)

$$y = y_0 + A_1 \cdot e^{-\frac{(x-2 \cdot 10^{-8})}{t_1}} + A_2 \cdot e^{-\frac{(x-2 \cdot 10^{-8})}{t_2}} \quad (V.3)$$

En donde se consideró como centro de cada exponencial el valor $2 \cdot 10^{-8}$ (en unidades de segundos). Este valor corresponde a donde se inician en el eje x, los datos medidos. En esta ecuación los parámetros ajustables son y_0 , A_1 , A_2 , t_1 y t_2 .

Finalmente, con los valores obtenidos para los tiempos de decaimiento (t_1 y t_2), y considerando la solubilidad del ligando (a presiones parcial de 1 atm y a diferentes temperaturas), se obtuvieron los valores de k_{on} . Por otro lado, a partir de las amplitudes (A_1 y A_2) asociados a ambos tiempos de decaimientos, es posible obtener las porcentajes de la porción de la traza analizada, descriptas por un tiempo u otro. Esto permite establecer las posibles conformaciones o microestados del sistema para el cual el ligando se une con un tiempo característico u otro.

Los resultados de k_{on} para las tres proteínas considerando las incertezas asociadas al ajuste, se muestran a continuación en la Tabla 1

I

<i>Ph-TrHbO (1 atm CO)</i>						
Temp (K)	kon ₁	Δkon ₁	kon ₂	Δkon ₂	F1 (%)	F2 (%)
274	1,4E+05	4E+01	6,4E+05	3E+02	53,5	46,5
278	1,8E+05	6E+00	2,3E+06	5E+02	79,5	20,5
283	2,7E+05	7E+00	1,4E+06	2E+02	79,6	20,4
293	7,2E+05	1E+01	2,7E+06	2E+02	78,8	21,2
303	1,6E+06	1E+01	5,7E+06	3E+02	89,5	10,5
313	2,8E+06	2E+01	8,0E+06	4E+02	89,1	10,9

II

<i>Bs-TrHbO (1 atm CO)</i>		
Temp (K)	kon ₁	Δkon ₁
283	2,8E+05	4E-01
293	5,1E+05	5E-01
303	8,8E+05	7E-01
313	1,3E+06	1E+00
323	1,8E+06	1E+00

III

<i>Tf-TrHbO (1 atm CO)</i>						
Temp (K)	kon ₁	Δkon ₁	kon ₂	Δkon ₂	F1 (%)	F2 (%)
283	2,2E+05	4E+00	4,4E+04	4E+00	78,3	21,7
293	4,9E+05	4E+00	8,4E+04	5E+00	84,3	15,7
303	1,0E+06	7E+00	1,7E+05	8E+00	87,2	12,8
313	1,8E+06	9E+00	3,6E+05	1E+01	88,6	11,4
323	2,8E+06	1E+01	5,4E+05	2E+01	90,8	9,2

Tabla 1: valores de k_{on} obtenidos del ajuste biexponencial de la zona bimolecular de la traza temporal, para las tres proteínas estudiadas.

Para corroborar los valores obtenidos se realizaron también ajustes sobre la traza completa, utilizando una suma de exponenciales (análogas a la de la ecuación (V.3)) utilizando como función de ajuste 1, 2, 3, 4, y 5 exponenciales.

Para corroborar la bondad del ajuste y evitar una posible sobre-estimación de la cantidad de parámetros ajustables, se procedió a realizar sobre cada ajuste un test comparativo de Akaike.^[21]

Se obtuvo, en general, que la mejor función que ajustaba (el menor valor del test de Akaike) contenía 4 términos exponenciales. De este ajuste, se tomaron como las constantes cinéticas de unión de ligando las primeras dos o una (dependiendo de la proteína) y se las comparó con las obtenidas anteriormente.

También con esos valores se realizó el ajuste de Eyring (que se describirá a continuación). En todos los casos, los valores obtenidos de k_{on} , fracciones y parámetros de

activación, coinciden con los del ajuste biexponencial mencionado anteriormente considerando sus incertezas.

