Biblioteca Digital
F C E N - U B A

BIBLIOTECA CENTRAL LUIS F LELOIR
BIBLIOTECA CENTRAL LELOIR
FACULTAD DE CIENCIAS EXACTAS Y NATURALES UBA

Tesis Doctoral

Estudio fisicoquímico y bioinformático de la familia de hemoglobinas truncadas

Bustamante, Juan Pablo

2015-11-13

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:
Bustamante, Juan Pablo. (2015-11-13). Estudio fisicoquímico y bioinformático de la familia de hemoglobinas truncadas. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires.

Cita tipo Chicago:
Bustamante, Juan Pablo. "Estudio fisicoquímico y bioinformático de la familia de hemoglobinas truncadas". Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires. 2015-11-13.

EXACTAS UBA
Facultad de Ciencias Exactas y Naturales

UBA
Universidad de Buenos Aires

Dirección: Biblioteca Central Dr. Luis F. Leloir, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires. Intendente Güiraldes 2160 - C1428EGA - Tel. (++54 +11) 4789-9293

Contacto: digital@bl.fcen.uba.ar

UNIVERSIDAD DE BUENOS AIRES

Facultad de Ciencias Exactas y Naturales

Departamento de Química Biológica

Estudio fisicoquímico y bioinformático de la familia de hemoglobinas truncadas

Tesis presentada para optar al título de Doctor de la Universidad de Buenos Aires en el área Química Biológica

Juan Pablo Bustamante

Director de Tesis: Prof. Dr. Darío A. Estrin

Director Asistente: Dr. Leonardo Boechi

Consejero de Estudios: Prof. Dr. Marcelo Martí

Buenos Aires, noviembre de 2015

Estudio fisicoquímico y bioinformático de la familia de hemoglobinas truncadas

Resumen

Las hemoproteínas son proteínas que poseen un grupo hemo unido generalmente de forma covalente. Participan en transporte de gases, en reacciones de transferencia de electrones y catálisis de reacciones redox. Un grupo muy particular dentro de estas proteínas son las hemoglobinas truncadas (trHbs), que constituyen una familia interesante de referencia para estudios de estructura y función de proteínas, debido a su plegado conservado y pequeño tamaño. Las trHbs están presentes en los tres super reinos de la vida, bacteria, arquea y varios organismos eucariotas. Han sido clasificadas en tres grupos, conocidos como N, O y P. Si bien la mayoría de trHbs no poseen una función asignada, se conoce que ésta generalmente depende de su afinidad por ligandos pequeños, es decir, de la relación entre la velocidad a la cual captan y liberan ligandos, caracterizadas por constantes de velocidad k_{on} y k_{off}, respectivamente. De allí la relevancia de abocarse al estudio de los determinantes moleculares de tal afinidad.

Los procesos de captación y liberación de ligandos han sido extensamente estudiados durante los últimos años por diferentes grupos de investigación, incluido el grupo donde se desarrolló la presente tesis, utilizando enfoques teóricos y experimentales. Sin embargo, solo se estudiaron algunos de los procesos involucrados, tales como el rol de túneles e interacciones de puentes de hidrógeno, encontrándose únicamente tendencias cualitativas respecto a constantes k_{on} y k_{off}.

En la presente tesis, a través de la aplicación de diversas técnicas computacionales, se profundizó en la descripción anterior mediante la inclusión de dos determinantes moleculares cruciales en el proceso de captación de ligandos: las cavidades internas de las proteínas y el rol de las moléculas de agua presentes en el sitio activo. Adicionalmente, se construyó un modelo matemático, validado a través de un riguroso análisis estadístico, que permite predecir de manera cuantitativa las constantes k_{on} y k_{off} de O_2.

Dado que la mayoría de trHbs no poseen una función asignada, se extendió el alcance del modelo propuesto a toda la familia de trHbs, prediciendo valores para una gran cantidad de trHbs (sobre un total de ~1100 secuencias proteicas). Se asignaron posibles funciones moleculares y se analizó a esta familia de proteínas en un contexto evolutivo-funcional, caracterizando a cada grupo filogenético. De manera adicional, el estudio evolutivo reveló que hay un grupo de trHbs que comparte características comunes que lo diferencian

significativamente del resto de los grupos conocidos (N, O y P), por lo que se lo clasificó como un cuarto grupo, denominado Q.

Finalmente, puesto que una de las proteínas más estudiadas durante el desarrollo de la presente tesis (la trHb O de *Thermobifida fusca*, Tf-trHbO) es una proteína termoestable perteneciente a una actinobacteria termófila, se estudiaron las bases moleculares de su termoestabilidad comparándola con una trHb mesoestable: la trHb O de *Mycobacterium tuberculosis*. Se encontró que la alta estabilidad térmica de la Tf-trHbO se explica, en gran parte, por el cambio de un único aminoácido que altera la estructura de un loop, haciéndolo más flexible y promoviendo la formación de puentes salinos.

En su conjunto, esta tesis presenta una profunda caracterización de dos factores cruciales en el proceso de captación de ligandos en hemoglobinas truncadas, que, incluidos en un modelo matemático junto a una descripción del proceso de liberación de ligandos como ha sido previamente descripta, permiten explicar valores experimentales de k_{on} y k_{off} de O_2 con una precisión muy alta. La extensión del modelo a toda la familia de trHbs ha permitido asignar posibles funciones moleculares a una gran cantidad de miembros de estas globinas. Complementado con un análisis evolutivo y bioinformático, se ha logrado además obtener una detallada caracterización de cada grupo de trHbs, descubriendo incluso un grupo nuevo. Por último, el estudio teórico sobre la termoestabilidad de la Tf-trHbO abre las puertas a posteriores estudios que analicen las bases moleculares de la adaptación a condiciones extremas.

Palabras clave: hemoglobinas truncadas, cavidades internas, sitios de hidratación, dinámica molecular, QM/MM, predicción de constantes cinéticas, predicción secuencia-estructura-función, filogenia, evolución molecular, bioinformática, termoestabilidad.

Physicochemical and bioinformatic study of the truncated hemoglobin family

Abstract

Hemeproteins are proteins that possess a heme group generally covalently bound. They participate in gas transport, electron transfer reactions and catalysis of redox reactions. A particular group within these proteins is the truncated hemoglobins (trHbs), which constitute an interesting benchmark group for structure-to-function studies, due to its conserved fold and small size. The trHbs are present in all three superkingdoms of life, bacteria, archeas and several eukaryotes. They have been classified into three groups, labeled as N, O and P. While most trHbs do not have an assigned function, it is well known that generally their function depends on their affinity for small ligands, i.e. the balance between capture and release ligand rates, characterized by the k_{on} and k_{off} rate constants, respectively. Hence, the importance of focusing on the study of the molecular determinants of this affinity.

The ligand capture and release processes have been extensively studied throughout in the last several years, by different research groups, including the group in which this thesis has been developed, with theoretical and experimental approaches. However only some involved processes, such as the role of tunnels and hydrogen bond interactions, were studied finding merely qualitative trends in k_{on} and k_{off} rate constants.

In the present thesis, through the application of a variety of computational techniques, we incorporated to the previous description two key determinant factors of the ligand capture process: protein internal cavities and the role of water molecules in the active site. Additionally, validated through a rigorous statistical analysis, a mathematical model for the quantitative prediction of O_2 k_{on} and k_{off} rate constants was built.

Due to the fact that most trHbs do not have an assigned function, we extended our proposed model to the whole family, predicting values for a large amount of trHbs (over ~1100 protein sequences). In this context, possible molecular functions were assigned, as well as a protein family analysis in an evolutionary-functional scenario, characterizing each phylogenetic group. Furthermore, the evolutionary study revealed a novel trHb group that shares common features that differentiate it from the rest known groups (N, O and P), so it was classified in a fourth group, labeled as Q.

Finally, given that one of the most studied proteins along this thesis (the trHb O from *Thermobifida fusca,* Tf-trHbO) is a thermostable protein from a thermophilic actinobacteria, the

molecular basis of its thermostability was studied, comparing it with a mesostable trHb: the trHb O from *Mycobacterium tuberculosis*. It has been found that the high thermal stability of Tf-trHbO, is mainly explained by a single amino acid change that alters a loop structure making it more flexible, which also promotes the formation of salt bridges.

Overall, in this thesis we present for the truncated hemoglobin family a deep characterization of two key determinant factors of the ligand capture process that, included in a mathematical model together with a description of the ligand release process that has been previously made, allows us to explain experimental O_2 k_{on} and k_{off} values with high precision. The extension of the model to the whole trHb family enabled putative molecular function assignments. Complemented by an evolutionary and bioinformatics analysis, a detailed characterization of each trHb group was also achieved, describing a novel group. Last but not least, the theoretical study of the Tf-trHbO thermostability is a first step towards further in-depth studies of the molecular basis of adaptation to extreme conditions.

Keywords: truncated hemoglobins, internal cavities, hydration sites, molecular dynamics, QM/MM, kinetic constants prediction, sequence-structure-function prediction, phylogeny, molecular evolution, bioinformatics, thermostability.

A Vicky, la mujer de mi vida, y a León, mi pequeño gran amor.

A mamá, papá y Belu, por su incondicional acompañamiento y amor profundo.

A la Tita, por ser un ejemplo de vida y enseñarme a vivir siempre con alegría.

Índice general

Resumen -

Abstract -

I. Introducción **13**

I.1 Biología / bioinformática estructural ... 13

I.2 Hemoproteínas ... 13

I.3 Superfamilia de globinas ... 14

I.4 Hemoglobinas truncadas (trHbs) ... 15

I.5 Objetivos y lineamientos ... 17

I.6 Referencias ... 18

II. Métodos generales de simulación computacional **21**

II.1 Métodos basados en la mecánica cuántica (QM) ... 22

II.1.1 Teoría del Funcional de la Densidad (DFT) ... 23

II.2 Métodos basados en la mecánica clásica (MM) ... 24

II.2.1 Campo de fuerzas clásico ... 24

II.2.2 Modelos para el solvente ... 25

II.2.2.1 Solvente explícito: TIP3P ... 26

II.2.3 Condiciones periódicas de contorno y radio de corte ... 26

II.2.4 Parámetros asociados ... 28

II.3 Métodos híbridos cuántico-clásicos (QM/MM) ... 28

II.4 Exploración de la superficie de energía potencial ... 29

II.4.1 Minimización de la energía ... 30

II.4.2 Simulaciones de dinámica molecular clásica ... 31

II.4.3 Detalles técnicos. Configuración de los sistemas y parámetros de simulación ... 34

II.4.4 Detalles técnicos. Energías de unión de O_2 (ΔE_{O_2} y $\Delta\Delta E_{O_2}$) ... 36

II.4.5 Muestreo de migración de ligandos explícitos versus implícitos ... 37

II.4.6 Detalles técnicos. Topología y perfiles de energía libre de migración de ligandos ... 40

II.5 Referencias 40

Resultados y discusión **43**

III. Cavidades internas y moléculas de agua: factores determinantes en la captación de ligandos **45**

III.1 Introducción 45

III.2 Métodos computacionales 48

III.2.1 Configuración de los sistemas y parámetros de simulación 48

III.2.2 Análisis de los perfiles de energía libre de migración de ligandos en las proteínas Tf-trHbO y Mt-trHbN 48

III.2.3 Definición, identificación y caracterización de sitios de agua (SA) .. 49

III.3 Resultados y discusión 49

III.3.1 Sistema de cavidades permanentes y transitorias 49

III.3.2 Modulación del proceso de captación de ligandos 52

III.3.3 G8: una posición estructural estratégica para la modulación de cavidades internas y moléculas de agua 59

III.4 Conclusiones 62

III.5 Referencias 63

IV. Un modelo cuantitativo para la captación y liberación de O_2 **67**

IV.1 Introducción 67

IV.2 Métodos computacionales 68

IV.2.1 Configuración de los sistemas y parámetros de simulación 68

IV.2.2 Topología y perfiles de energía libre de migración de O_2 68

IV.2.3 Energía de unión de O_2 (ΔE_{O_2} y $\Delta\Delta E_{O_2}$) 68

IV.3 Resultados 68

IV.3.1 Modelado de la constante de captación del ligando (k_{on}) 68

IV.3.2 Modelado de la constante de liberación del ligando (k_{off}) 71

IV.3.3 Predicción de la constante de captación del ligando (k_{on}) 73

IV.3.4 Predicción de la constante de liberación del ligando (k_{off}) 74

IV.3.5 Validación estadística 75

IV.4 Discusión 76

IV.5 Conclusiones 78

IV.6 Referencias 79

V. Relaciones evolutivas y funcionales en la familia de trHbs 81

V.1 Introducción 81

V.2 Métodos computacionales 82

V.2.1 Análisis filogenético basado en secuencias proteicas 82

V.2.1.1 Fuente de datos e identificación de secuencias de trHbs 82

V.2.1.2 Alineamientos múltiples de secuencias (AMS) 82

V.2.1.3 Análisis filogenético 83

V.2.1.4 Análisis adicionales 83

V.2.2 Modelado y simulación 84

V.2.2.1 Modelos por homología. Configuración de los sistemas y parámetros de simulación clásica 84

V.2.2.2 Topología y perfiles de energía libre de migración de O_2 84

V.2.2.3 Energía de unión de O_2 (ΔE_{O_2} y $\Delta\Delta E_{O_2}$) 84

V.2.2.4 Determinación de las constantes de captación y liberación de O_2 (k_{on} y k_{off}) 85

V.3 Resultados 86

V.3.1 Identificación de aminoácidos claves que definen la afinidad por ligandos pequeños 86

V.3.2 Revisión del árbol filogenético de las trHbs y análisis de composición aminoacídica 88

V.3.3 Modelado de toda la familia proteica 91

V.3.4 Análisis global de las constantes cinéticas y la afinidad por O_2 92

V.3.5 Combinación de resultados fisicoquímicos y filogenéticos 93

V.4 Discusión 95

V.4.1 Razones estructurales subyacentes de la afinidad por ligandos pequeños 96

V.4.2 De la estructura a la función 96

V.4.3 Distribución funcional basada en organismos 98

V.5 Conclusiones 99

V.6 Referencias 99

VI. Bases moleculares de la adaptación a altas temperaturas en la trHbO de *T. fusca* 103

VI.1 Introducción 103

VI.2 Métodos computacionales 105

VI.2.1 Configuración de los sistemas y parámetros de simulación 105

VI.3 Resultados y discusión 105

VI.3.1 Dinámica de las proteínas frente al cambio de temperatura 105

VI.3.2 Comportamiento de proteínas mutantes *in silico* 108

VI.3.3 Impacto en interacciones polares 109

VI.3.4 Mediciones de desnaturalización de proteínas *wild type* y mutantes intercambiadas 110

VI.4 Conclusiones 111

VI.4.1 Flexibilidad frente al cambio de temperatura 111

VI.4.2 Interacciones polares frente al cambio de temperatura 112

VI.4.3 Influencia de la identidad de la posición E3 sobre el cambio de temperatura 113

VI.5 Referencias 113

VII. Conclusiones y perspectivas generales 115

VIII. Otras publicaciones 119

IX. Anexo 121

X. Agradecimientos 139

XI. Abreviaturas 141

I. Introducción

I.1 Biofisicoquímica y bioinformática

La biofisicoquímica es un área multidisciplinaria que, mediante el estudio de procesos químicos y físicos, busca comprender fenómenos biológicos a distintos niveles: atómico y molecular. Se ocupa de estudiar la identidad y las interacciones entre moléculas pequeñas, medianas, macromoléculas y complejos macromoleculares formados por lípidos, azúcares, proteínas y ácidos nucleicos.

Por otra parte, la bioinformática, un área relativamente emergente en Argentina, surge en la década de los '90 con la masiva aparición de grandes volúmenes de datos biológicos que se van depositando en almacenes digitales. La bioinformática constituye también un espacio multidisciplinario con dos pilares fundamentales: la biología y la informática, nutriéndose además de la química, la matemática y la estadística. Este campo de la ciencia busca estudiar y procesar datos biológicos para dar respuesta a distintos interrogantes a escalas microscópicas y macroscópicas.

La conjunción de la biofisicoquímica y la bioinformática, daría lugar a lo que hoy muchos llaman bioinformática estructural o biología estructural computacional. En la presente tesis de doctorado es esta interacción entre áreas de la ciencia la que permite el estudio desde casos particulares a generales, extrayendo interpretaciones y conclusiones para toda una familia de proteínas. A su vez, la sinergia entre estas áreas del conocimiento también facilita enfoques que van desde lo general a lo particular, como en el diseño racional de drogas partiendo de inmensas bibliotecas de posibles compuestos seleccionados como blancos terapéuticos.

I.2 Hemoproteínas

Las proteínas desempeñan un papel fundamental para la vida y son las biomoléculas más versátiles y diversas conocidas. Dentro de la vasta cantidad de proteínas que existen, un conjunto muy importante son las metaloproteínas, aquellas que contienen un ión metálico como cofactor y, dentro de éstas, están las hemoproteínas, que contienen un grupo prostético denominado hemo, con un Fe como átomo central. El grupo hemo (figura I.1.A) contiene un anillo tetrapirrólico cíclico con sustituyentes laterales (figura I.1.B). En el centro de este anillo tetrapirrólico se encuentra un átomo de hierro que se encuentra generalmente en estado

ferroso (FeII) o férrico (FeIII), y que está coordinado a los cuatro nitrógenos de los pirroles. El pirrol (figura I.1.C) es un compuesto químico orgánico aromático y heterocíclico, que consiste en un anillo de cinco miembros que responde a la fórmula C_4H_5N.