Los valores de k_{on} para las tres proteínas a temperatura ambiente reportados, concuerdan con los valores medidos, $6.9 \cdot 10^5 \text{ M}^{-1} \text{ s}^{-1}$ para Ph-TrHbO a 25 °C, mientras que $9.3 \cdot 10^5$ y $8.7 \cdot 10^4 \text{ M}^{-1} \text{ s}^{-1}$ para Tf-TrHbO a 20 °C, y $2.3 \cdot 10^5 \text{ M}^{-1} \text{ s}^{-1}$ para Bs-TrHbO a 20 °C.

Para Ph y Tf el ajuste biexponencial resultó ser considerablemente mejor que el monoexponencial, describiendo más precisamente la curva para los datos medidos. Por el contrario, para Bs las trazas se ajustaron bien con una monoexponencial.

A partir de las amplitudes se pueden calcular el porcentaje de la traza temporal que es descrita por cada tiempo de decaimiento obtenido. Estos porcentajes se muestran en las columnas 5 y 6 de la Tabla 1. Se puede observar que para Ph estas fracciones rondan el 50% para temperaturas bajas (274K), mientras que para temperaturas altas los porcentajes se vuelcan mayoritariamente hacia una relación 90%, 10% para kon_1 y kon_2 respectivamente. Para Tf la temperatura más baja de medida es 283K, a esta temperatura, los valores de los porcentajes es similar a Ph. Finalmente a temperatura alta (323K) los porcentajes son 90% 10% análogamente como ocurre con Ph.

El comportamiento general con la temperatura para los porcentajes descriptos anteriormente puede interpretarse como el sistema presentando dos poblaciones: una asociada a la unión de ligando desde solvente con una constante cinética kon_1 y la otra asociada a la unión de ligando desde solvente con una constante cinética kon_2 . Esta relación se ve modulada por la temperatura, de manera de aumentar el porcentaje de la población con que se une con una constante cinética kon_1 a medida que se aumenta la temperatura.

Este comportamiento no es presentado para Bs en donde el ligando CO se reúne desde el solvente con un solo tiempo característico.

Con los valores de k_{on} en función de la temperatura se procedió a realizar un ajuste con la ecuación de Eyring para obtener los valores de $\Delta H^\ddagger$, $\Delta S^\ddagger$ y finalmente $\Delta G^\ddagger$.

La forma funcional elegida para la ecuación de ajuste, surge de considerar la ecuación de Eyring en su definición más global (sin ninguna linealización), es decir:

$$k_{on} = \left(\frac{k_b T}{h} \right) \cdot e^{\frac{\Delta G^\ddagger}{RT}} = \left(\frac{k_b T}{h} \right) \cdot e^{\frac{\Delta H^\ddagger}{RT}} \cdot e^{\frac{\Delta S^\ddagger}{R}} \quad (V.4)$$

En donde se consideró que $\Delta G^\ddagger = \Delta H^\ddagger - T \cdot \Delta S^\ddagger$, con $\Delta H^\ddagger$ y $\Delta S^\ddagger$ independientes de la temperatura.

Esta forma funcional resulta ser la más precisa en términos de incertezas y de variabilidad en las soluciones para los parámetros ajustados.^[22]

Para poder decidir si los valores experimentales ajustan correctamente al modelo de Eyring, se decidió utilizar una forma funcional para el ajuste con la menor cantidad de parámetros ajustables y restringiendo los valores de k_b , h y R a los valores tabulados, de manera de establecer el modelo matemático lo más restrictivo posible.

A partir de lo anterior, la forma funcional para el ajuste que se utilizó fue la siguiente:

$$y = \left(\frac{x}{4,79926 \cdot 10^{-11}} \right) \cdot e^{-\frac{B}{1,98709369 \cdot x}} \cdot e^{\frac{A}{1,98709369}} \quad (V.5)$$

Donde $4,79926 \cdot 10^{-11}$ corresponde al valor numérico de $(k_b/h)^{-1}$ en unidades de Kelvin, y $1,98709369$ al valor numérico de R en unidades de cal/K.mol. A y B son los parámetros a ajustar.

Con esta forma funcional, el ajuste dará como resultados valores los parámetros de activación asociados a cada k_{on} , $\Delta H^\ddagger$ de activación (parámetro A) y $\Delta S^\ddagger$ de activación (parámetro B) con sus incertezas asociadas. La Figura 8 muestra los resultados del ajuste de los valores de k_{on} en función de la temperatura utilizando la ecuación de ajuste (V.5).

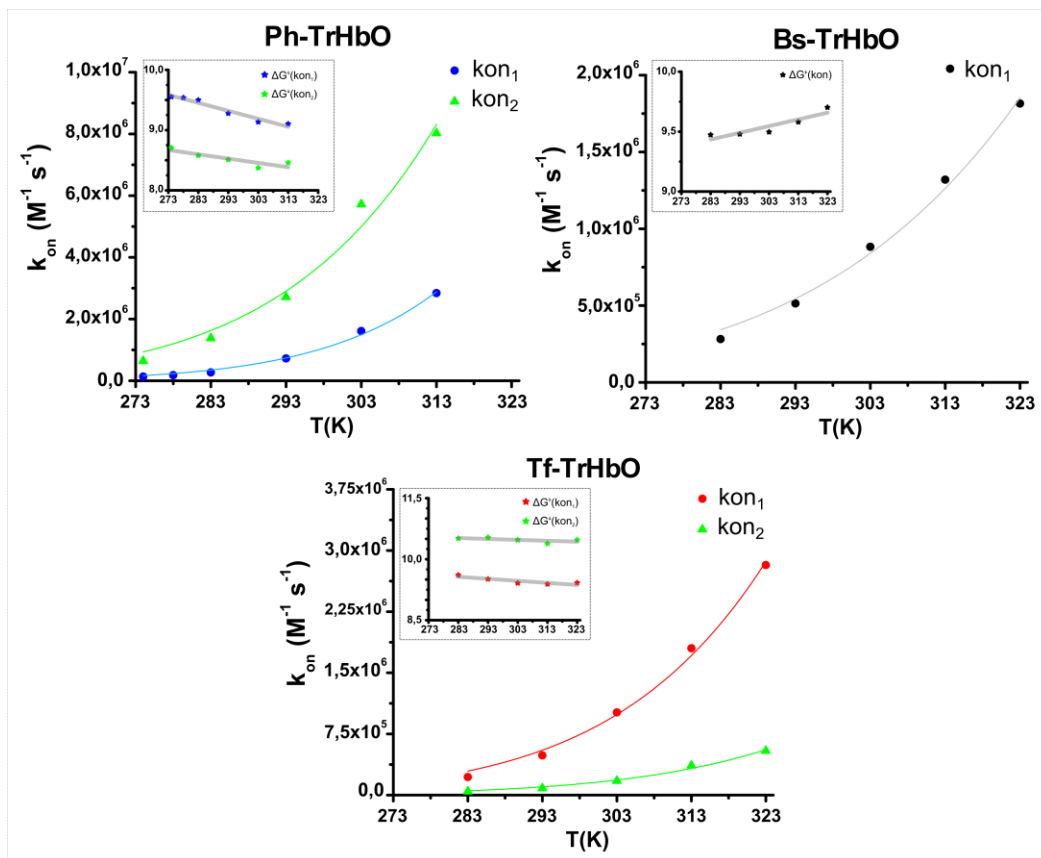


Figura 8: analisis de Eyring para los valores de k_{on} en función de la temperatura para las tres proteínas estudiadas. En el inset, se muestra la dependencia lineal de $\Delta G^\ddagger$ con la temperatura

Como control adicional, se calcularon los valores de $\Delta G^\ddagger$ calculados a partir de los valores de medidos de k_{on} y de la ecuación de Eyring (primera igualdad de la ecuación (V.4)) y se graficaron en función de la temperatura.

Estos resultados se muestran como inset en la Figura 8. Como puede observarse para todos los casos la relación entre $\Delta G^\ddagger$ y la temperatura es lineal dando evidencia de que $\Delta H^\ddagger$ y $\Delta S^\ddagger$ son aproximadamente constantes en todo el rango de temperatura estudiado, y por lo tanto validando las suposiciones del modelo de Eyring utilizado.

Los valores de $\Delta H^\ddagger$ y $\Delta S^\ddagger$ obtenidos del ajuste global de Eyring se muestran en la Tabla 2, en esta tabla se muestra también el valor $\Delta G^\ddagger$ a 298 K.