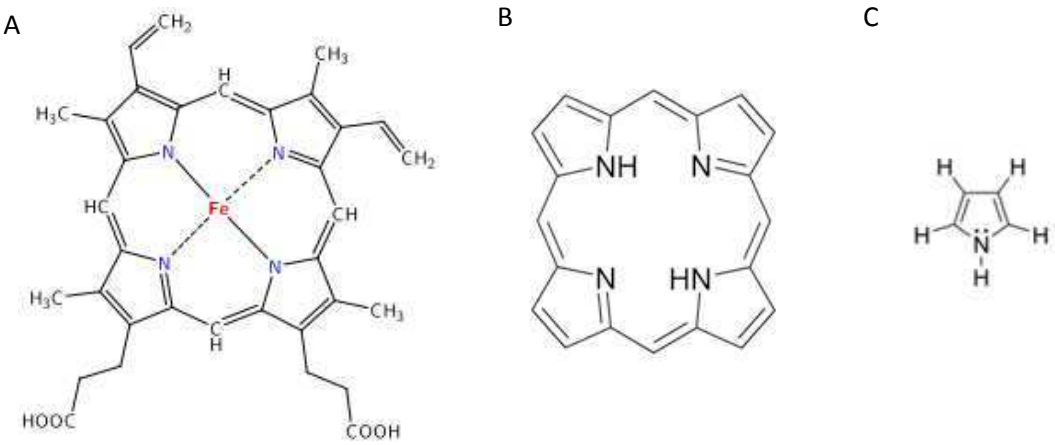


Figura I.1. Estructuras químicas del grupo hemo (A) y sus constituyentes: un anillo tetrapirrólico cíclico con sustituyentes laterales (B) y el pirrol (C).

I.3 Superfamilia de globinas

La clasificación anterior no considera relaciones evolutivas entre proteínas. Cualquier proteína conteniendo un grupo hemo es llamada hemoproteína, aunque dichas proteínas no posean un parentesco evolutivo. Si por el contrario se realiza una clasificación evolutiva de proteínas considerando sus secuencias aminoacídicas, aperece un grupo bien diferenciado de proteínas solubles que contienen un grupo hemo y un plegado muy particular, denominadas globinas.

La superfamilia de globinas ofrece un amplio, diverso y extensamente estudiado grupo de biomoléculas, cuya función está íntimamente ligada a la afinidad y reactividad por ligandos pequeños (principalmente O_2, pero también NO, CO, NO_2 y S_2H)[1,2]. Las Hbs pueden tener otras funciones además del almacenamiento y transporte de oxígeno, incluyendo funciones de censado y enzimáticas, como catalasas y peroxidasas[3]. Las globinas con alta afinidad por O_2 funcionan usualmente como enzimas redox de O_2 (y otras especies reactivas de oxígeno y nitrógeno), como las hemoglobinas truncadas de micobacterias[4,5]. Globinas con afinidad moderada por O_2 en general actúan como proteínas de almacenamiento y transporte de O_2, como la mioglobina y la hemoglobina de mamíferos[6,7], mientras que las globinas con baja afinidad por O_2 funcionan como sensores redox de NO y CO, como la guanilato ciclasa soluble o los sensores acoplados a globinas (GCS)[8].

I.4 Hemoglobinas truncadas

Las hemoglobinas truncadas (trHbs), también llamadas Hbs 2/2, son uno de los tres linajes dentro de la superfamilia de globinas, junto con las globinas de dominio único 3/3 (globinas SD), flavohemoglobinas y las GCS 3/3 y protoglobinas. La familia de trHbs es la única presente en los tres super reinos de la vida, bacteria, arquea y varios organismos eucariotas[9,10]. Basado en un análisis filogenético realizado con 111 secuencias disponibles en el año 2006[11], y sobre la caracterización reportada por Wittenberg y otros investigadores cuatro años antes[9], las trHbs han sido agrupadas en tres grupos, conocidos como N, O y P (también llamados I, II y III), dando lugar a la notación trHbN, trHbO y trHbP, respectivamente. Estas proteínas se caracterizan por su simplificado y único plegado caracterizado por 2 α-hélices sobre 2 α-hélices (2/2) (figura I.2A) y por su reducido tamaño relativo de entre un 75 y 80% respecto a las hemoglobinas de vertebrados 3/3[9]. Dada su estructura tridimensional, las trHbs se organizan en bloques definidos por α-hélices y loops nombrados en orden alfabético, como se muestra en la figura I.2A, con la finalidad de facilitar la descripción textual y llegar a una rápida identificación de la posición espacio-estructural a la cual se está refiriendo. Utilizando esta notación, el plegado 2/2 se da entre las hélices B y E sobre las G y H del plegado típico de globinas. Estos dos pares de hélices se disponen en el espacio en forma antiparalela y se ensamblan de manera tal de rodear y resguardar al grupo hemo del solvente, conteniéndolo en una cavidad hidrofóbica. Dentro de ésta, el plano que describe el anillo porfirínico define los sitios proximal y distal. El grupo hemo, además de ser contenido por una cavidad hidrofóbica, es retenido por la proteína formando un enlace de quinta coordinación entre el átomo de hierro del mismo y una histidina característica del sitio proximal perteneciente a la hélice F, la HisF8. Por otro lado, hay varios aminoácidos polares encargados de estabilizar los propianatos de este cofactor metálico, los cuales se identificarán y describirán en el capítulo V.

Como se comentó en la sección anterior, las globinas poseen funciones relacionadas con la afinidad y reactividad por ligandos pequeños. El valor de dicha afinidad está determinado por el cociente entre las constantes cinéticas de asociación (k_{on}) y disociación (k_{off}). Estas constantes corresponden a los procesos moleculares de captación y liberación, respectivamente, de ligandos pequeños. Para que ambos procesos se lleven a cabo, la participación del grupo hemo es crucial. Sin embargo, puesto que éste está enterrado en el interior de estas globinas, está protegido del ambiente acuoso y, por ende, no está directamente accesible a ligandos en solución. Tal es así que, en el proceso de captación, los ligandos deben encontrar al menos un "camino" que conecte el solvente con el sitio activo de una globina, un canal de migración. Una vez que un ligando logra migrar desde el solvente hacia el sitio activo a través de estos túneles

o canales, debe desplazar otros posibles ligandos que se encuentren en el sitio activo (ya sea unidos al átomo Fe o no, ligandos externos o aminoácidos proteicos) y, finalmente, formar el enlace Fe-ligando y ser estabilizado por interacciones con la proteína[12–14]. Por otro lado, la liberación de los ligandos involucra el proceso inverso: la ruptura de las interacciones ligando-proteína y su escape hacia el solvente[13,15].

Ambos procesos moleculares de captación y liberación de ligandos pequeños han sido extensamente estudiados en trHbs tanto por el grupo de investigación donde se ha desarrollado presente tesis como por otros grupos, utilizando enfoques teóricos y experimentales[2,5,14,17–23]. Sin embargo, solo algunos de los procesos involucrados fueron estudiados, tales como el rol de túneles e interacciones de puentes de hidrógeno en el proceso de migración de ligandos hacia y desde el sitio activo. En todas las oportunidades, se estudiaron casos aislados de a una o dos proteínas a la vez, y hasta el momento los resultados de las simulaciones no han sido puestos a prueba frente a todos los datos experimentales disponibles. Por otro lado, solamente se encontraron tendencias cualitativas respecto a datos experimentales de k_{on} y k_{off}.

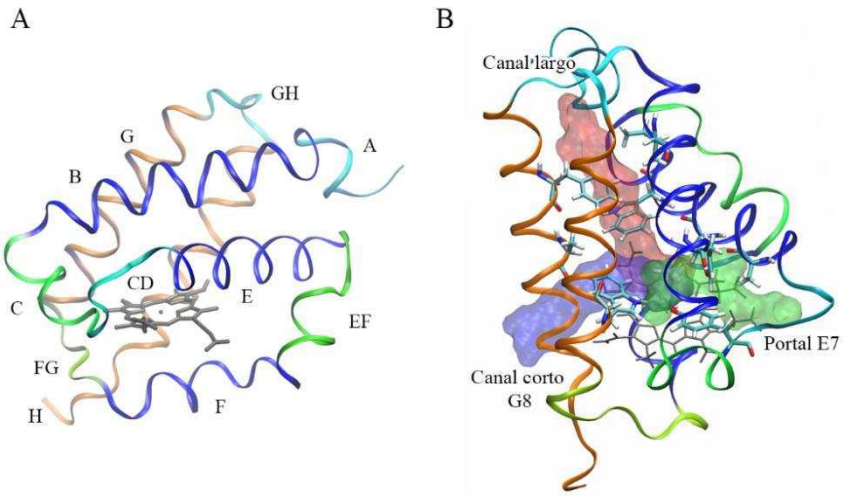


Figura I.2. A) Plegado típico de la estructura de las trHbs y los bloques estructurales comúnmente usados en diferentes colores. B) Representación esquemática mostrando los tres posibles canales para ingreso y egreso de ligandos en trHbs: canal largo, canal corto G8 y portal E7 (en rojo, azul y verde, respectivamente).

En cuanto a la migración de ligandos pequeños desde el solvente hacia el sitio activo, las trHbs presentan tres posibles túneles o canales topológicamente diferentes, denominados canal largo, canal corto G8 y portal E7 (la ruta principal de ingreso y egreso de ligandos en mioglobina y hemoglobina [24,25]), como puede observarse en la Figura I.2B. Algunos miembros de esta familia de proteínas presentan solo uno de estos túneles "abiertos", mientras que otros poseen dos de los tres descriptos.

Por otro lado, el entorno del sitio activo, también llamado sitio distal, está definido por cinco posiciones estructurales denominadas, al igual que la HF8, de acuerdo a la posición en la estructura secundaria a la que pertenecen: B10, CD1, E7, E11 y G8[9], esquematizada en la figura I.2. La identidad de estos aminoácidos da lugar a características distintivas en las trHbs, puesto que modulan la afinidad de acuerdo a cuán fuertemente estabilicen a los ligandos exógenos. Si

bien resulta tentador analizar la contribución de cada aminoácido por separado, actúan como un grupo al momento de definir la reactividad por cada ligando, por lo que es dicho grupo quien modula la afinidad. Esto está evidenciado por los ejemplos mostrados en la figura I.3, donde la presencia de aminoácidos capaces de formar puentes de hidrógeno (como la YB10 en el caso presente) no necesariamente significa que ellos actuarán estableciendo este tipo de interacción de corto alcance con los ligandos allí presentes. En la trHb P *de Campylobacter jejuni* (Cj-trHbP), la presencia de HE7 compite por el grupo OH de la YB10, mientras que el WG8 expulsa al O_2 lejos de ésta, atrayéndolo hacia él. Por otro lado, en un caso donde no hay W en G8 y H en E7, en la trHb N de *Mycobacterium tuberculosis* (Mt-trHbN), la YB10 se encuentra estabilizando estrechamente al ligando unido mediante un puente de hidrógeno.

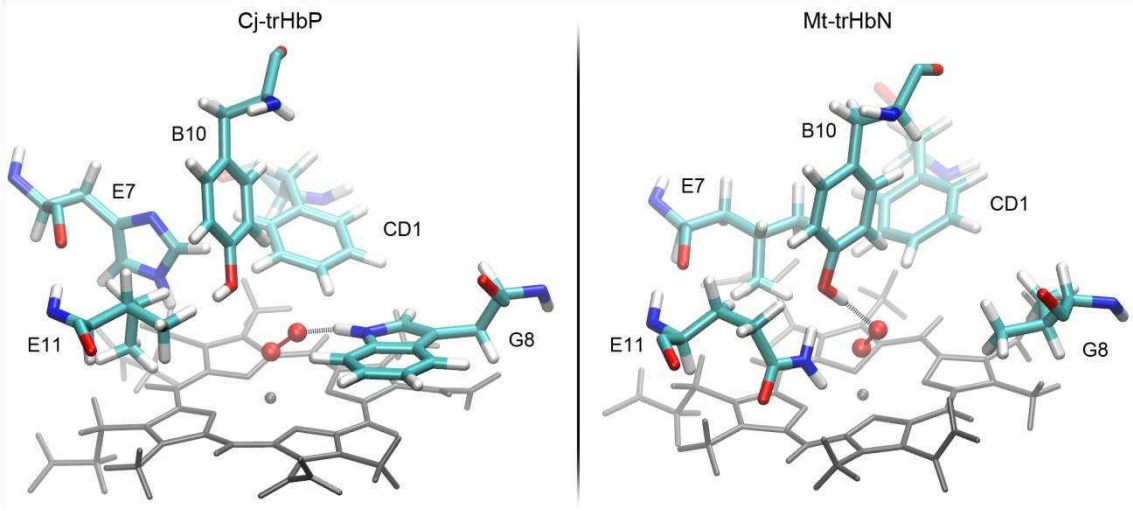


Figura I.3. Representación esquemática de los sitios activos de Cj-trHbP (panel izquierdo) y Mt-trHbN (panel derecho). Se muestran el grupo hemo (en gris), el O_2 (en rojo) unido mediante el enlace de sexta coordinación al átomo de Fe central del grupo hemo y los cinco aminoácidos claves del sitio distal. Las interacciones de puentes de hidrógeno para los ejemplos presentes se encuentran señalados con líneas punteadas.

I.5 Objetivos y lineamientos

El principal objetivo de esta tesis doctoral consistió en comprender cuáles son y cómo se modulan los determinantes moleculares de la captación y liberación de ligandos en trHbs. Se comenzó estudiando los determinantes moleculares involucrados en estos procesos en dos de los miembros de esta familia, la trHb O de *Thermobifida Fusca* (Tf-trHbO) y la trHb N de *Mycobacterium tuberculosis* (Mt-trHbN). Posteriormente se diseñó un modelo fisicoquímico que, incluyendo relaciones cuantitativas entre estos determinantes moleculares, permitió estimar cuantitativamente las constantes cinéticas experimentales de k_{on} y k_{off} de O_2. Con este poder predictivo, se abordó un estudio evolutivo-molecular de toda la familia de trHbs, y se

asignó a cada miembro de la familia una posible función en base a su capacidad para captar y liberar O_2.

Dado que uno de los modelos de trHb elegidas corresponde a la Tf-trHbO, una proteína termoestable, se realizó un estudio preliminar de las bases moleculares de su termoestabilidad. Se espera que sea un punto de partida para posteriores estudios que profundicen sobre las bases moleculares de la termoestabilidad en estas proteínas.

Para alcanzar estos objetivos propuestos, se aplicaron diversas técnicas computacionales de muestreo y análisis basados en dinámica molecular clásica, cálculos híbridos cuántico-clásicos, construcción de modelos matemáticos, análisis y alineamiento de secuencias, construcción de árboles filogenéticos y análisis evolutivos, desarrollados en detalle en el capítulo correspondiente a métodos (II).

1.6 Referencias

(1) Milani, M.; Pesce, A.; Nardini, M.; Ouellet, H.; Ouellet, Y.; Dewilde, S.; Bocedi, A.; Ascenzi, P.; Guertin, M.; Moens, L.; Friedman, J. M.; Wittenberg, J. B.; Bolognesi, M. *J. Inorg. Biochem.* **2005**, *99* (1), 97–109.

(2) Capece, L.; Boechi, L.; Perissinotti, L. L.; Arroyo-Mañez, P.; Bikiel, D. E.; Smulevich, G.; Marti, M. A.; Estrin, D. A. *Biochimica et Biophysica Acta - Proteins and Proteomics*. **2013**, *1834* (9), 1722-1738.

(3) Vinogradov, S. N.; Tinajero-Trejo, M.; Poole, R. K.; Hoogewijs, D. *BBA - Proteins Proteomics* **2013**, *1834* (9), 1789–1800.

(4) Ouellet, H.; Ouellet, Y.; Richard, C.; Labarre, M.; Wittenberg, B.; Wittenberg, J.; Guertin, M. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (9), 5902–5907.

(5) Milani, M.; Pesce, A.; Ouellet, Y.; Ascenzi, P.; Guertin, M.; Bolognesi, M. *EMBO J.* **2001**, *20* (15), 3902–3909.

(6) Kendrew, J. C.; Dickerson, R. E.; Strandberg, B. E.; Hart, R. G.; Davies, D. R.; Phillips, D. C.; Shore, V. C. *Nature* **1960**, *185* (4711), 422–427.

(7) Muirhead, H.; Perutz, M. F. *Nature* **1963**, *199* (4894), 633–638.

(8) Hou, S.; Freitas, T. A. K.; Larsen, R.; Piatibratov, M.; Sivozhelezov, V.; Yamamoto, A.; Meleshkevitch, E.; Zimmer, M.; Ordal, G.; Alam, M. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (16), 9353–9358.

(9) Wittenberg, J. B.; Bolognesi, M.; Wittenberg, B. a; Guertin, M. *J. Biol. Chem.* **2002**, *277* (2), 871–874.