		$\Delta H^\ddagger$ (Kcal/mol)	$\Delta S^\ddagger$ (cal/K.mol)	T. $\Delta S^\ddagger$ (298 K) (Kcal/mol)	$\Delta G^\ddagger$ (298 K) (Kcal/mol)	R ²
Ph-TrHbO	kon ₁	11,8(±0,7)	9(±2)	2,6(±0,6)	9(±1)	0,99423
	kon ₂	9(±1)	2(±4)	1(±1)	8(±2)	0,97271
Bs-TrHbO	kon ₁	7,1(±0,5)	-8(±1)	-2,4(±0,3)	9,5(±0,8)	0,99035
Tf-TrHbO	kon ₁	9,7(±0,6)	1(±2)	0,3(±0,6)	9(±1)	0,99399
	kon ₂	10,1(±0,9)	-1(±3)	-0,3(±0,9)	10(±2)	0,98547

Tabla 2: parámetros de activación: $\Delta H^\ddagger$, $\Delta S^\ddagger$ y $\Delta G^\ddagger$ (298 K) asociados a cada k_{on} para las tres proteínas estudiadas.

V.4 Discusión

Los resultados muestran que para todas las proteínas estudiadas el comportamiento para las constantes cinéticas de unión de ligando CO responden al modelo de Eyring en el rango de temperaturas estudiado. Es decir, que para el proceso estudiado no se distingue un comportamiento diferencial entre estas proteínas pertenecientes a organismos adaptados y no adaptados frente a cambios de temperatura.

Según los resultados entonces, las tres proteínas estudiadas presentan un mismo mecanismo de unión (para cada una) independientemente de la temperatura estudiada, preservando las características principales de los canales así como del sitio activo de la proteína frente a cambios de temperatura. Esto implica que si existe una modulación por parte de la temperatura, no parece estar ligada al proceso de migración y unión de ligandos.

Los parámetros de activación $\Delta H^\ddagger$ y $\Delta S^\ddagger$ son aproximadamente constantes para el rango de temperaturas estudiado.

Al comparar los valores absolutos de $\Delta H^\ddagger$ y $-T \cdot \Delta S^\ddagger$ se evidencia que para la primera parte del proceso de unión (desde reactivos hasta el estado de transición) el proceso es fundamentalmente entálpico para las tres proteínas, con mayores valores para la relación entre $\Delta H^\ddagger$ y $\Delta S^\ddagger$ para Ph y Tf en comparación con Bs.

Como se ha mencionado, los factores más relevantes que modulan la velocidad el proceso de unión de pequeños ligandos en hemoproteínas están asociadas a dos características que pueden ser interpretadas independientemente: 1) migración del ligando a través de canales proteicos y 2) la unión del ligando al centro metálico (que puede incluir en algunos casos el desplazamiento de moléculas de agua u otros aminoácidos en el sitio activo). Estos dos factores han mostrado ser los de mayor peso en la descripción de los valores de k_{on} , y por consiguiente los determinantes del mecanismo de migración y unión de ligandos.^[16,23]

Los efectos térmicos permiten obtener valores de k_{on} mayores debido a la mayor facilidad para superar las barreras que tanto la matriz proteica (canales) como el estado de coordinación del hemo (moléculas de agua en el sitio activo), y no a un cambio en el mecanismo de migración/unión. Si así fuera, existiría un comportamiento que se aleja de lo establecido por la relación de k versus la temperatura que establece Eyring, debido a que $\Delta H^\ddagger$ y $\Delta S^\ddagger$ cambiarían significativamente producto de un cambio de mecanismo.^[2]

Lo anterior podría incluso evidenciarse como una pérdida en la linealidad en el gráfico de $\Delta G^\ddagger$ en función de la temperatura.

Es probable que si bien caracterizar el estado de transición en estos sistemas no sea intuitivo a primera vista, es probable que el mecanismo de entrada (en relación a la modulación de canales de migración y al contenido de agua en el sitio activo) se mantenga.

Desde un punto de vista estructural, se ha mostrado mediante técnicas de simulación computacional que algunas hemoproteínas presentan zonas en su estructura que son más afectadas que otras, con el incremento de la temperatura^[24]. En estos casos, las zonas móviles mostraron estar alejadas de los canales de migración y del sitio activo, y se presentan en este nuevo contexto como un mecanismo de atenuación ante el efecto de las perturbaciones térmicas.

Cabe aclarar que si bien existen trabajos donde se estudia el efecto de la temperatura en las proteínas, este capítulo establece una interpretación más directa de la relación entre el comportamiento estructural de las proteínas adaptadas y no adaptadas con su actividad, frente a cambios de temperatura.^[13,14,24]

V.5 Referencias

- (1) Madigan, M. T.; Martinko, J. M.; Brock, T. D. *Brock Biology of Microorganisms*, 11.ed., internat. ed.; Prentice Hall: Englewood Cliffs, New Jersey, **2006**.
- (2) Wang, W.; Roberts, C. J. Non-Arrhenius Protein Aggregation. *The AAPS Journal* **2013**, *15* (3), 840–851.
- (3) Thirumangalathu, R.; Krishnan, S.; Bondarenko, P.; Speed-Ricci, M.; Randolph, T. W.; Carpenter, J. F.; Brems, D. N. Oxidation of Methionine Residues in Recombinant Human Interleukin-1 Receptor Antagonist: Implications of Conformational Stability on Protein Oxidation Kinetics[†]. *Biochemistry* **2007**, *46* (21), 6213–6224.
- (4) Pan, B.; Abel, J.; Ricci, M. S.; Brems, D. N.; Wang, D. I. C.; Trout, B. L. Comparative Oxidation Studies of Methionine Residues Reflect a Structural Effect on Chemical Kinetics in rhG-CSF[†]. *Biochemistry* **2006**, *45* (51), 15430–15443.
- (5) Nellis, D. F.; Michiel, D. F.; Jiang, M.-S.; Esposito, D.; Davis, R.; Jiang, H.; Korrell, A.; Knapp, G. C.; Lucernoni, L. E.; Nelson, R. E.; Pritt, E. M.; Procter, L. V.; Rogers, M.; Sumpter, T. L.; Vyas, V. V.; Waybright, T. J.; Yang, X.; Zheng, A. M.; Yovandich, J. L.; Gilly, J. A.; Mitra, G.; Zhu, J. Characterization of Recombinant Human IL-15 Deamidation and Its Practical Elimination through Substitution of Asparagine 77. *Pharmaceutical Research* **2012**, *29* (3), 722–738.
- (6) Falconer, R. J.; Marangon, M.; Van Sluyter, S. C.; Neilson, K. A.; Chan, C.; Waters, E. J. Thermal Stability of Thaumatin-Like Protein, Chitinase, and Invertase Isolated from Sauvignon Blanc and Semillon Juice and Their Role in Haze Formation in Wine. *Journal of Agricultural and Food Chemistry* **2010**, *58* (2), 975–980.
- (7) Andrews, J. M.; Roberts, C. J. Non-Native Aggregation of α -Chymotrypsinogen Occurs