(10) Vinogradov, S. N.; Hoogewijs, D.; Bailly, X.; Arredondo-Peter, R.; Gough, J.; Dewilde, S.; Moens, L.; Vanfleteren, J. R. *BMC Evol. Biol.* **2006**, *6*, 31.

(11) Vuletich, D. a; Lecomte, J. T. J. *J. Mol. Evol.* **2006**, *62* (2), 196–210.

(12) Goldbeck, R. a; Bhaskaran, S.; Ortega, C.; Mendoza, J. L.; Olson, J. S.; Soman, J.; Kliger, D. S.; Esquerra, R. M. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (5), 1254–1259.

(13) Olson, J.; Phillips, G. *J. Biol. Inorg. Chem.* **1997**, *2*, 544–552.

(14) Ouellet, Y.; Milani, M.; Couture, M.; Bolognesi, M.; Guertin, M. *Biochemistry* **2006**, *45* (29), 8770–8781.

(15) Martí, M. a; González Lebrero, M. C.; Roitberg, A. E.; Estrin, D. a. *J. Am. Chem. Soc.* **2008**, *130* (5), 1611–1618.

(16) Pond, M. P.; Majumdar, A.; Lecomte, J. T. J. *Biochemistry* **2012**, *51* (29), 5733–5747.

(17) Arroyo-Mañez, P.; Bikiel, D. E.; Boechi, L.; Capece, L.; Di Lella, S.; Estrin, D. A.; Martí, M. A.; Moreno, D. M.; Nadra, A. D.; Petruk, A. A. *Biochim. Biophys. Acta - Proteins Proteomics* **2011**, *1814* (8), 1054–1064.

(18) Boechi, L.; Martí, M. A.; Milani, M.; Bolognesi, M.; Luque, F. J.; Estrin, D. A. *Proteins Struct. Funct. Genet.* **2008**, *73* (2), 372–379.

(19) Boechi, L.; Mañez, P. A.; Luque, F. J.; Marti, M. a; Estrin, D. a. *Proteins* **2010**, *78* (4), 962–970.

(20) Forti, F.; Boechi, L.; Estrin, D. A.; Marti, M. A. *J. Comput. Chem.* **2011**, *32* (10), 2219–2231.

(21) Arroyo-Mañez, P.; Bikiel, D. E.; Boechi, L.; Capece, L.; Di Lella, S.; Estrin, D. a; Martí, M. a; Moreno, D. M.; Nadra, A. D.; Petruk, A. a. *Biochim. Biophys. Acta* **2011**, *1814* (8), 1054–1064.

(22) Pesce, A.; Milani, M.; Nardini, M.; Bolognesi, M. *Methods Enzymol.* **2008**, *436* (08), 303–315.

(23) Milani, M.; Pesce, A.; Ouellet, Y.; Dewilde, S.; Friedman, J.; Ascenzi, P.; Guertin, M.; Bolognesi, M. *J. Biol. Chem.* **2004**, *279* (20), 21520–21525.

(24) Perutz, M. F.; Mathews, F. S. *J. Mol. Biol.* **1966**, *21* (1), 199–202.

(25) Scott, E. E.; Gibson, Q. H.; Olson, J. S. *J. Biol. Chem.* **2001**, *276*, 5177–5188.

II. Métodos generales de simulación computacional

Aquí se presentarán los métodos que serán utilizados en diversos capítulos de la presente tesis, los relacionados con el muestreo y análisis de sistemas de biomoléculas basados en dinámica molecular clásica y cálculos híbridos cuántico-clásicos. Aquellos métodos empleados para la resolución de problemáticas particulares, como la construcción de modelos matemáticos, análisis y alineamiento de secuencias, construcción de árboles filogenéticos y análisis evolutivos, serán descriptos en detalle en los capítulos correspondientes a su utilización.

Al utilizar métodos de simulación computacional es importante tener presente que éstos se basan en modelos, es decir, en representaciones (aproximaciones) del mundo real que tienen características y limitaciones que los definen y diferencian entre sí, principalmente debido al nivel de aproximación de cada uno. Es preciso entonces identificar primero cuál de estos métodos es el indicado para dar respuesta a los interrogantes fenomenológicos que se planteen. Por un lado, los modelos más detallados, provenientes de la química cuántica (QM, por sus siglas en inglés, *Quantum Mechanics*) consideran de manera explícita la densidad electrónica de los átomos intervinientes en un sistema, por lo que permiten estudiar y resolver problemas referentes a mecanismos de reacciones químicas (formación y ruptura de enlaces covalentes), transferencia de cargas, estados electrónicamente excitados, espectroscopías IR, Raman, RMN, y UV. Por otro lado, los modelos de mecánica molecular o clásica (MM, por sus siglas en inglés, *Molecular Mechanics*), caracterizados por un nivel de detalle menor, consideran a los átomos como esferas rígidas con cargas puntuales, por lo que ya no describen cambios en la distribución electrónica. Estos modelos MM son utilizados para resolver problemas referentes a la dinámica de sistemas de biomoléculas (proteínas, lípidos, carbohidratos, ácidos nucleicos) dependiente del tiempo y del entorno, cambios conformacionales pequeños, difusión de ligandos, determinación de factores moleculares que intervienen en una función biológica particular, interacciones proteína-proteína, proteína-ligandos, proteína-inhibidores, proteína-ADN.

Dado que los modelos MM describen a los sistemas de manera mucho más aproximada que los modelos QM, el costo computacional asociado a la simulación de sistemas con un modelo u otro varia inmensamente. Mientras que en general suelen estudiarse sistemas QM de un tamaño aproximado de entre 50 y 150 átomos con tiempos de simulación de la escala de

decenas a cientos de ps, los sistemas MM suelen abarcar hasta 3 órdenes de magnitud más (150.000 átomos), y tiempos de la escala de cientos de nanosegundos.

En las secciones siguientes se hará una breve descripción sobre cómo funcionan los métodos cuánticos, clásicos e híbridos cuántico-clásicos, seguido de técnicas de muestreo de la superficie de energía potencial de los sistemas.

II.1 Métodos basados en la mecánica cuántica (QM)

Cuando se desea predecir las propiedades de sistemas moleculares microscópicos, considerando la distribución electrónica de un sistema, se deben resolver las ecuaciones de la mecánica cuántica, que tienen como ecuación fundamental a la ecuación de Schrödinger[1]:

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

Ésta es una ecuación diferencial cuya solución consiste en obtener la función de onda, que da la descripción matemática de un sistema de partículas cuánticas, sujetas al potencial dado. La función de onda contiene toda la información necesaria para calcular las propiedades del sistema.

El objetivo básico de la química cuántica es resolver la ecuación de Schrödinger para un sistema molecular. Resulta imposible resolver de forma exacta esta ecuación para un sistema molecular, y se deben buscar soluciones aproximadas. El primer paso en el camino a la solución consiste en suponer que el potencial es independiente del tiempo (lo cual es cierto si no hay fuerzas externas). A partir de esto se deduce una ecuación independiente del tiempo, cuya solución solo depende de las coordenadas espaciales de las partículas, que además está unívocamente relacionada con la energía del sistema. Si se define ahora al operador hamiltoniano como:

$$\hat{\mathrm{H}} = -\frac{\hbar^2}{2m}\nabla^2 + V \qquad \text{(II.2)}$$

la ecuación de Schrödinger se reduce a:

$$\hat{\mathrm{H}}\Psi(r) = E\Psi(r) \qquad \text{(II.3)}$$

La ecuación de Schrödinger es, por lo tanto, una ecuación diferencial en derivadas parciales de autovalores, en donde un operador actúa sobre la autofunción, obteniendo como resultado a la función multiplicada por un escalar. La segunda simplificación que se debe tener en cuenta es la denominada aproximación de Born-Oppenheimer, que consiste en desacoplar el movimiento de los núcleos del de los electrones. Este procedimiento se justifica ya que la velocidad relativa de los núcleos con respecto a la de los electrones resulta ser despreciable. Así, se asume que los electrones se mueven en un potencial de cargas positivas estáticas dadas por

las posiciones nucleares. El problema queda ahora reducido a obtener la función de onda para el siguiente operador denominado hamiltoniano electrónico, para un sistema compuesto por M núcleos y N electrones:

$$\hat{\mathrm{H}}_{el} = -\sum_{i=1}^{N} \frac{1}{2}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M} \frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N} \frac{1}{r_{ij}} \qquad \text{(II.4)}$$

En este operador se pueden distinguir tres términos, el primero representa la energía cinética electrónica, el segundo es el potencial asociado a la interacción núcleo-electrón y el tercero es el potencial de la interacción electrón-electrón. Si bien con las aproximaciones mencionadas se ha simplificado el problema, aún no se obtuvo un método que permita resolver la ecuación de Schrödinger. La clave para resolver de manera aproximada ecuaciones diferenciales de este tipo es utilizar el principio o teorema variacional. Éste postula que al evaluar la energía con una función de prueba (diferente de aquella que es solución exacta de la ecuación que no se conoce), la energía que se obtendrá será siempre mayor a la energía real (la correspondiente a la función de onda que es solución exacta) del sistema, o sea:

$$< \Psi'(r)|\hat{\mathrm{H}}|\Psi'(r) \geq E_0 \qquad \text{(II.5)}$$

donde E_0 representa la energía real del sistema que corresponde a evaluar:

$$< \Psi(r)\big|\hat{\mathrm{H}}\big|\Psi(r) = E_0 \qquad \text{(II.6)}$$

y la función de onda es la función de onda real del sistema, solución exacta de la ecuación de Schrödinger para el estado fundamental. Esto permite proponer como estrategia para encontrar soluciones aproximadas de la ecuación de Schrödinger, la búsqueda de aquellas funciones que minimicen la energía del sistema. Dado el operador hamiltoniano, existe un conjunto infinito de soluciones (cada una representada por una autofunción y su autovalor) que se pueden ordenar de manera creciente según:

$$\hat{\mathrm{H}}|\Psi_i\rangle = \varepsilon_i|\Psi_i\rangle \qquad i = 0,1,2,\ldots \qquad \varepsilon_0 \leq \varepsilon_1 \leq \ldots \leq \varepsilon_i$$

II.1.1 Teoría del Funcional de la Densidad (DFT)

Un formalismo que se ha popularizado en los últimos años debido a su eficiencia computacional y a la calidad de los resultados obtenidos, se conoce como teoría del funcional de la densidad (DFT)[2]. La teoría de los funcionales de la densidad se basa en los teoremas de Hohenberg y Kohn formulados en los años 60 que permiten describir el problema electrónico usando la densidad de carga electrónica. El primero de los teoremas establece que la función de onda, y por ende cualquier propiedad del sistema, se halla unívocamente determinada por la densidad electrónica del sistema.

En el contexto de DFT la energía E en función de la densidad está dada por:

$$E[\rho] = T[\rho] + V_{ee}[\rho] + V_{ne}[\rho] \quad \text{(II.7)}$$

donde los términos corresponden a la energía cinética (T), la interacción electrón-electrón (V_{ee}) y la interacción núcleo-electrón (V_{ne}). Usando diferentes definiciones de cada uno de los términos anteriores, y utilizando la aproximación de Kohn y Sham que considera un sistema de referencia de electrones no interactuantes, la fórmula final para el funcional de la densidad puede escribirse como:

$$E[\rho] = T_s[\rho] + V_{ne}[\rho] + \frac{1}{2}\iint \frac{\rho(r_1)\rho(r_2)}{r_{12}} dr_1 dr_2 + E_{xc}[\rho] \quad \text{(II.8)}$$

Con esta definición se puede calcular la energía para todos los términos, excepto por la relativamente pequeña contribución del término de intercambio y correlación (E_{xc}), cuya forma funcional no es conocida. La teoría del funcional de la densidad, sumado a la aproximación de Kohn y Sham se basa en un funcional exacto por definición, donde los efectos debidos a la naturaleza cuántica del electrón, el intercambio y la correlación son descriptos por el término E_{xc}. Sin embargo, como la dependencia funcional de E_{xc} con la densidad es desconocida, esto implica que la precisión y calidad de un cálculo basado en DFT estará íntimamente relacionada con la aproximación utilizada para darle forma al funcional de intercambio y correlación. Existen muchos funcionales de intercambio y correlación. Los más utilizados en sistemas bioquímicos son: BLYP, B3LYP, PBE[2–4].

II.2 Métodos basados en la mecánica clásica (MM)

II.2.1 Campo de fuerzas clásico

El campo de fuerzas es una expresión de la energía potencial dependiente de las coordenadas de los núcleos y una serie de parámetros[5]. La expresión del campo de fuerzas de AMBER[6,7], el cual fue utilizado para los cálculos realizados en este trabajo, contiene dos clases de contribuciones, las contribuciones de unión y las de no unión. Las de unión se computan para átomos formando un enlace covalente o como máximo a dos átomos de distancia, e incluyen penalidades 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 por una función periódica. Las contribuciones de no unión se computan entre átomos que están en diferentes moléculas o a más de 3 enlaces de distancia, y se dividen en interacciones electrostáticas basadas en la ley de Coulomb y en un potencial de 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:

$$E(r^N) = \sum_{enlaces} \frac{k_{ri}}{2}(r_i - r_0)^2 + \sum_{ángulos} \frac{k_{\theta i}}{2}(\theta_i - \theta_0)^2 + \sum_{torciones} \frac{V_n}{2}(1 + \cos(n\omega - \gamma)) + \sum_{i=1}^{N}\sum_{j=i+1}^{N}(4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}) \quad \text{(II.9)}$$

El primer término refiere a la contribución de estiramiento de los enlaces, el cual se describe a través de un potencial armónico, donde k_{ri} corresponde a la constante de fuerza asociada a la unión *i-ésima* y r_0 a la distancia de equilibrio de la misma unión. El potencial armónico resulta una buena y sencilla aproximación al potencial real en la zona cercana al mínimo. El segundo término corresponde a las contribuciones dadas por las flexiones angulares. Éstas también se representan con un potencial armónico de constante $k_{\theta i}$, y valor de equilibro θ_0. El término asociado a las torsiones o ángulos diedros presenta tres parámetros: V_n corresponde a la barrera energética asociada a la rotación entre dos mínimos sucesivos, n representa la multiplicidad del potencial, es decir, el número de mínimos encontrados entre 0 y 360°, y por último γ corresponde a la fase de la función sinusoidal. El último término de la ecuación (II.9) 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, contiene los parámetros ε_{ij} y σ_{ij}, asociados a la profundidad del pozo de energía y a la distancia entre los átomos correspondiente al mínimo de energía, respectivamente. La parte electrostática tiene como parámetros las cargas sobre los átomos. En el campo de fuerzas de AMBER, la distribución de cargas se representa a través de cargas puntuales q_i ubicadas en las coordenadas de cada átomo. Estas cargas se obtienen de manera de reproducir el potencial electrostático de la molécula, calculado a partir de cálculos cuánticos.

II.2.2 Modelos para el solvente

La gran mayoría de los fenómenos químicos de interés ocurren en solución acuosa; no es de extrañar entonces el esfuerzo que químicos y físicos han hecho para desarrollar modelos que permitan simular esta vital sustancia.

Dentro de los componentes utilizados para modelar al agua se encuentran fluidos dieléctricos, cargas puntuales fijas, cargas fluctuantes, dipolos puntuales, potenciales tipo Lennard-Jones, potenciales armónicos de enlace, entre otros. La utilización de dichos componentes, así como la parametrización de los mismos, dependen en buena medida de qué propiedad interesa reproducir y con qué nivel de precisión en relación al costo computacional asociado. A continuación se describirá uno de los modelo de solvente explícito más conocidos, el cual fue utilizado en esta tesis para describir la solvatación en diferentes simulaciones.

II.2.2.1 Solvente explícito: TIP3P

Modelar de forma explícita el solvente consiste en considerar a cada una y todas las moléculas de agua que constituyen el sistema de forma explícita, considerando su estructura molecular con el grado de descripción que el modelo imponga. En las simulaciones clásicas en las que estos modelos explícitos son ampliamente empleados, la estructura molecular del agua está dada por coordenadas de los núcleos de los átomos que la forman, distancias y ángulos de equilibrio y cargas puntuales, es decir, parámetros. En la presente tesis se ha utilizado el modelo explícito TIP3P, el cual presenta tres sitios interactuantes, que corresponden a las posiciones de los núcleos de la molécula de agua. A cada átomo se le asigna una carga puntual (-0,834 e para el oxígeno, y +0,417 e para cada uno de los átomos de hidrógeno), y el átomo de oxígeno toma además los parámetros de Lennard-Jones (σ = 3,15061 Å, ε= 636,4 $J\ mol^{-1}$). Las distancias interatómicas O-H se fijan en 0,9572 Å, al igual que el ángulo HOH, en 104,52°, correspondiéndose con las observadas experimentalmente[5,8].

II.2.3 Condiciones periódicas de contorno y radio de corte

Al realizar una simulación con solvente explícito hay una cantidad finita de moléculas de solvente en un sistema dado, por este motivo, si el soluto de interés sufre traslaciones y/o cambios conformacionales posiblemente al menos una parte del mismo quede sin solvatar de manera adecuada (figura II.1).