- through Nucleation and Growth with Competing Nucleus Sizes and Negative Activation Energies[†]. *Biochemistry* **2007**, 46 (25), 7558–7571.
- (8) Yoshioka, S.; Tajima, S.; Aso, Y.; Kojima, S. Inactivation and Aggregation of Beta-Galactosidase in Lyophilized Formulation Described by Kohlrausch-Williams-Watts Stretched Exponential Function. *Pharmaceutical Research* **2003**, 20 (10), 1655–1660.
 - (9) Yoshioka, S.; Aso, Y.; Izutsu, K.; Kojima, S. Is Stability Prediction Possible for Protein Drugs? Denaturation Kinetics of Beta-Galactosidase in Solution. *Pharmaceutical Research* **1994**, 11 (12), 1721–1725.
 - (10) Waseem, A.; Salahuddin, A. Anomalous Temperature-Dependence of the Specific Interaction of Concanavalin A with a Multivalent Ligand-Dextran. *Biochimica et Biophysica Acta (BBA)* **1983**, 746 (1–2), 65–71.
 - (11) Mehrotra, S.; Balaram, H. Methanocaldococcus Jannaschii Adenylosuccinate Synthetase: Studies on Temperature Dependence of Catalytic Activity and Structural Stability. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2008**, 1784 (12), 2019–2028.
 - (12) Matagne, A.; Jamin, M.; Chung, E. W.; Robinson, C. V.; Radford, S. E.; Dobson, C. M. Thermal Unfolding of an Intermediate Is Associated with Non-Arrhenius Kinetics in the Folding of Hen Lysozyme. *Journal of Molecular Biology* **2000**, 297 (1), 193–210.
 - (13) Tölgyesi, F.; Ullrich, B.; Fidy, J. Tryptophan Phosphorescence Signals Characteristic Changes in Protein Dynamics at Physiological Temperatures. *Biochimica et Biophysica Acta (BBA)* **1999**, 1435 (1–2), 1–6.
 - (14) Chen, B. L.; Baase, W. A.; Schellman, J. A. Low-Temperature Unfolding of a Mutant of Phage T4 Lysozyme. 2. Kinetic Investigations. *Biochemistry* **1989**, 28 (2), 691–699.
 - (15) Oliveberg, M.; Tan, Y. J.; Fersht, A. R. Negative Activation Enthalpies in the Kinetics of Protein Folding. *Proceedings of the National Academy of Science U.S.A.* **1995**, 92 (19), 8926–8929.
 - (16) Bustamante, J. P.; Abbruzzetti, S.; Marcelli, A.; Gauto, D.; Boechi, L.; Bonamore, A.; Boffi, A.; Bruno, S.; Feis, A.; Foggi, P.; Estrin, D. A.; Viappiani, C. Ligand Uptake Modulation by Internal Water Molecules and Hydrophobic Cavities in Hemoglobins. *The Journal of Physical Chemistry B* **2014**, 118 (5), 1234–1245.
 - (17) Mozzarelli, A.; Viappiani, C. Oxygen Binding and Sensing Proteins. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2013**, 1834 (9), 1683.
 - (18) Abbruzzetti, S.; Spyraakis, F.; Bidon-Chanal, A.; Luque, F. J.; Viappiani, C. Ligand Migration through Heme protein Cavities: Insights from Laser Flash Photolysis and Molecular Dynamics Simulations. *Physical Chemistry Chemical Physics* **2013**, 15 (26), 10686.
 - (19) Abbruzzetti, S.; Faggiano, S.; Spyraakis, F.; Bruno, S.; Mozzarelli, A.; Astegno, A.; Dominici, P.; Viappiani, C. Oxygen and Nitric Oxide Rebinding Kinetics in Nonsymbiotic Hemoglobin AHb1 from Arabidopsis Thaliana. *IUBMB Life* **2011**, 63 (12), 1094–1100.
 - (20) Abbruzzetti, S.; Bruno, S.; Faggiano, S.; Grandi, E.; Mozzarelli, A.; Viappiani, C. Time-Resolved Methods in Biophysics. 2. Monitoring Haem Proteins at Work with Nanosecond Laser Flash Photolysis. *Photochemical & Photobiological Sciences* **2006**, 5 (12), 1109.
 - (21) Akaike, H. A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control* **1974**, 19 (6), 716–723.
 - (22) Lente, G.; Fábíán, I.; Poë, A. J. A Common Misconception about the Eyring Equation. *New Journal of Chemistry* **2005**, 29 (6), 759.
 - (23) Bustamante, J. P.; Szretter, M. E.; Sued, M.; Martí, M. A.; Estrin, D. A.; Boechi, L. A Quantitative Model for Oxygen Uptake and Release in a Family of Heme proteins. *Bioinformatics* **2016**, btw083.
 - (24) Bustamante, J. P.; Bonamore, A.; Nadra, A. D.; Sciamanna, N.; Boffi, A.; Estrin, D. A.; Boechi, L. Molecular Basis of Thermal Stability in Truncated (2/2) Hemoglobins. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2014**, 1840 (7), 2281–2288.

VI. CONCLUSIONES GENERALES

Los sistemas biorelevantes, como las hemoproteínas, representan un continuo desafío en la fisicoquímica actual. La complejidad de estos sistemas y la aplicabilidad de los resultados obtenidos a partir de estudios realizados sobre ellos, destacan continuamente la importancia del trabajo interdisciplinar.

Para el caso de las proteínas, el objetivo de comprender de manera global los aspectos funcionales y estructurales, a menudo requiere del uso conjunto de diferentes tipos de herramientas como técnicas, metodologías, experimentos y desarrollos teóricos que se enmarcan en distintas disciplinas.

El continuo aprendizaje de los alcances y las limitaciones de cada tipo de herramienta y la intercomunicación entre los resultados obtenidos a partir de distintas disciplinas permite un efecto sinérgico que finalmente acerca a una interpretación global, paradigmática y contextualizada acerca del sistema estudiado.

En este contexto, el presente trabajo de tesis intentó integrar diversas áreas: desarrollos teóricos/matemáticos, la utilización de métodos computacionales a partir de la química teórica y la realización de experimentos.