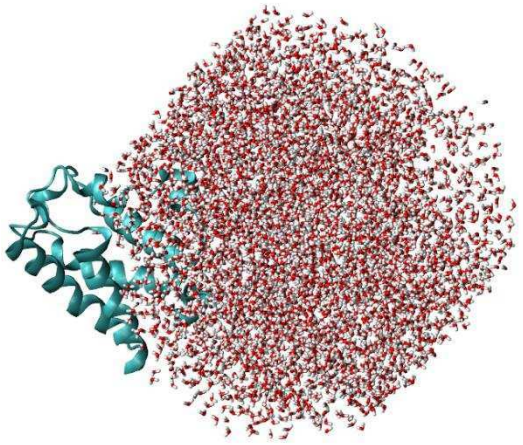


Figura II.1. Representación esquemática de un sistema sin condiciones periódicas de borde. Puede observarse que la macromolécula ha quedado parcialmente fuera de la caja de solvente debido a traslaciones durante la simulación.

Estableciendo condiciones periódicas de contorno el sistema se ubica dentro de una celda y se considera que la misma se repite infinitas veces en el espacio tridimensional. El objetivo de este esquema es lograr que cualquier partícula dentro de esta caja tenga el mismo entorno que el que tendría en el seno de un sistema macroscópico. De esta manera, si la proteína o molécula de interés lograse "escapar" del solvente en el cual está inmersa, se encontrará con el solvente presente en la celda contigua y permanecerá solvatada (figura II.2).

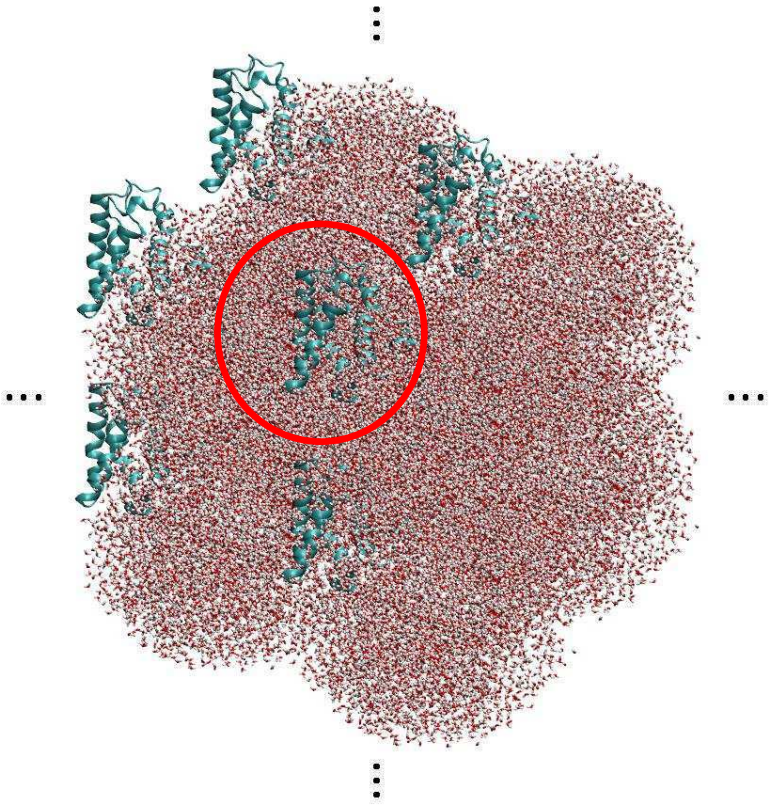


Figura II.2. Representación esquemática de un sistema con condiciones periódicas de borde. La celda central se repite en todas las direcciones. A pesar de haber sido desplazada parcialmente de su caja de solvente, puede apreciarse que la proteína (resaltada en rojo) se encuentra inmersa en moléculas de agua de su propia caja y de cajas contiguas.

El problema pasa a ser cómo calcular las interacciones, ya que si el sistema es infinito, las interacciones también lo son. Para lidiar con esta suma infinita de términos se utilizan dos estrategias diferentes. Para los términos de no unión, se utiliza un radio de corte, tal que para los átomos que se encuentran a una distancia mayor que este radio sus interacciones no se calculan. En el caso de las interacciones de Lennard-Jones, al ser de corto alcance decaen mucho más rápido ($1/r^6$) de lo que crece el número de interacciones (r^2), y esto no genera un problema siempre que el radio de corte sea lo suficientemente grande. En el caso de las interacciones Coulómbicas, el problema es más complejo ya que este tipo de interacciones decrece como la inversa de las distancia, resultando de mayor alcance.

Para resolver esta situación, se utiliza una metodología denominada sumas de Ewald[9], un método originalmente desarrollado para el estudio de cristales iónicos, que en su implementación se denominaba *Particle Mesh Ewald*[10,11]. 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. Y en una segunda etapa, 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[12].

II.2.4 Parámetros asociados

Para definir un campo de fuerzas, además de la expresión de la energía potencial se requiere una gran cantidad de parámetros asociados a cada uno de los sumandos de la ecuación. Para el caso de simulación de proteínas, el campo de fuerzas de AMBER contiene los parámetros asociados a los aminoácidos naturales (en sus distintos estados de protonación, según corresponda) presentes en la mayoría de las proteínas. En este contexto, puede hablarse de la transferibilidad del campo de fuerza, es decir, el hecho de que el mismo conjunto de parámetros puede utilizarse para moléculas asociadas. Por ejemplo, los parámetros para un aminoácido de tirosina serán los mismos en todas las proteínas estudiadas, y no será necesaria su parametrización en cada proteína que se desee estudiar. En el campo de fuerzas de AMBER los átomos dentro de cada aminoácido reciben un nombre de átomo y un tipo de átomo. Los tipos de átomo permiten que no sea necesario incluir parámetros para cada átomo incluido en cada uno de los aminoácidos. Distintas especies pueden poseer el mismo tipo de átomo, y los parámetros asociados a las uniones, los ángulos y los diedros se dan por tipo de átomo, en lugar de para cada átomo en particular. Si bien en el campo de fuerzas se incluyen parámetros para todos los aminoácidos naturales y otras moléculas comúnmente encontradas en los sistemas proteicos, para incluir grupos prostéticos o azúcares o sus sustituyentes, por ejemplo, será necesario obtener los parámetros necesarios para realizar el cálculo.

II.3 Métodos híbridos cuántico-clásicos (QM/MM)

Hasta aquí se han explicado brevemente los fundamentos teóricos de los métodos cuánticos (QM) y los métodos clásicos (MM), así como sus características y limitaciones asociadas. Los

métodos híbridos QM/MM permiten estudiar reacciones químicas en sistemas complejos como las proteínas o complejos macromoleculares. Estos métodos describen a una porción del sistema (un soluto o un sitio activo de una enzima, por ejemplo) mediante métodos de estructura electrónica (subsistema cuántico) y al resto del entorno (macromolécula y/o solvente) por métodos clásicos (subsistema clásico). El hamiltoniano total y la energía del sistema tienen tres contribuciones:

$$E_{TOT} = E_{QM} + E_{MM} + E_{QM-MM} \quad \text{(II.10)}$$

En este trabajo de tesis, se utilizó el programa *Hybrid* que ha sido desarrollado en el grupo del Dr. Estrin[13]. Este programa calcula la contribución cuántica a nivel de DFT, teniendo en cuenta sólo los átomos cuánticos, pero considerando a los electrones inmersos en el potencial generado no sólo por los núcleos de éstos, sino además por el potencial generado debido a las cargas parciales de los átomos clásicos. La contribución clásica es calculada con el campo de fuerzas de AMBER sin modificaciones. El término crítico es el del acoplamiento cuántico-clásico. Este término también está compuesto por tres contribuciones y se calcula de acuerdo a:

$$E_{QM-MM} = \sum_{i=1}^{C} q_i \int \frac{\rho(r)}{|r-\tau_i|} dr + \sum_{i=1}^{C} \sum_{\alpha=1}^{Q} \frac{q_i Z_\alpha}{|R_\alpha - \tau_i|} + E_{QM-MM}^{LJ} \quad \text{(II.11)}$$

Siendo C el número de átomos clásicos con cargas parciales q_i y Q el número de átomos cuánticos con índice i, r es la coordenada electrónica, τ_i es la posición del núcleo cuántico y R_α es la posición del núcleo clásico. El primer término representa la interacción entre la densidad electrónica cuántica y las cargas parciales de los átomos clásicos. El segundo representa la interacción entre los núcleos de los átomos cuánticos y las cargas parciales clásicas. Estos dos términos representan la interacción electrostática entre los dos subsistemas. El tercer término representa la interacción de Van Der Waals y repulsiva a corto alcance entre los átomos de ambos sistemas y se calcula usando un potencial de Lennard-Jones idéntico al usado en el campo de fuerzas clásico.

II.4 Exploración de la superficie de energía potencial

La superficie de energía potencial (SEP), definida en el contexto de la aproximación de Born-Oppenheimer[1,2], es una función escalar que define el valor de la energía potencial del sistema en función de las coordenadas de los núcleos. Por ejemplo, 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. Para sistemas sistemas grandes (como por ejemplo una proteína) la SEP resulta muy compleja y es imposible conocerla en su totalidad. En esta sección se describirán distintos métodos desarrollados para explorar la SEP de forma eficiente, que

permiten concentrar la atención en el área de la SEP que se desea explorar. Si bien existe una gran cantidad de metodologías para realizar esta tarea, se desarrollarán solamente aquellas aplicadas en los cálculos realizados en este trabajo.

II.4.1 Minimización de la energía

En los mínimos de la SEP se encuentran las configuraciones o conformaciones estables. Para encontrarlos, debe aplicarse un algoritmo de minimización. Debido a la complejidad con la que varía la SEP con las coordenadas cartesianas, se utilizan métodos numéricos para obtener los mínimos de la función. Dentro de los diferentes métodos, pueden distinguirse los métodos que usan las derivadas de la función a minimizar, y aquellos que no. Dentro de los primeros, existen métodos que utilizan solo la derivada primera (que es nula en la posición del mínimo), llamados de primer orden, y los que utilizan además las derivadas segundas (deben ser positivas), llamados de segundo orden. Los métodos utilizados en este trabajo corresponden a métodos de derivadas de primer orden, en particular los métodos llamados *Conjugate Gradients* y *Steepest Descents*[5]. 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 movimientos en dirección paralela a la fuerza neta con 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 un incremento de la energía se procede a reducir la longitud del paso.

El método *Conjugate Grandients*, por su parte, se diferencia del método de *Steepest Descents* en que la dirección en la que se realiza 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 apuntan siempre ir a la reducción de la energía, por lo que el resultado final dependerá fuertemente de la configuración de partida. En otras palabras, simplemente a través de optimizaciones de geometría como las descriptas 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.4.2 Simulaciones de dinámica molecular

En los casos de la presente tesis, la complejidad de la superficie de energía potencial involucra la existencia de gran cantidad de mínimos locales. Esto implica que no existe un solo mínimo representativo de cada sistema, sino que existen una serie de conformaciones, que estarán presentes en mayor o menor proporción según sus energías relativas.

Cuando se realiza una medida experimental, todas las conformaciones relevantes estarán representadas en el ensamble de moléculas presente en el sistema. Una manera de representar este ensamble es a través de la reconstrucción de la evolución temporal de la dinámica del sistema, asumiendo que el sistema ha evolucionado un tiempo suficiente para que haya podido explorar las distintas conformaciones disponibles (convergencia de la simulación). La hipótesis ergódica[5] postula que el promedio temporal puede reemplazarse por el promedio sobre las configuraciones de un sistema dado, dando origen así a las simulaciones basadas en evolución temporal, diferenciándose de aquellas que buscan explorar el colectivo de configuraciones espaciales. En una simulación de dinámica molecular (DM), se realiza la reconstrucción de la evolución temporal de la dinámica de un sistema a través de la resolución de las ecuaciones de movimiento de Newton. En otras palabras, la trayectoria se obtiene a través de la resolución de las ecuaciones diferenciales asociadas a:

$$\frac{d^2x_i}{dt^2} = \frac{F_{xi}}{m_i} \qquad \text{(II.12)}$$

donde se describe el movimiento de una partícula de masa m_i en la coordenada x_i, con F_{x_i} la fuerza en esa dirección.

La fuerza sobre cada una de las partículas cambia cuando la partícula modifica su posición, y es función del movimiento de todas las otras partículas. La resolución analítica del problema no es posible, por lo que debe recurrirse a métodos numéricos. El método implementado en el programa AMBER se denomina algoritmo de Verlet[14], basado en el método de las diferencias finitas. En este método, 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 a partir 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 estratégicamente las expansiones de Taylor de las posiciones a diferentes tiempos, según:

$$r(t+\delta t) = r(t) + v(t)dt + \frac{a(t)}{2}\delta t^2 + \cdots \quad \text{(II.13)}$$

$$r(t-\delta t) = r(t) - v(t)dt + \frac{a(t)}{2}\delta t^2 + \cdots \quad \text{(II.14)}$$

Sumando miembro a miembro las expresiones II.13 y II.14, y despreciando los términos de orden superior al segundo, se obtiene:

$$r(t+\delta t) \approx 2r(t) - r(t-\delta t) + a(t)\delta t^2 \quad \text{(II.15)}$$

Las velocidades, necesarias para la estimación de la energía cinética, no aparecen explícitamente, pero pueden obtenerse de la resta de las mismas expresiones:

$$v(t) \approx \frac{r(t+\delta t) - r(t-\delta t)}{2\delta t} \quad \text{(II.16)}$$

Debe prestarse especial atención a la elección del intervalo de tiempo δt. Cuanto menor sea éste, mayor será la cantidad de pasos necesarios para realizar una simulación de una dada cantidad de tiempo, con lo que incrementará el costo computacional. Sin embargo, un intervalo de tiempo demasiado grande dará origen a inestabilidades en el algoritmo de integración. Por otra parte, debe elegirse un δt que sea menor a 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 de manera clásica, los movimientos más rápidos están asociados a las vibraciones de los enlaces, que se encuentran en el orden de los fs (1 fs equivale a $10^{-15}\ s$). 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 δt y así acelerar los cálculos consiste en utilizar el algoritmo SHAKE[15], que permite fijar estas uniones en sus posiciones de equilibrio.

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

$$K = \sum_{i=1}^{N} \frac{m_i v_i^2}{2} = \frac{3k_B T}{2} N \quad \text{(II.17)}$$

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. Sin embargo, este método no refleja la naturaleza fluctuante de la temperatura alrededor de un promedio, y por ello no resulta ser el mejor de los enfoques de control térmico de las simulaciones.

Uno de los métodos más utilizados para regular la temperatura en una simulación de DM es el algoritmo de Berendsen[16]. 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.17 que si las velocidades se escalan por un factor λ, el cambio asociado en la temperatura resulta:

$$\Delta T = (\lambda^2 - 1)T(t) \qquad \text{(II.18)}$$

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:

$$\frac{dT(t)}{dt} = \frac{1}{\tau}[T_{baño} - T(t)] \qquad \text{(II.19)}$$

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 δt, corresponde a:

$$\Delta T = \frac{\delta t}{\tau}[T_{baño} - T(t)] \qquad \text{(II.29)}$$

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

$$\lambda^2 = 1 + \frac{\delta t}{\tau}\left[\frac{T_{baño}}{T(t)} - 1\right] \qquad \text{(II.30)}$$

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 δt elegido otorga buenos resultados para los sistemas estudiados como en esta tesis.

Por otra parte, es también común realizar simulaciones de DM 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[16]. El volumen de la celda se escala entonces en un factor γ, lo que equivale a escalar las coordenadas atómicas en $\gamma^{1/3}$. Resulta entonces:

$$\gamma = 1 + \kappa\frac{\delta t}{\tau_P}[P - P_{baño}] \qquad \text{(II.31)}$$

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

Luego de realizar una simulación de DM, el resultado principal que se obtiene es la trayectoria de cada uno de los átomos que conforman el sistema a lo largo del tiempo simulado. La visualización de la misma a través de algún programa de representación gráfica aporta gran cantidad de información sobre el sistema. Con ella se pueden analizar la evolución de parámetros geométricos claves, como distancias, ángulos, ángulos diedros y otros parámetros

de relevancia estructural como 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 este trabajo fue el *Visual Molecular Dynamics* (VMD), el cual presenta una gran cantidad de herramientas muy útiles para el análisis de una simulación de DM[17].

RMSD y RMSF. Un parámetro que permite estudiar la estabilidad de la estructura a lo largo de la simulación es la raíz cuadrada de la desviación cuadrática media (RMSD, por sus siglas en inglés, *Root Mean Squared Deviation*) respecto a una estructura de referencia, la cual se calcula de la siguiente manera:

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\delta_i^2} \qquad \text{(II.32)}$$

En la expresión II.32, 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 será la estructura de referencia. Realizando este cálculo para cada uno de los eventos de la simulación, se obtiene el RMSD en función del tiempo. La estructura de referencia puede ser la estructura cristalográfica, por ejemplo, o una estructura promedio, lo que otorga una noción de la magnitud de las fluctuaciones del sistema a lo largo de la dinámica, es decir, como una función del tiempo.

Si en este caso el promedio es sobre el tiempo, se puede tener una noción de las fluctuaciones, ahora por partícula:

$$RMSF = \sqrt{\frac{1}{T}\sum_{t_j=1}^{T}[x_i(t_j) - \tilde{x}_i]^2} \qquad \text{(II.33)}$$

donde T es el tiempo total de promedio y $\tilde{x}_i$ es la posición de referencia de la partícula i. Esta es la expresión para calcular la raíz cuadrada de la fluctuación cuadrática media (RMSF), la cual resulta muy útil para comparar las magnitudes de las fluctuaciones de cada partícula del sistema en un rango de tiempo de simulación determinado, como una función de cada partícula. En general, en el caso de proteínas, se calculan tanto el RMSD como el RMSF tomando como referencia las posiciones de los átomos de C_α del esqueleto proteico.