En particular, se profundizó en la mejora en la estimación de perfiles de energía libre usando la técnica de dinámica molecular guiada; el estudio de la migración, unión y especiación de las especies de sulfuro; y el estudio del efecto de la temperatura en los procesos de migración y unión de pequeños ligandos en hemoproteínas. Estos enfoques, desarrollados a lo largo de los capítulos III, IV y V, centrados en la migración y unión de pequeños ligandos en hemoproteínas, resultaron de gran utilidad para la descripción de las bases moleculares y estructurales de los procesos químicos estudiados.

El capítulo III, presenta un estudio teórico acerca de distintos estimadores usados en las simulaciones de dinámica molecular guiada. Se mostró además, que la caracterización del tipo de distribución de trabajos resulta crucial para la determinación de un perfil de energía libre preciso. Finalmente, se aplicó este estudio a la migración y unión del anión superóxido en MnSOD, demostrando que la elección del estimador de los perfiles de energía libre calculados resulta crucial, pudiendo llevar a conclusiones químicas erróneas y discrepancias con los experimentos. Además, como se mostró en el capítulo IV, para el perfil de migración de las

especies de sulfuro en hemoglobina I de *L. pectinata*, cuando los valores de trabajo presentan dispersiones pequeñas (del orden de $k_B T$), las diferencias entre los estimadores no son significativas.

En el capítulo IV, además del análisis mencionado anteriormente, se abordó un estudio acerca de la migración, unión y especiación del sulfuro de hidrógeno (H_2S) en hemoproteínas, tomando como caso paradigmático la interacción de las especies de sulfuro con HbI de *L. pectinata*. En este capítulo, se usó una batería de técnicas de simulación computacional para postular un mecanismo de interacción de la especie de sulfuro con dicha proteína.

Finalmente, en el capítulo V, se decidió ampliar el estudio de los procesos de migración y unión de ligandos en hemoproteínas incluyendo un estudio de la influencia de la temperatura en los procesos reactivos mencionados. Se abordó entonces un estudio acerca de cómo afecta la temperatura en la migración y unión de ligandos en tres hemoglobinas truncadas pertenecientes a organismos adaptados a diferentes temperaturas.

Los resultados mostraron que las tres proteínas estudiadas no evidenciaron un comportamiento diferencial en las constantes de unión (k_{on}), y además que éstas respondieron al modelo de Eyring. Esto permitió concluir que cada proteína presenta un mismo mecanismo de unión independientemente de la temperatura, y preserva los factores principales que modulan la migración y unión de ligandos. Finalmente, se postuló que si la actividad de estas hemoproteínas frente a pequeños ligandos, están moduladas apreciablemente por la temperatura, dicha modulación no se evidencia en la etapa de migración y unión de los ligandos.

En resumen, las preguntas planteadas, los diversos enfoques abordados y las diversas herramientas utilizadas en esta tesis, nos permitieron aportar nueva información acerca de los puntos de vista a tener en cuenta para estudiar los procesos reactivos mencionados y adquirir además un entrenamiento profundo, tanto en el empleo de diferentes técnicas, como en la interpretación de los resultados, objetivos fundamentales en mi formación académica.

VII. PUBLICACIONES

VII.1 Publicaciones enmarcadas en los temas de la presente tesis

- Nicoletti FP, Comandini A, Bonamore A, Boechi L, **Boubeta FM**, Feis A, Smulevich G, Boffi A. *Sulfide binding properties of truncated hemoglobins*. Biochemistry. **2010** Mar 16;49(10):2269-78.
- Bieza SA*, **Boubeta FM***, Feis A, Smulevich G, Estrin DA, Boechi L, Bari SE. *Reactivity of inorganic sulfide species toward a heme protein model*. Inorg Chem. 2015 Jan 20;54(2):527-33.
- **Boubeta FM**, Bari SE, Estrin DA, Boechi L. *Access and Binding of H₂S to Hemoproteins: The Case of Hbl of Lucina pectinata*. J Phys Chem B. **2016** Sep 15;120(36):9642-53.
- Arrar M*, **Boubeta FM***, Szretter Noste ME, Rodriguez D, Sued M, and Boechi L. *On the optimal use of the Jarzynski equality*. Arrar M & **Boubeta FM** contributed equally. **Trabajo enviado.**
- **Boubeta FM**, Abbruzzetti S, Giordano Daniela, Estrin DA, Boffi A, Verde C, Viappiani C, and Boechi L. *Temperature-independence of CO binding mechanism in heme proteins from adapted and non adapted organisms*. **Trabajo en preparación.**