II.4.3 Detalles técnicos. Configuración de los sistemas y parámetros de simulación

En todos los casos, se asumió que los estados de protonación de los aminoácidos corresponden al pH fisiológico (Asp y Glu negativamente cargados, Lys y Arg positivamente

cargados), todas las His expuestas al solvente fueron protonadas en el átomo N-δ, así como la HisF8, que está coordinada al Fe del grupo hemo. Puesto que distintos estados de protonación de una His presente en la posición estructural CD1 pueden causar diversos patrones de redes de puentes de hidrógeno, en estos casos el estado de protonación fue cuidadosamente elegido basado en dos aspectos principales: i) estructuras cristalográficas que sugieren un patrón de red de puentes de hidrógeno dado[18,19] y ii) correlación entre constantes de disociación experimentales y calculadas para aquellos casos en los cuales se disponga de información experimental. Una vez completos, los sistemas fueron inmersos en una caja octahédrica de moléculas de agua pre equilibrada con aproximadamente ~4900 moléculas de agua TIP3P, configurando la distancia mínima entre la proteína y el extremo de la caja en 10Å.

Todos los parámetros usados para aminoácidos corresponden a los determinados por el campo de fuerzas ff99SB de AMBER[20], excepto para el grupo hemo que corresponden a los desarrollados[21] y ampliamente utilizados en varios estudios de hemoproteínas por parte del grupo de investigación en el cual se enmarca dicha tesis doctoral[22–28]. Todas las simulaciones se hicieron con condiciones periódicas de contorno con un radio de corte de 9Å y con el método de sumas de Ewald para tratar las interacciones electrostáticas. Las uniones covalentes involucrando átomos de hidrógeno se mantuvieron en sus distancias de equilibrio utilizando el algoritmo SHAKE, mientras que la temperatura y la presión se mantuvieron constantes con el termostato y baróstato de Berendsen, respectivamente, como está implementado en el paquete de AMBER12[7]. El protocolo de equilibración consiste en: i) calentamiento lento de todo el sistema de 0 a 27°C durante 20ps a volumen constante, con restricciones harmónicas de 80 Kcal por molÅ^2 para todos los átomos de C_α, (ii) equilibración de la presión para todo el sistema simulado durante 1ns a 27°C con los mismos átomos restringidos. Luego de estos dos pasos, se simularon dinámicas moleculares a temperatura constante de entre 50 y 200ns, dependiendo los casos de estudio (ver en cada capítulo de resultados).

Las proteínas mutantes *in silico*, mutantes de un único aminoácido o incluso dos o tres para algunas formas wt, se construyeron comenzando de las estructuras cristalinas extraídas de la base de datos PDB y modificadas luego usando el módulo tLEaP del paquete de AMBER12[7]. Estas estructuras de proteínas mutantes se equilibraron y simularon siguiendo el mismo protocolo usado que para las formas wt. Todas las estructuras presentadas se han mantenido estables durante la escala temporal de las simulaciones, como lo evidencia el análisis de RMSD (datos no mostrados).

II.4.4 Detalles técnicos. Energías de unión de O_2 (ΔE_{O_2} y $\Delta\Delta E_{O_2}$)

Para obtener las energías de unión se realizaron cálculos QM/MM del complejo proteína-O_2 unido. Las estructuras iniciales para estos cálculos se obtuvieron de simulaciones de DM. Se partió de fotos de dichas dinámicas basadas en un análisis de la estructura y dinámica de patrones de puentes de hidrógeno estabilizando al O_2 unido. La estructura obtenida correspondiente a una de esas fotos, por pertenecer a una DM a temperatura ambiente, normalmente suele ser enfriada lentamente a 0 K (-273°C), con el objetivo de relajar la estructura y llegar a un mínimo de energía de dicha conformación. Partiendo de la estructura congelada se llevaron a cabo optimizaciones de geometría con un esquema QM/MM usando un algoritmo de gradiente conjugado, a un nivel de teoría de DFT, utilizando la implementación del programa SIESTA (*Spanish Initiative for the Electronic Structure of Thousands of Atoms,* por sus siglas en inglés) desarrollado por el grupo de investigación en el cual se enmarca dicha tesis doctoral[13,22,23], con el funcional de intercambio y correlación PBE. Todos los cálculos fueron realizados usando el funcional de aproximación de gradiente generalizado propuesto por Perdew y otros[4].

Durante las corridas de QM/MM solo se les permitió moverse libremente a los aminoácidos localizados a menos de 10 Å del centro reactivo del grupo hemo. El subsistema de átomos cuánticos fue el integrado por el Fe, el ligando distal y el grupo imidazol de la histidina proximal (HF8). El resto de la unidad proteica, junto con las moléculas de agua, correspondieron al sistema clásico. La interfase entre los subsistemas cuántico (QM) y clásico (MM) se trató con el método del átomo unión (*link*, del inglés).

Las energías de unión de O_2 (ΔE_{O_2}, $kcal\ mol^{-1}$) se calcularon de la siguiente manera:

$$\Delta E_{O_2} = E_{Prot-O_2} - \left(E_{O_2} + E_{Prot}\right) \qquad \text{(II.34)}$$

donde E_{Prot-O_2} es la energía de la proteína oxigenada, E_{Prot} es la energía de la proteína desoxigenada y E_{O_2} es la energía de una molécula aislada de O_2.

Las proteínas oxigenadas se simularon en el estado singulete de spin, las desoxigenadas en el estado quinteto y el oxígeno libre en el triplete, siendo éstos los estados fundamentales conocidos para cada caso. Todas las simulaciones se llevaron a cabo con la aproximación de spin no restringida. Estas estrategias metodológicas han sido amplia y satisfactoriamente usadas por el grupo de investigación para estudiar afinidades por O_2 (así como también otros ligandos) en trabajos publicados con anterioridad[29].

Sin embargo, se sabe que los cálculos de ΔE_{O_2}, que también se pueden pensar como energías de disociación de oxígeno, son significativamente sobreestimados debido al hecho de que hay

involucrada una transición de bajo (singulete) a alto (quinteto) estado de spin, y el método de DFT sobreestima la energía del salto de spin, favoreciendo configuraciones de bajo estado energético[30]. Por otro lado, los valores de ΔE_{O_2} se computan para la estructura optimizada, es decir, la mejor conformación posible a -273°C, mientras que los valores cinéticos se calculan a temperatura ambiente. Por último, debido a errores intrínsecos en los métodos de QM/MM basados en DFT, las energía calculadas dependen fuertemente del funcional de intercambio y correlación y de los parámetros de inicio. Todas estas cuestiones pueden ser parcialmente corregidas normalizando la energía de unión de oxígeno. Para lograrlo, se ha definido el $\Delta\Delta E_{O_2}$, el cual corresponde al ΔE_{O_2} (energía de unión de oxígeno previamente descripta por la ecuación II.34) y la diferencia entre ΔE_{hemo}, energía de unión de oxígeno a un hemo con un grupo imidazol aislado en vacío (la cual da un valor de $22\ Kcal\ mol^{-1}$) y el valor de $k_{off}{}_{hemo_libre}$ ($10^4 s^{-1}$)[22,31].

II.4.5 Muestreo de migración de ligandos explícitos versus implícitos

En algunos casos, los procesos que se desean estudiar por dinámica clásica requieren de tiempos de simulación muy elevados (migración de una hormona a través de una membrana lipídica, migración del oxígeno dentro de la hemoglobina humana, etc.). Inclusive si uno realizase simulaciones suficientemente largas como para lograr que el proceso ocurra en el tiempo de simulación, no sería posible para obtener parámetros termodinámicos y cinéticos del proceso. Es necesario, en estos casos, utilizar técnicas computacionales que aceleran el muestreo. Algunas de ellas utilizan diferente temperatura para la proteína y el ligando, otras utilizan optimizaciones durante la dinámica molecular[32], y hay otras que no contemplan el ligando explícitamente[33]. Sin embargo, en estas técnicas de ligando explícito solo es posible estudiar un posible canal de migración por vez.

En la presente tesis doctoral, los canales de migración, así como también la energía libre asociada a los perfiles correspondientes, fueron computados con el método de muestreo del ligando implícito (ILS, por sus siglas en inglés, *Implicit Ligand Sampling*)[33,34], el cual permite estudiar todos los posibles canales de migración de manera simultánea. Este método procesa una simulación de DM en ausencia del ligando, utilizando una molécula de éste como sonda de manera implícita[34]. Utilizar ILS implica asumir que la presencia del ligando no afecta a la estructura global de la proteína, ni la topología del canal a medida que éste lo atraviese, lo cual ya ha sido previamente testeado para hemoproteínas[35].

Este método computa el potencial de fuerza media (PMF, por sus siglas en inglés, *potential of mean force*) correspondiente a ubicar una pequeña molécula como O_2, CO, entre otras, en todas las posibles ubicaciones dentro de una proteína. El PMF describe el costo en energía libre de tener una partícula localizada en una posición espacial dada, integrada sobre todos los grados de libertad del sistema, excepto los de la posición del ligando, y corresponde a una cantidad que indica cuáles áreas de la proteína son accesibles por el ligando y a qué valor de energía libre asociado. El método de ligando implícito para computar ligandos monoatómicos, diatómicos (y también más complejos) dentro de una proteína asume que estas partículas son muy pequeñas, y que por lo tanto interactúan débilmente con la matriz proteica, por lo que no afectan a la estructura global de la proteína, ni la topología del canal a medida que éste lo atraviese. Debido a esto, es posible analizar la dinámica de la proteína en ausencia del ligando y tratar la presencia de éste como una perturbación débil, y así, incluso, obtener resultados con una precisión aceptable. El resultado de utilizar este procedimiento es un mapa tridimensional completo de las regiones favorables y canales de migración asociados para ligandos pequeños dentro de una proteína.

A continuación, se derivará una expresión para obtener el PMF de un ligando implícito. El PMF $\mathrm{W}(r)$, el cual en este caso representa al valor de la energía libre de Gibbs de ubicar el ligando en una posición espacial específica r, está directamente relacionado con la probabilidad $\rho(r)$ de encontrar el ligando en esa posición, y está definido como:

$$\mathrm{W}(r) = -k_B T \ln\left[\frac{\rho(r)}{\rho_0}\right] \tag{II.35}$$

donde ρ_0 es un factor de normalización arbitrario.

A temperatura (T) y presión (P) constantes, la distribución de densidad de probabilidad del ligando $\rho(r)$ se puede expresar como:

$$\rho(r) = \frac{\int dV \int d^{3N}p \int d^{3N}q \int d^3p' \int d^3r'\, e^{-\beta[H(p,q,p',r')+PV]}\delta^3(r'-r)}{\int dV \int d^{3N}p \int d^{3N}q \int d^3p' \int d^3r' e^{-\beta[H(p,q,p',r')+PV]}} \tag{II.36}$$

donde $\int d^{3N}p \int d^{3N}q$ se refiere a la integración sobre todos los grados de libertad de la proteína del sistema de referencia (el cual incluye el solvente circundante), y $\int d^3p' \int d^3r'$ es la integración sobre los grados de libertad del ligando; $H(p,q,p',r')$ es el hamiltoniano para el sistema proteína-ligando, V es el volumen que ocupa el sistema, y se define $\beta = (k_B T)^{-1}$.

A continuación se procederá a derivar el PMF correspondiente a la migración del ligando, tratando el efecto de éste como una perturbación sobre el ensamble de referencia del estado de la proteína sin ligando. Considerando ahora los grados de libertad del ligando, de manera separada de aquellos correspondientes al resto del sistema proteína-ligando, la densidad de la probabilidad de encontrar al ligando adopta la siguiente forma:

$$\rho(r) = \frac{\int dV \int d^{3N}p \int d^{3N}q \int dp' \int d\Omega \int d^3r' e^{-\beta[H(p,q,p',r',\Omega)+PV]} \delta^3(r'-r)}{\int dV \int d^{3N}p \int d^{3N}q \int dp' \int d\Omega \int d^3r' e^{-\beta[H(p,q,p',r',\Omega)+PV]}} \quad \text{(II.37)}$$

donde r es el centro de masa del ligando, p' es el grado de libertad del momento del ligando y Ω corresponde a todos los grados de libertad de las coordenadas restantes del ligando.

Frente a la presencia de un ligando con grados de libertad interna, el hamiltoniano para el sistema de la proteína de referencia (H_0) será modificado por una cantidad igual a la energía de interacción proteína-ligando $\Delta E(q,r,\Omega)$, la energía cinética $K(p')$ y también a la energía potencial interna del ligando $U(\Omega)$:

$$\mathrm{H}(p,q,p',r',\Omega) = \mathrm{H}_0(p,q) + \Delta E(q,r,\Omega) + U(\Omega) + K(p') \quad \text{(II.38)}$$

Insertando el hamiltoniano perturbado (ecuación II.38) en la expresión de la densidad de probabilidad del ligando (ecuación II.37), se obtiene

$$\rho(r) = \frac{\int dV \int d^{3N}p \int d^{3N}q \int d\Omega\, e^{-\beta[\mathrm{H}_0(p,q)+PV])]} e^{-\beta[\Delta E(q,r,\Omega)+U(\Omega)]} \int dp\prime e^{-\beta K(p\prime)}}{\int dV \int d^{3N}p \int d^{3N}q \int d\Omega \int d^3r\prime e^{-\beta[\mathrm{H}_0(p,q)+PV])]} e^{-\beta[\Delta E(q,r,\Omega)+U(\Omega)]} \int dp\prime e^{-\beta K(p\prime)}} \quad \text{(II.39)}$$

Buscando expresar estas notaciones en términos de un ensamble isotérmico-isobárico (NTP) promedio sobre todos los estados de la proteína de referencia, dicho promedio de cualquier observable general $A(r)$ puede definirse como:

$$< A(r) >_{NTP} = \frac{\int dV \int d^{3N}p \int d^{3N}q\, e^{-\beta[H(p,q)+PV]} A(p,q,r)}{\int dV \int d^{3N}p \int d^{3N}q\, e^{-\beta[H(p,q)+PV]}} \quad \text{(II.40)}$$

Luego, usando esta definición del ensamble NTP promedio, la distribución de probabilidad del ligando queda de la siguiente manera:

$$\rho(r) = \frac{<\int d\Omega\, e^{-\beta[\Delta E(r,\Omega)+U(\Omega)]}>_{NTP}}{<\int d^3r\prime \int d\Omega\, e^{-\beta[\Delta E(r\prime,\Omega)+U(\Omega)]}>_{NTP}} \quad \text{(II.41)}$$

Ahora, insertando dicha expresión en la definición de PMF, ecuación (II.35) e imponiendo que el PMF debe ser 0 cuando el ligando está en vacío (definido por $\Delta E(q,r,\Omega) = 0$), se obtiene:

$$\mathrm{W}(r) = -k_B T \ln\left[\frac{<\int d\Omega\, e^{-\beta[\Delta E(r,\Omega)+U(\Omega)]}>_{NTP}}{<\int d\Omega\, e^{-\beta U(\Omega)}>_{NTP}}\right] \quad \text{(II.42)}$$

Para el caso de ligando diatómicos, como los tratados en la presente tesis, los enlaces de unión se mantienen fijos, por lo que los únicos grados de libertad Ω que restan son aquellos específicos de la orientación del ligando. En este caso, la energía interna del ligando $U(\Omega)$ es constante, por lo que el término que lo contiene en la ecuación II.42 se cancela. Por lo tanto, la expresión para el PMF de la ecuación II.42, deriva en la siguiente expresión simplificada, utilizada por el método de ILS:

$$\mathrm{W}(r) = -k_B T \ln\left[\frac{<\int d\Omega\, e^{-\beta[\Delta E(r,\Omega)]}>_{NTP}}{<\int d\Omega>}\right] \quad \text{(II.43)}$$

II.4.6 Detalles técnicos. Topología y perfiles de energía libre de migración de ligandos

Los cálculos de ILS se han realizado considerando una grilla rectangular, con una resolución de 0.5Å, la cual incluye toda la caja de la simulación (proteína y solvente). La sonda usada fue, en algunos casos, una molécula de CO, y en otros, una de O_2. Se utilizaron 5000 eventos (*frames*, del inglés) tomadas de simulaciones de producción de DM. La duración de las simulaciones varía entre 50 y 200ns, según el estudio realizado (ver en cada capítulo de resultados). Los valores para el tamaño de la grilla, la resolución y la cantidad de fotos han sido ampliamente muestreados en un trabajo previo[35]. Los análisis de los cálculos de ILS se llevaron a cabo con un programa desarrollado en el lenguaje de programación TCL, en conjunto con la Lic. Victoria Gisel Dumas y el Dr. Marcelo Martí. Dicho programa permite determinar, para cada caso, la magnitud / profundidad de los pozos (mínimos energéticos) y las barreras energéticas escaladas, por lo que la energía libre del ligando en el solvente ha sido configurada en cero.

II.6 Referencias

(1) Szabo, A.; Ostlund, N. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; In New York: Mc GrawHill Publishing Company; 1989.