* Ambos autores contribuyeron igualmente a este trabajo

VIII. AGRADECIMIENTOS

A mis directores de tesis, a Darío y Leo, quienes a partir de sus consejos y a través de extensos debates científicos me han podido transmitir parte de su vasto conocimiento y experiencia científica, en esta etapa especial entre muchas otras pertenecientes al continuo aprendizaje que representa la vida científica.

En especial, al gran amigo en que se convirtió Leo, que empujó durante más de cinco años para que se materializaran todos los aspectos que implican presentar una tesis Doctoral. A vos gracias hermano!

A las investigadoras del Instituto del Cálculo, a Mariela, Maru y Daniela, por las incansables discusiones matemáticas tan gratificantes y educativas.

Quisiera también agradecer también a los tres jurados: al Dr Skaf, a la Dra Pickholz y al Dr Olabe, que han aceptado ser jurados de esta tesis, aun teniendo en algunos casos que viajar desde lejos.

A varios investigadores del DQIAyQF/INQUIMAE, especialmente a Sara Bari, excelente persona que he tenido el placer de conocer desde antes y durante la tesis, con quien he podido aprender muchísimo con el adicional de haber disfrutado mucho durante el proceso de aprendizaje, su eterna predisposición, su humor, y su conocimiento han generado preciados tesoros en mi persona.

Por otro lado, quisiera también agradecerle a varios profesores que han dictado con enorme paciencia y entusiasmo los cursos de doctorado: Gabriela Lagorio, Daniel Laria, Damian Scherlis y Enrique San Román.

A Nano, por haberme aconsejado en todo momento, en los buenos y malos, en temas laborales y personales. También a su hermano, a Lolo, por ayudarme y aconsejarme sin ningún tipo de compromiso en variadas oportunidades, su ayuda resultó fundamental durante todo el capítulo de cinética.

Dentro del grupo de Modelado Molecular con el cual compartí más de 5 años, a Will, Ari, JuanPis Bustamante, Uri, Diego, Fran, Facto, Lourdes, Yami, Lu, Laia, Clau, Lula, Jhonny, Mauro, Eli, Lucas, Fede I, Fede P, Nico, Vicky, Juan Pablo M., Juan P. Arcón, y Marce. Con todos

hemos compartido en más de una oportunidad conversaciones científicas muy gratificantes, y muchísimos otros momentos espectaculares como juntadas, mates, salidas, y tantas otras.

En particular, a Noosh por su incansable ayuda, y por las discusiones científicas que día a día nos motivaron, a toda hora, y usando todos los medios de comunicación existentes. Junto con Leo y Emi creo que forman el trío más genuinamente optimista que conocí.

A la ANPCyT, al CONICET, a la Universidad de Buenos Aires y a la Unión Europea, por el apoyo en becas, equipamiento y congresos a los que asistí. Al DQIAyQF y al INQUIMAE por el espacio concedido. También a todas las entidades que financiaron en gran medida los viajes internacionales realizados.

A Alejandra y Alberto del departamento de QI por la increíble eficiencia en su trabajo, llegando compartir en varias circunstancias momentos de una relación más familiar que laboral.

Quisiera agradecer a todos los grupos extranjeros que amablemente me recibieron en las pasantías que realicé en sus laboratorios: Cristiano Viappiani y Stefania Abbruzzetti (Parma-Italia), y a los chicos de Firenze: Barbara, Alessandro y Maria Angela (Florenzia-Italia)

En los aspectos personales más íntimos, solo haré una pequeña mención porque cuando la palabra escrita no alcanza, es conveniente ahorrarlas esperando ansioso un afectivo mano a mano: a mis Viejos sobre todo, que me bancaron largos días de ausencia, entre viajes, congresos, estudios, etc. A Tami, mi compañera de todos los días, la que más toleró todo de mí, la que me devuelve a la realidad y me lleva a todos lados. A todos mis amigos: a la banda de la facultad, y la banda del barrio. Ellos saben quiénes son y que significan para mí.