(2) Levine, I. *Quantum Chemistry*, 5th edition; Hall, P., Ed.; New York - USA, 2001.

(3) Becke, A. *J. Chem. Phys.* **1993**, *98* (7), 5648.

(4) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77* (18), 3865–3868.

(5) Leach, A. *Molecular Modelling: Principles and Applications*; EMA, P. E., Ed.; 2001.

(6) Ponder, J.; Case, D.; Valerie, D. In *Advances in Protein Chemistry*; Press, A., Ed.; 2003; pp 27–85.

(7) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham III, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91* (1-3), 1–41.

(8) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79* (2), 926–935.

(9) Ewald, P. *Ann. Phys.* **1921**, *369*, 253–287.

(10) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(11) Essmann, U.; Perera, L.; Berkowitz, M.; Darden, T.; Lee, H.; Pedersen, L. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(12) Darden, T.; Perera, L.; Li, L.; Pedersen, L. *Structure* **1999**, *7*, R55–R60.

(13) Crespo, A.; Scherlis, D. A.; Martí, M. A.; Ordejón, P.; Roitberg, A. E.; Estrin, D. A. *J. Phys. Chem. B* **2003**, *107* (49), 13728–13736.

(14) Verlet, L. *Phys. Rev.* **1967**, *159* (98).

(15) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23* (3), 327–341.

(16) Berendsen, H.; Postma, J.; Van Gunsteren, W.; DiNola, A.; Haak, J. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(17) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14* (1), 33–38.

(18) Giordano, D.; Pesce, A.; Boechi, L.; Bustamante, J. P.; Caldelli, E.; Howes, B. D.; Riccio, A.; di Prisco, G.; Nardini, M.; Estrin, D.; Smulevich, G.; Bolognesi, M.; Verde, C. *Fed. Eur. Biochem. Soc. J.* **2015**.

(19) Pesce, A.; Nardini, M.; LaBarre, M.; Richard, C.; Wittenberg, J. B.; Wittenberg, B. A.; Guertin, M.; Bolognesi, M. *Biochim. Biophys. Acta - Proteins Proteomics* **2011**, *1814* (6), 810–816.

(20) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21* (12), 1049–1074.

(21) Marti, M. A.; Capece, L.; Bidon-Chanal, A.; Crespo, A.; Guallar, V.; Luque, F. J.; Estrin, D. A. *Methods Enzymol.* **2008**, *437*, 477–498.

(22) Marti, M. A.; Crespo, A.; Capece, L.; Boechi, L.; Bikiel, D. E.; Scherlis, D. A.; Estrin, D. A. *J. Inorg. Biochem.* **2006**, *100* (4), 761–770.

(23) 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. *Phys. Chem. Chem. Phys.* **2006**, *8* (48), 5611–5628.

(24) Forti, F.; Boechi, L.; Bikiel, D.; Martí, M. A.; Nardini, M.; Bolognesi, M.; Viappiani, C.; Estrin, D. A.; Luque, F. J. *J. Phys. Chem. B* **2011**, *115* (46), 13771–13780.

(25) Capece, L.; Lewis-ballester, A.; Marti, M. A.; Estrin, D. A.; Yeh, S. *Biochemistry* **2011**, *50*, 10910–10918.

(26) 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. *Biochemistry* **2011**, *50* (19), 3946–3956.

(27) Giordano, D.; Boechi, L.; Samuni, U.; Vergara, A.; Martı, M. A.; Estrin, A.; Friedman, J. M.; Mazzarella, L.; Prisco, G.; Grassi, L. *FEBS J.* **2009**, *276*, 2266–2277.

(28) Perissinotti, L. L.; Marti, M. A.; Doctorovich, F.; Luque, F. J.; Estrin, D. A. *Biochemistry* **2008**, *47* (37), 9793–9802.

(29) Capece, L.; Boechi, L.; Perissinotti, L. L.; Arroyo-Mañez, P.; Bikiel, D. E.; Smulevich, G.; Marti, M. A.; Estrin, D. A. *Biochimica et Biophysica Acta - Proteins and Proteomics*. **2013**, *1834* (9), 1722-1738.

(30) Scherlis, D.; Cococcioni, M.; Sit, P.; Marzari, N. *J. Phys. Chem. B* **2007**, *111*, 7384–7391.

(31) Scott, E. E.; Gibson, Q. H.; Olson, J. S. *J. Biol. Chem.* **2001**, *276*, 5177–5188.

(32) Guallar, V.; Lu, C.; Borrelli, K.; Egawa, T.; Yeh, S. *J. Biol. Chem.* **2009**, *284* (5), 3106–3116.

(33) Cohen, J.; Olsen, K. W.; Schulten, K. *Methods Enzymol.* **2008**, *437* (07), 439–457.

(34) Cohen, J.; Arkhipov, A.; Braun, R.; Schulten, K. *Biophys. J.* **2006**, *91* (5), 1844–1857.

(35) Forti, F.; Boechi, L.; Estrin, D. A.; Marti, M. A. *J. Comput. Chem.* **2011**, *32* (10), 2219–2231.

Resultados

y

discusión

III. Cavidades internas y moléculas de agua: factores determinantes en la captación de ligandos

III.1 Introducción

En los últimos años muchos grupos de investigación han estudiado en profundidad la importancia de los canales de migración para el ingreso de ligandos pequeños, así como la relevancia de aminoácidos del sitio activo que los estabilicen. En este contexto, dicho capítulo profundizará en el proceso de captación de ligandos (medido experimentalmente como la constante k_{on}), estudiando otros procesos involucrados, tales como la migración de éstos a través de los diferentes sistemas de cavidades y túneles que presentan las trHbs, y las interacciones con las moléculas de aguas que intervienen en el sitio activo. Para esto, se han elegido dos trHbs pertenecientes a los grupos O y N como modelos para estudiar dichos factores, la Tf-trHbO y la Mt-trHbN.

Las cavidades internas han sido descriptas hace ya más de tres décadas por Tilton y Kuntz en las dos globinas más conocidas y estudiadas, la hemoglobina y la mioglobina[1]. Mediante experimentos realizados bajo atmósfera de Xe, identificaron sitios de las proteínas en los que cuales los átomos de Xe quedaban retenidos, sitios llamados cavidades internas. En el caso de la mioglobina, los extensos estudios que siguieron dicho descubrimiento han revelado su importancia, siendo capaces no solo de albergar ligandos de manera transitoria, sino de ser parte de un sistema de túneles y cavidades de migración[2–4]. Estudios cristalográficos sobre algunas trHbs (pertenecientes al grupo N, *Mycobacterium tuberculosis*, *Chlamydomonas eugametos*, y *Paramecium caudatum*) sugirieron también la importancia de estas cavidades internas en el proceso de captación de ligandos[5].

Por otro lado, actualmente se conoce que las moléculas de agua juegan un rol activo en los procesos de reconocimiento biomolecular. Muchas veces, esto es debido a la desolvatación parcial del ligando, y otras, a la región del receptor involucrada en el proceso de reconocimiento al ligando, un hecho que a menudo involucra el reordenamiento de moléculas de aguas cercanas o dentro del sitio activo[6–10]. Dicha reorganización conduce a cambios sutiles, delicados, en la red

de puentes de hidrógeno con el agua[11], que tendrán consecuencias directas sobre el proceso global de unión, especialmente cuando las moléculas de agua están embebidas en la cavidad de reconocimiento del ligando[12,13]. En este último caso, el desplazamiento de moléculas de agua retenidas puede promover efectos favorables o desfavorables sobre el proceso global de unión, dependiendo de las propiedades fisicoquímicas de las superficies de las biomoléculas y del sistema de cavidades que posean. En investigaciones teóricas y experimentales previas, ya se ha observado la presencia de moléculas de aguas en el sitio distal de la Tf-trHbO, en cercano contacto con la YB10[14]. La presencia de solvente en el sitio activo, aunque no coordinado al grupo hemo, puede imponer barreras cinéticas a la unión del ligando debido a la necesidad de intercambiar moléculas de solvente con el ligando, como fue previamente reportado por el grupo liderado por Olson para la mioglobina [15] y por Ouellet y otros investigadores para la Mt-trHbN [16].

Experimentalmente, hay dos técnicas ampliamente utilizadas para medir la k_{on} y k_{off} de ligandos en proteínas: flujo detenido (*stopped flow*, en inglés) y flash fotólisis. La primera consiste en realizar mezclados muy veloces de los reactivos, los cuales serán detectados y medidos durante la reacción de unión. En la segunda, se hace incidir un pulso de luz sobre la proteína que posee el ligando coordinado al Fe del grupo hemo, logrando que el mismo se libere, y realizando posteriores mediciones sobre la re-unión de este ligando al Fe. El mismo principio de flash fotólisis es el que usa la técnica de absorción transitoria ultra rápida, utilizada para cubrir rangos temporales del orden de los femtosegundos a picosegundos (10^{-15}s a 10^{-12}s).

En la primera sección de resultados del presente capítulo (sección III.3.1) se mostrarán estudios computacionales con detalle molecular donde un sistema de cavidades permanentes y transitorias conformando el portal E7 ha sido identificado como el canal responsable del acceso de ligandos en la proteína modelo considerada, la Tf-trHbO[17]. En base a los resultados allí presentes, ha sido posible proponer un esquema cinético asociado al proceso de captación. Dichos estudios teóricos se han complementados con distintos experimentos cinéticos y espectroscópicos, de flash fotólisis con un láser pulsado y absorción transitoria ultra rápida llevados a cabo por los grupos de los Dres. Cristiano Viappiani (Universitá degli studi di Parma, Italia) y Paolo Foggi (Universitá di Firenze, Italia) a fin de medir la cinética de re-unión del ligando en el rango de 10^{-12}s a 10^{-1}s. Sin embargo, debido a las limitaciones presentes en los rangos de tiempos medibles por las técnicas experimentales mencionadas, si bien la combinación de ellas fue primordial para estudiar la cinética de re-unión prácticamente completa, hay un pequeño rango temporal que ha quedado inaccesible por ambas técnicas espectroscópicas entre 10^{-9} y 10^{-8}s. Independientemente de esta cuestión técnica, se mostrará cómo la complementariedad del estudio del proceso de captación desde la teoría y los experimentos ha

permitido validar satisfactoriamente el esquema cinético basado en el sistema de cavidades permanentes y transitorias propuesto.

Con el objetivo de estudiar de qué manera las cavidades internas y las moléculas de agua contribuyen al proceso de captación, definiendo así las constantes de velocidad asociadas, se realizaron mutaciones puntuales en el sitio distal de la Tf-trHbO. El sitio activo de esta trHb está caracterizado por el invariante enlace de pentacoordinación entre el grupo hemo y la histidina proximal, y por una cavidad distal altamente polar, en la cual la YB10, YCD1 y WG8 proveen tres potenciales donantes de puentes de hidrógeno para estabilizar ligandos exógenos[17]. El WG8 ha sido reportado como el aminoácido más involucrado en una fuerte estabilización de ligandos exógenos como sulfuro (S^-) y fluoruro (F^-) en el estado férrico del hierro hémico, y monóxido de carbono (CO) y oxígeno molecular (O_2) en el estado ferroso[14,18,19], junto con la YCD1 que también ha presentado un rol importante, pero no tan significativo. Por consiguiente, en la sección "III.3.2 Modulación del proceso de captación de ligandos" se presentarán resultados correspondientes a dos mutaciones, una mutación única, WG8F, y una triple, WG8F-YB10F-YCD1, donde todas las interacciones polares del sitio activo han sido removidas. Como en la sección anterior, se mostrarán resultados de experimentos de cinéticas de re-unión para cada caso, realizados por los mismos grupos colaboradores. Para comprender las bases estructurales y dinámicas de los cambios experimentales observados en las mediciones de cinéticas de re-unión tanto de la proteína wt como de las dos mutantes realizadas, se han llevado a cabo dos enfoques computacionales basados en simulaciones de DM sin ligandos explícitos: i) análisis de las barreras de energía libre asociadas a la migración del ligando desde el solvente al sitio activo atravesando un canal, y ii) caracterización probabilística de regiones del espacio específicas, localizadas dentro del sitio activo, conteniendo moléculas de agua que posean un potencial efecto sobre el proceso global de captación.

Habiendo profundizado en el estudio de la importancia de las cavidades internas y las moléculas de agua retenidas en el proceso de captación de ligandos, en la sección III.3.3 se hará especial hincapié en una posición estructural que es capaz de modular ambos factores. Respecto a la Tf-trHbO, la Mt-trHbN presenta características del sitio activo y de los canales de migración marcadamente diferentes. Posee dos canales, el largo y el corto G8, que son los que permiten el acceso a ligandos desde el solvente al sitio activo. Con los dos aminoácidos polares YB10 y EE11 y la presencia de una valina en G8, la estabilización de ligandos es muy distinta respecto a la correspondiente en la Tf-trHbO. La posición estructural G8 se encuentra espacialmente en un lugar estratégico, puesto que está en la zona de encuentro de ambos canales, largo y corto G8, y también posee un rol fundamental en estabilización de ligandos exógenos. Se mostrará cómo este aminoácido es capaz de alterar significativamente tanto la dinámica de las cavidades

internas como la estabilización de moléculas de aguas presentes en el sitio activo, generando cambios apreciables en las constantes k_{on}. Para esto, se realizarán simulaciones de DM y un análisis equivalente al de la sección anterior para la Tf-trHbO, siempre acompañado de mediciones experimentales realizadas por colaboradores (en este caso, de cinéticas obtenidas por la técnica de flujo detenido), sobre la proteína wt y dos mutantes únicas de la Mt-trHbN, yendo desde la valina (un aminoácido pequeño), a uno muy voluminoso y rígido, como el triptófano, pasando por la fenilalanina.

Los estudios aquí presentados revelarán cómo el proceso de captación de ligandos pequeños en trHbs está regulado por un delicado balance a nivel molecular entre aminoácidos claves en la arquitectura de cavidades internas y canales y aquellos involucrados en la retención de moléculas de agua en el sitio activo.

III.2 Métodos computacionales

III.2.1 Configuración de los sistemas y parámetros de simulación

Las estructuras de inicio corresponden a las estructuras obtenidas por cristalografía de rayos X de la Tf-trHbO (código de la base de datos *Protein Data Bank* 2BMM[17]) y de la Mt-trHbN (código de PDB 1IDR[20]). Tanto la protonación de los aminoácidos, como la construcción de la caja de aguas y el resto de las condiciones de los sistemas se establecieron según el protocolo presentado en la sección II.4.3 del capítulo Métodos. El tiempo de simulación de las DM de producción fue de 100ns para las proteínas wt y mutantes WG8F y WG8F-YB10F-YCD1F de la Tf-trHbO y de 200ns para las proteínas wt y mutantes únicas VG8F y VG8W de la Mt-trHbN.

III.2.2 Análisis de los perfiles de energía libre de migración de ligandos en las proteínas Tf-trHbO y Mt-trHbN

Los cálculos de ILS sobre estos sistemas proteicos se realizaron siguiendo el protocolo presentado en la sección II.4.3 del capítulo Métodos. La sonda utilizada, en todos los casos, fue una molécula de CO, puesto que en estudios experimentales de reactividad de hemoproteínas es el ligando comúnmente más utilizado, debido a su fácil manejo, buen corrimiento del espectro UV-visible al unirse al hemo y estabilidad de la forma Fe-CO dentro de la proteína.

La energía libre para los perfiles de energía fue establecida en $0\ kcal\ mol^{-1}$ para el CO en el seno del solvente, a 9 Å del Fe hémico para el portal E7, en el caso de la Tf-trHbO, y a 13 y 24 Å del Fe hémico, para el canal corto G8 y para el canal largo, respectivamente, de la Mt-trHbN.

III.2.3 Definición, identificación y caracterización de sitios de agua (SA)

Los sitios de agua (SA) corresponden a regiones específicas de la proteína de interés, albergando una molécula de agua con un valor de probabilidad mayor al de una molécula rodeada por otras aguas en el seno del solvente. Como se ha comprobado previamente en otros trabajos[21–23], estas regiones pueden identificarse calculando la probabilidad de encontrar una molécula de agua dentro de una región especial definida durante simulaciones de DM. El volumen de la región establecida ha sido definida de manera arbitraria en 1 $Å^3$, y las coordenadas del centro del SA corresponden a la posición promedio de todos los átomos de O de las moléculas de agua que visitaron el SA durante la simulación. Se ha considerado que una molécula de agua está ocupando un SA cuando la distancia entre la posición de su átomo de O y el centro del SA es menor a 0,6 Å. Una vez identificado, para todos los SAs, se calcularon los siguientes parámetros:

a) *Probabilidad de encontrar una molécula de agua (PEA).* Corresponde a la probabilidad de encontrar una molécula de agua en la región definida por el SA y normalizada respecto a la probabilidad de estar en el seno del solvente albergando una molécula en una esfera del mismo volumen definido por un SA a una temperatura y presión correspondientes. Así, la PEA se usa como un radio de corte para decidir cuáles SAs serán considerados para una caracterización. Serán considerados solo los SA con valores de PEA mayores a 2.

b) *Energía potencial asociada con la interacción de moléculas de agua dentro de un SA con la proteína y el resto del solvente.* Las contribuciones calculadas entre una molécula de agua localizada dentro de un SA y el resto de la proteína, y con las otras moléculas de solvente, se calcularon teniendo en cuenta un radio de corte para las interacciones de 8 Å, el cual ya ha sido validado en otro trabajo[21]. Para cada SA, se calculó la energía de interacción promedio sobre los últimos 90ns de cada simulación. Los errores estándares para todos los promedios calculados presentan una variabilidad menor a 0,01 $kcal\ mol^{-1}$.

III.3 Resultados y discusión

III.3.1 Sistema de cavidades permanentes y transitorias

Se llevaron a cabo simulaciones de DM seguidas de cálculos de ILS para estudiar la migración de ligandos pequeños en la Tf-trHbO, en su forma *wild type* (wt, del inglés, que significa tipo natural o silvestre).

Los análisis del estudio teórico mostraron que hay solo un sistema de cavidades definiendo un túnel a través del cual pequeños ligandos como el CO o el O_2 pueden ingresar y unirse al grupo hemo (figura III.1). Este túnel resulta ser el denominado portal E7, previamente encontrado en otras trHbs[24–26]. En este caso, dicho túnel presenta tres cavidades asociadas que fueron encontradas analizando extensivamente la simulación de DM y los resultados arrojados por el análisis de ILS.

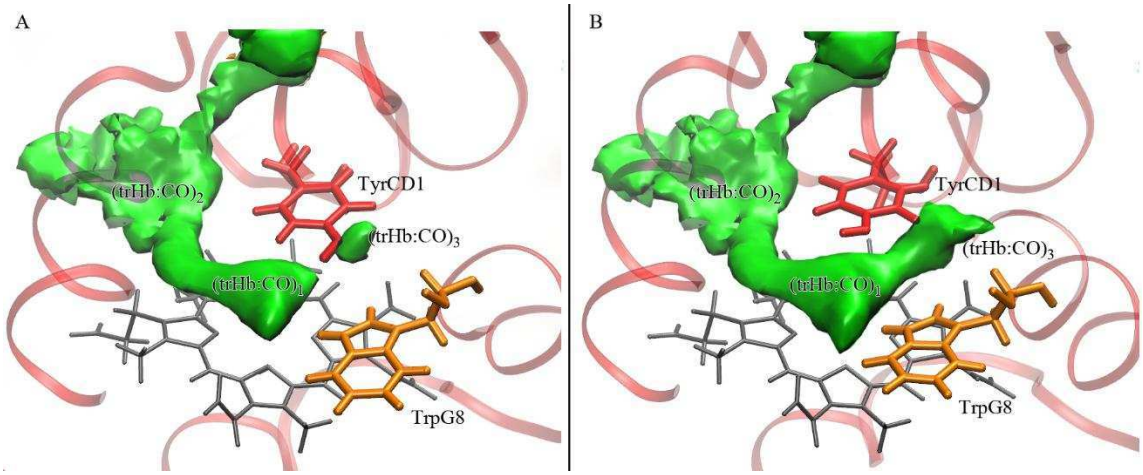
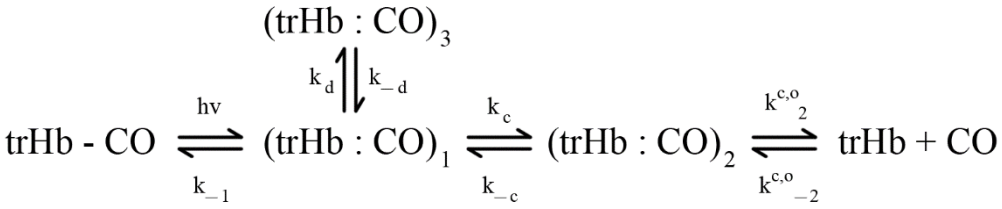


Figura III.1. Representaciones del sitio distal y del canal de ingreso / egreso de ligandos pequeños conectado por tres cavidades en la Tf-trHbO. Se muestran dos conformaciones espaciales (en A y B) de la YCD1 (en rojo). Nótese que la segunda conformación de la YCD1 (B) permite el acceso a la cavidad $(trHb:CO)_3$.

Se identificaron dos sitios de unión permanentes sobre el canal de ingreso / egreso, ambos localizados cerca del grupo hemo, $(trHb : CO)_1$ y $(trHb : CO)_2$, quedando el primero sobre el átomo de Fe. También se observó una cavidad adicional transitoria, $(trHb : CO)_3$, la que, sin embargo, mostró estar desconectada del sitio distal, debido a que la YCD1 bloquea el pasaje entre $(trHb : CO)_1$ y $(trHb : CO)_3$ (Figura III.1A). Monitoreando la evolución temporal de la torsión de la cadena lateral de la YCD1, se observó una conformación diferente en la cual se abre el posible túnel que conecta $(trHb : CO)_1$ y la cavidad pequeña $(trHb : CO)_3$ (Figura III.1B). Esto ha sido atribuido al desplazamiento de la YCD1 en dirección opuesta al WG8. Ambas conformaciones espaciales de la YCD1 han sido muestreadas de manera significativa durante la simulación (ver Anexo, figura A.III.1), indicando que la tercera cavidad puede ser accesible por ligandos pequeños desde el sitio distal.

Partiendo de este estudio se propuso el esquema cinético presente en el esquema III.1. Usando dicho esquema, los investigadores de los grupos experimentales colaboradores armaron y resolvieron numéricamente un sistema de ecuaciones diferenciales para evaluar el ajuste sobre los datos experimentales y calcular las constantes microscópicas que definen la velocidad con la que, en este caso el CO, migra desde una cavidad a otra, incluso hasta y desde el solvente.

La cinética observada al realizar el experimento de flash fotólisis puede observarse en la figura III.2, junto a la superposición del ajuste logrado partiendo del esquema cinético propuesto.



Esquema III.1. Esquema cinético microscópico propuesto para la cinética de re-unión del CO en la Tf-trHbO. (trHb : $CO)_1$ y (trHb : $CO)_2$ indican, respectivamente, los sitios primario y secundario de unión, luego del CO fotodisociado dentro de la cavidad distal. Mientras que (trHb : $CO)_3$ representa la cavidad temporal accesible desde el sitio activo.

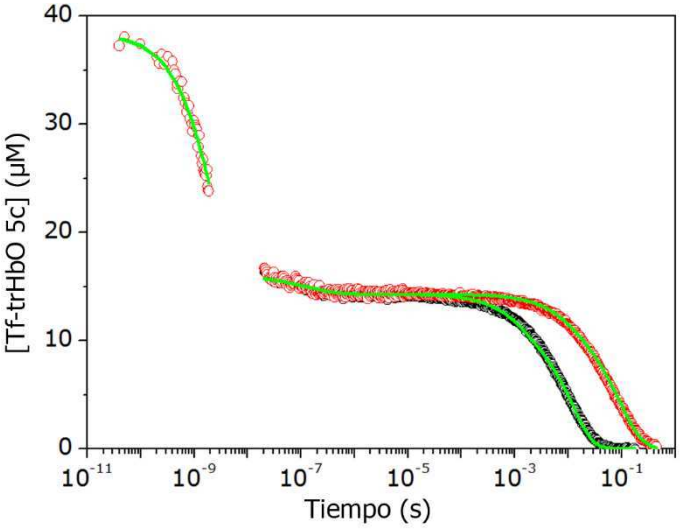


Figura III.2. Resultados del análisis global del curso completo de la cinética de re-unión de CO en la Tf-trHbO combinando espectroscopia de absorción transitoria en escala de femtosegundos y flash fotólisis en escala de nanosegundos. A medida que pasa el tiempo se observa cómo disminuye la concentración de Tf-trHbO pentacoordinada [Tf-trHbO 5c], puesto que se vuelve a unir el CO liberado con la fotólisis. *T* = 20 °C. Las curvas negra y roja corresponden a presiones de 1 y 0.1 atm de CO, respectivamente. Los ajustes (líneas verdes) se observan superpuestos a los datos experimentales (círculos).

Como puede observarse en la figura III.2, luego de que la dinámica del estado excitado se complete, se observa la aparición de la recombinación geminada desde la cavidad distal (del inglés, *geminate recombination*). Ésta sucede dentro de la escala temporal de cientos de picosegundos, correspondiente a la región del primer decaimiento de las curvas observadas. La constante temporal aproximada de 2ns y su gran amplitud de esta fase de la cinética (57% del cambio en absorbancia total), sugiere que la barrera para recombinación es similar a las reportadas para la trHbO de *Bacillus subtilis* (Bs-trHbO)[24] y la trHbO de *Mycobacterium tubercolosis* (Mt-trHbO)[27,28]. Las tres proteínas, Tf-trHbO, Bs-trHbO y Mt-trHbO exhiben una arquitectura del sitio distal con aminoácidos polares conservados, los cuales definen el patrón de la red de puentes de hidrógeno que estabilizan al ligando unido al Fe.

La constante de asociación (k_{on}) para la Tf-trHbO fue estimada en $8{,}7 \cdot 10^4\ M^{-1}s^{-1}$ y $9{,}3 \cdot 10^5\ M^{-1}s^{-1}$ para la conformación lenta (~80%) y rápida (~20%), respectivamente, a 20°C. Estos valores aumentan a $1{,}63 \cdot 10^5\ M^{-1}s^{-1}$ y $2{,}75 \cdot 10^6\ M^{-1}s^{-1}$ a 25°C, los cuales pueden compararse con la determinación previa de k_{on} de O_2, estimada en $9 \cdot 10^5\ M^{-1}s^{-1}$[17]. Este valor

corresponde a la conformación más poblada (conformación lenta), e indica que la $k_{on_{O_2}}$ es un orden de magnitud más grande que la $k_{on_{CO}}$, según estas mediciones y estimaciones realizadas.

Hasta aquí, se ha descripto con un detalle molecular el túnel con sus cavidades permanentes y transitorias responsables del acceso de ligandos en la Tf-trHbO, el portal E7. Se propuso un esquema cinético asociado al proceso de captación sobre el cual el grupo colaborador armó y resolvió numéricamente un sistema de ecuaciones diferenciales. Evaluaron el ajuste sobre los datos experimentales y calcularon las constantes microscópicas que definen la velocidad con la que el ligando migra desde una cavidad a otra, incluso hasta y desde el solvente. Habiéndose alcanzado un ajuste realmente satisfactorio sobre la curva experimental, los resultados aquí presentados sugieren fuertemente que el esquema cinético propuesto es consistente con los datos experimentales cinéticos.

III.3.2 Modulación del proceso de captación de ligandos

Considerando las conclusiones de la sección anterior, el objetivo de esta sección es profundizar sobre cómo las cavidades internas y las moléculas de aguas retenidas en el sitio activo son capaces de modular el proceso de captación. Para esto, se estudiaron las consecuencias que pueden generar las siguientes mutaciones de aminoácidos distales sobre la cinética de re-unión: mutación única del WG8 por fenilalanina (WG8F) y mutación triple WG8F-YB10F-YCD1F (también denominada 3F), donde todas las interacciones polares del sitio activo han sido removidas, a fin de obtener una descripción más acabada de la influencia del agua en el proceso de unión del ligando.

Los experimentos de cinéticas de re-unión, realizados por los mismos grupos colaboradores que en la sección anterior, arrojan la curva presentada en la figura III.3, en el rango de algunos picosegundos a cientos de milisegundos. Los datos están mostrados como fracción de proteína sin ligando (N(t) = ΔA(t)/ΔA(t_0) donde t_0 es el tiempo al cual la recombinación geminada comienza luego de la relajación del estado excitado) como función del tiempo de decaimiento luego de la excitación. Solo para poder comparar, sobre la misma gráfica se reporta también la curva para la Tf-trHbO wt,

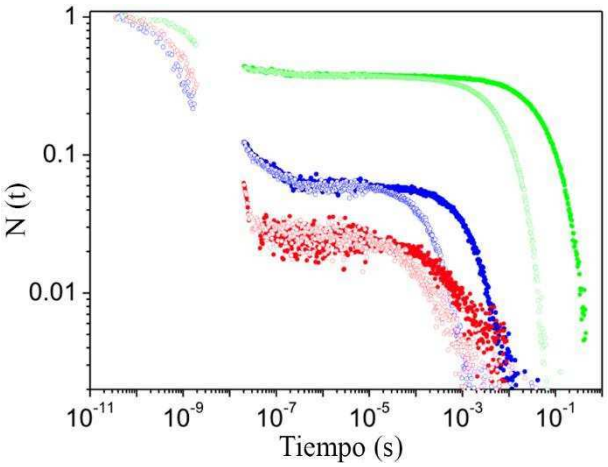


Figura III.3. Cinéticas de re-unión para las proteínas wt (verde), y mutantes WG8F (azul) y 3F (rojo) de la Tf-trHbO. Mediciones realizadas con 1 (círculos huecos) y 0.1 (círculos rellenos) atm de CO. T=20°C; λ=436 nm. N(t) = ΔA(t)/ΔA(t_0) donde t_0 es el tiempo al cual la recombinación geminada comienza luego de la relajación del estado excitado.

presentada ya en la sección previa. Para ambas mutantes, puede observarse que la recombinación geminada resulta ser más rápida que para la Tf-trHbO wt y además, posee una mayor amplitud. Las moléculas de CO disociadas se re-unen en esta fase de recombinación geminada en la mutante WG8F con la constante más rápida. La amplitud de la fase de recombinación geminada, estimada en ~1 µs, es de 97% en la mutante 3F y 94% en la WG8F, en contraste con el valor de 57% observado en la forma wt de la Tf-trHbO.

Al analizar la cinética completa de re-unión a partir del esquema III.1, los colaboradores modelaron las curvas de re-unión optimizando las constantes del esquema III.1, para cada caso. Como resultado, se extrajeron los valores ajustados, que pueden observarse en la tabla III.1.

Tabla III.1. Constantes microscópicas determinadas con el ajuste global en el curso total (a 1 y 0.1 atm de CO) de la cinética de re-unión del CO para las formas wt, WG8F y 3F de la Tf-trHbO a 20 °C.

	wt	WG8F	3F
k_{-1} ($10^8 s^{-1}$)	3.0±0.1	20.0±0.6	11.5±0.3
k_c ($10^8 s^{-1}$)	1.9±0.1	6.6±0.3	6.5±0.3
k_{-c} ($10^6 s^{-1}$)	2.0±0.2	84±8	200±20
k_d ($10^7 s^{-1}$)	3±1	8±2	8±2
k_{-d} ($10^6 s^{-1}$)	6±1	9±1	230±40
k^c_2 ($10^7 s^{-1}$)	3.9±0.2	2.0±0.1	0.9±0.1
k^c_{-2} ($10^6 M^{-1}s^{-1}$)	3.0±0.3	5.0±0.5	6.2±0.6
k^o_2 ($10^7 s^{-1}$)	9±4	-	-
k^o_{-2} ($10^7 M^{-1}s^{-1}$)	7±3	-	-

La presencia de moléculas de agua en el sitio activo de la Tf-trHbO ya fue previamente evidenciada en su estructura de rayos X[17]. Para que ingresen, es necesario contrarrestar y superar la gran barrera entrópica para el ingreso del solvente con interacciones electroestáticas internas favorables con aminoácidos polares. De acuerdo con las simulaciones de DM realizadas, las moléculas de agua ingresan al sitio activo a través del portal E7. Estas simulaciones sin ligando de la Tf-trHbO proveen una clara evidencia de que la cavidad distal alberga moléculas de agua que forman interacciones estables con las cadenas laterales del WG8 y la YCD1. Aunque ninguna de las moléculas de agua está coordinada al Fe hémico, una de ellas se encuentra prácticamente en la misma posición que si estuviese coordinada. La coordinación del CO al Fe requiere que al menos dicha molécula de agua sea desplazada, y es esperable que este proceso imponga una barrera de energía libre detectable. En este contexto, el análisis de la ocupación de moléculas del solvente en el sitio distal, principalmente determinadas por la dinámica de puentes de hidrógeno con aminoácidos distales, revela distintos valores de probabilidad de encontrar una molécula de agua en una región en particular dentro de la cavidad (figura III.4). En particular, se observan 3 moléculas de agua fuertemente estabilizadas por interacciones de puentes de hidrógeno en la proteína wt (figura III.4A), con una región bien definida (ver sitios de aguas SA1,

SA2 y SA3). El WG8 resulta ser el mayor responsable de la estabilización de estas moléculas en SA1, con la ayuda de la YCD1. Por otro lado, en ausencia de WG8F (proteínas WG8F y 3F), SA1 se desestabiliza, puesto que las moléculas del solvente pierden un donante muy importante de puentes de hidrógeno. Solo la YCD1 mantiene parcialmente un rol estabilizante para las moléculas de agua de SA1 en la mutante WG8F (figura III.4B). En el caso de la mutante 3F, todas

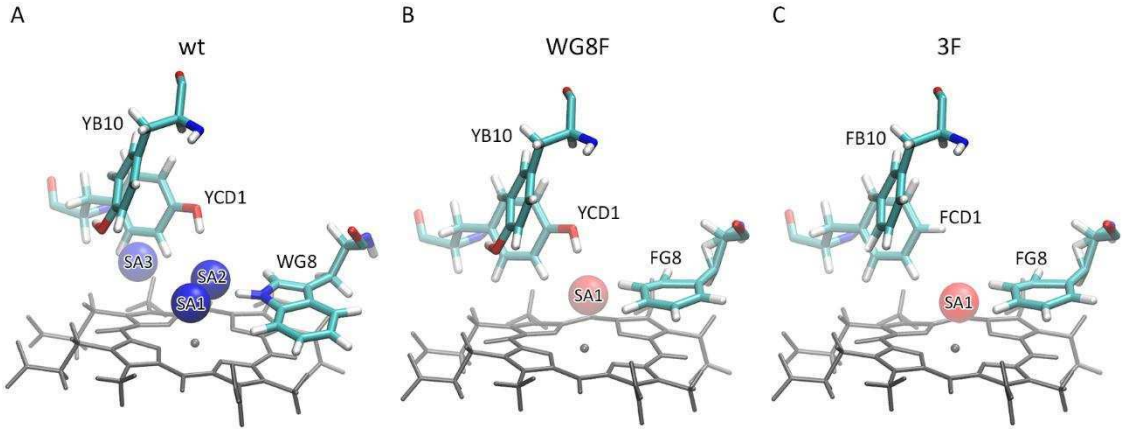


Figura III.4. Representación esquemática de los sitios de agua (SA) cercanos al grupo hemo. Los SAs se muestran como esferas de alta a baja probabilidad de contener una molécula de agua en una escala de azul a rojo. Los paneles A, B y C corresponden a las proteínas wt y a las mutantes WG8F y 3F de Tf-trHbO, respectivamente.

las interacciones polares en la cavidad distal se han perdido (figura III.4C), y cabe aclarar que la situación podría ser muy distinta dependiendo del estado de oxidación del Fe del grupo hemo. En un trabajo en colaboración con el grupo de la Dra. Giulietta Smulevich (Università degli Studi di Firenze – Italia), quienes realizaron experimentos de espectroscopía Raman resonante, se ha descripto una molécula de agua coordinada al grupo hemo en el estado férrico, y estabilizada por interacciones de moléculas de solvente adicionales[29]. Debido al hecho de que el Fe hémico se encuentra en su estado ferroso cuando está el CO unido, se ha asumido que cuando éste es fotodisociado, el Fe se mantiene en el mismo estado de oxidación. En ausencia de un ligando exógeno, la posición normalmente ocupada por un CO (u otro ligando pequeño) puede ser tomada por alguna molécula de agua cercana, dentro de la cavidad, como fue reportado en el trabajo en colaboración anteriormente citado[29]. Sin embargo, la estabilización de estas moléculas es marcadamente diferente, puesto que en el estado Fe(II) no hay moléculas de solvente coordinadas al metal. En este caso, dichas moléculas solo están estabilizadas débilmente por otras moléculas de agua, las cuales, de manera transitoria, están en contacto con ellas dentro de la cavidad distal.

Cabe notar que los aminoácidos del sitio activo atraviesan un pequeño reordenamiento estructural luego de la disociación de CO para alcanzar el equilibrio del estado hémico no unido. En este trabajo se ha asumido que la escala temporal de picosegundos a nanosegundos es

suficiente para alcanzar dicho reordenamiento estructural y la consecuente entrada de moléculas de agua a los SAs identificados basados en: i) solo se requieren pequeños reordenamientos de cadenas laterales de los aminoácidos y ii) es completamente factible en este corto tiempo que las moléculas de agua ocupen los SAs considerando que éstos están muy próximos a la cavidad distal en el estado Fe-CO unido, como ha sido previamente reportado por Droghetti y otros colaboradores[14].

Partiendo de las simulaciones de DM, también es posible obtener información termodinámica de los sistemas estudiados, como se detalla en la sección de métodos (III.2). Las moléculas de agua en SA1 muestran fuertes interacciones con la proteína wt, en comparación con las observadas para las mutantes estudiadas. La energía de interacción electroestática es cerca de $8.0\ kcal\ mol^{-1}$ más negativa para la wt que para la mutante WG8F y $16.0\ kcal\ mol^{-1}$ respecto a la mutante 3F. Por otro lado, la energía de interacción de Lennard-Jones es menos favorable en la forma wt que en ambas mutantes. Esto puede entenderse considerando que ambas mutantes poseen aminoácidos menos voluminosos y, por ende, ofrecen menos impedimento estérico en la cavidad. Sin embargo, esta contribución energética de $2{,}1\ kcal\ mol^{-1}$ en wt, $0{,}5\ kcal\ mol^{-1}$ en WG8F y $-1{,}6\ kcal\ mol^{-1}$ en 3F, no es suficiente para superar la estabilización electroestática favorable debido a la presencia de grupos polares capaces de interaccionar y establecer puentes de hidrógeno con las moléculas de agua. La energía de interacción global es más favorable para la proteína wt ($-14{,}8\ kcal\ mol^{-1}$) que para las mutantes WG8F ($-8{,}5\ kcal\ mol^{-1}$) o 3F ($-2{,}6\ kcal\ mol^{-1}$).

Se llevó a cabo un cálculo adicional, detallado en la sección de métodos (III.2), denominado probabilidad de encontrar agua (PEA), el cual refleja la probabilidad relativa de encontrar una molécula de agua dentro de un SA en comparación con el seno del solvente. En conformidad con el análisis estructural y energético, el PEA muestra que la probabilidad de encontrar una molécula de agua cerca del Fe hémico es mayor para la proteína wt que para las mutantes, siendo, los valores correspondientes, de 19,2, 11,6 y 2,5 para wt, WG8F y 3F, respectivamente. Esto significa que las moléculas de agua menos estabilizadas en el sitio distal serán más fácil de desplazar, dejando libre la cavidad para las moléculas de ligandos exógenos que puedan ingresar.

De igual manera, la energía de interacción correspondiente más favorable entre el agua y los aminoácidos del sitio activo sugiere que la barrera para la unión del CO debe ser mayor para la proteína wt que para las mutantes en las cuales el WG8 es reemplazado por un aminoácido que no provee de un puente de hidrógeno a la molécula de agua. Este resultado concuerda perfectamente con el escenario revelado por las curvas de re-unión de CO. Una sencilla comparación entre la recombinación geminada en la escala de picosegundos para las tres

proteínas (figura III.3 y tabla III.1) confirma que la re-unión en la forma wt ocurre con la constante aparente de tiempo más larga, es decir, con la barrera energética más grande. Para las mutantes estudiadas, la fase de la recombinación geminada es mayor, tanto en amplitud como en velocidad comparado con la proteína wt. En este sentido, la constante temporal decrece cuando se remueve al WG8 e incrementa ligeramente con fenilalaninas introducidas en las posiciones estructurales B10 y CD1. La tabla III.1 muestra que k_{-1} experimenta un aumento de 7 veces cuando el WG8 se cambia por fenilalanina.

Las curvas de re-unión para las mutantes muestran además diferentes fases de recombinación bimolecular (la fase siguiente a la recombinación geminada), que se originan principalmente por dos eventos secuenciales, seguido del límite difusional impuesto por la proteína y el ligando diatómico. En primer lugar, el ligando debe re-ingresar desde el seno del solvente a la cavidad distal de la proteína, a través del portal E7 y sus cavidades asociadas. En segundo lugar, el ligando encontrará moléculas de agua en el sitio distal, las cuales le impondrán una barrera de energía libre que necesitará superar para poder llegar al Fe hémico. Sobre la base del estudio teórico presente, se espera que dicha barrera sea diferente para las proteínas wt y las mutantes. La primera barrera energética, determinada por el portal E7 a través del cual debe migrar el ligando, se espera que esté fuertemente acoplada a la dinámica de la proteína y las cavidades internas que la compongan. Para analizar similitudes y diferencias del portal E7 de las mutantes, junto a la accesibilidad dinámica de la cavidad distal, comparándolo con el de la proteína wt, se utilizó el método de ILS. Además, con el programa POVME[30] se estimó el volumen de cada sistema de cavidades. En todos los casos se identificaron tres cavidades principales, $(trHb : CO)_1$, $(trHb : CO)_2$ y $(trHb : CO)_3$. El volumen total promedio calculado a lo largo de DMs de 90ns resultó ser prácticamente el mismo en la wt (134 $Å^3$) y mutante WG8F (137 $Å^3$), mientras que para la mutante 3F el volumen aumenta a 149 $Å^3$. La figura III.5 muestra los canales dinámicos muestreados en las simulaciones de DM y detectados con el programa ILS para las tres proteínas estudiadas. Un profundo análisis de las simulaciones incluyendo los resultados de ILS ha revelado que la mutante WG8F exhibe dos estados conformacionales que se llamarán, de aquí en adelante, I y II, mostrados en la figura III.5A. Durante el tiempo de simulación, las poblaciones correspondientes a ambas conformaciones fueron similares. La conformación I presenta un sistema de cavidades que permanece muy similar al observado para la proteína wt (figura III.5C) y, si bien la conexión entre las cavidades parece ser ligeramente diferente, también muestra una conexión abierta con el solvente, permitiendo así el intercambio del ligando con el exterior de la proteína. La conformación II implica un reordenamiento del sitio

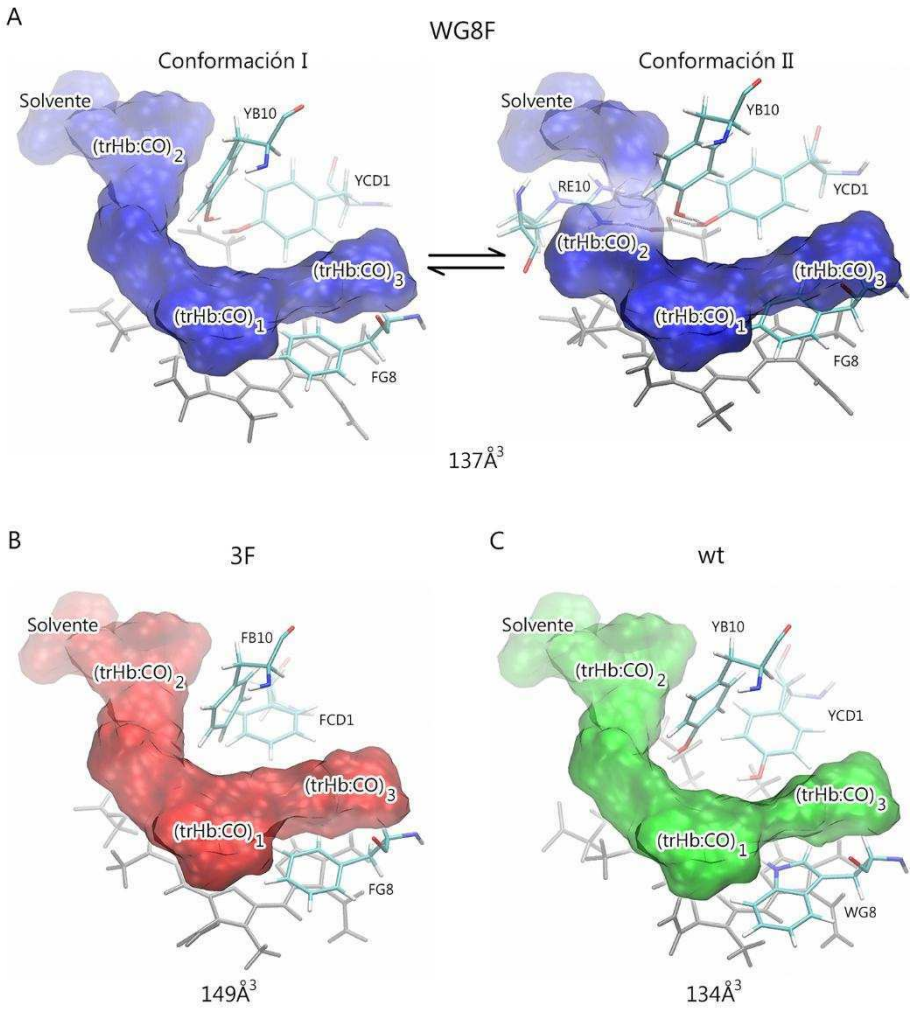


Figura III.5. Representaciones esquemáticas de los aminoácidos del sitio distal, el portal E7 y el sistema de cavidades asociado para las proteínas mutantes WG8F (A) y 3F (B) y wt (C) de la Tf-trHbO. En (A) puede observarse un equilibrio conformacional entre dos estados o conformaciones. Debajo de cada caso se muestran los volúmenes (en $Å^3$) estimados para cada sistema de cavidades.

distal en el cual la YCD1 es capaz de formar una red de puentes de hidrógeno con la RE10, la YB10 y un grupo propianato del hemo. En esta situación, la cavidad (trHb : CO)$_2$ no solo se achica, sino que además se acerca unos 3 Å a la cavidad (trHb : CO)$_1$, debido a la ruptura transiente y parcial del portal E7. Estas modificaciones en la conformación II mejoran la comunicación entre estas cavidades, pero resultan en un pasaje más estrecho hacia el solvente.

En busca de cuantificar más detalladamente la contribución de los aminoácidos mutados a la migración del ligando a través del portal E7, se determinaron los perfiles de energía libre para migración de CO, mostrados en la figura III.6. La energía libre fue fijada en valores de 0 a 9 Å del Fe hémico, donde el CO está completamente solvatado. Los perfiles calculados no presentan barreras significativas para la entrada del CO a la cavidad (trHb : CO)$_2$ a través del canal, en ninguno de los tres casos. Puede observarse una pequeña barrera para llegar a la cavidad (trHb : CO)$_1$, con cambios casi insignificantes entre proteínas cuando se considera la energía libre a aproximadamente 5 Å del Fe hémico. El pozo de energía libre a ~2 Å corresponde a la cavidad distal, muy polar en la proteína wt, y completamente apolar en la mutante 3F. Se observa una

diferencia apreciable en cuanto a la profundidad de este pozo, donde el CO está justo arriba del Fe para la formación del enlace de coordinación.

La comparación de las barreras energéticas para pasar desde la cavidad distal, (trHb : CO)$_1$, a la (trHb : CO)$_3$ sugiere valores similares, siendo la asociada a la mutante WG8F la más baja. El mínimo energético asociado a la cavidad en (trHb : CO)$_1$ de la proteína WG8F es el más alto en energía, debido a que el WG8, el principal aminoácido responsable de la estabilización del ligando, se ha perdido. La barrera para la reacción inversa, (trHb : CO)$_3$ $\rightarrow$ (trHb : CO)$_1$, es levemente afectada por las mutaciones, siendo la correspondiente a la Tf-trHbO 3F, la más baja. De acuerdo a los datos arrojados por los perfiles de energía libre, las barreras para el paso de (trHb : CO)$_1$ a (trHb : CO)$_3$ son generalmente menores que de (trHb : CO)$_1$ a (trHb : CO)$_2$, lo cual conduciría a un escape preferencial hacia (trHb : CO)$_3$ antes que (trHb : CO)$_2$, contribuyendo así a la alta proporción de recombinación geminada observada experimentalmente.

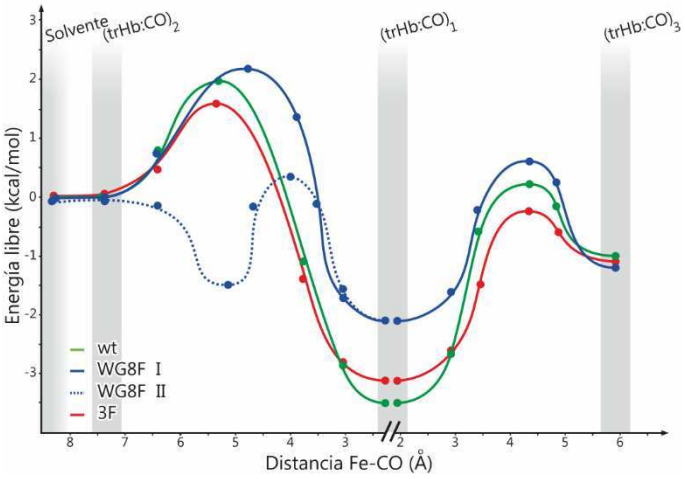


Figura III.6. Perfiles de energía libre a lo largo del portal E7, partiendo desde el solvente y migrando a través de las cavidades (trHb : CO)$_2$, (trHb : CO)$_1$ y (trHb : CO)$_3$. Se muestran los resultados para las proteínas wt y mutantes WG8F (ambas conformaciones) y 3F, en verde, azul y rojo, respectivamente. Los círculos representan los valores de energía libre calculados con ILS, y las líneas corresponden a un ajuste estimado sobre dichos valores calculados.

Es interesante notar que hay una clara variación en el perfil de energía libre para la conformación II de la mutante WG8F, respecto a la conformación I. Como se explicó anteriormente, esto es una consecuencia del reordenamiento del sitio distal en el cual la cavidad (trHb : CO)$_2$ se achica. Aunque las distintas barreras (comparando ambas conformaciones I y II) a lo largo del canal entre el solvente y la cavidad distal sugieren que la salida hacia y la re-unión desde el solvente deberían ocurrir con dos velocidades diferentes (dos constantes microscópicas distintas), no se observó dicha heterogeneidad en la fase bimolecular de la cinética de re-unión. Esto puede deberse al hecho de que ambos confórmeros se encuentran en un equilibrio de una escala temporal mucho más rápida que la de la re-unión bimolecular. Por lo tanto, las constantes microscópicas obtenidas reflejan valores promedio de las dos conformaciones.

Mientras que las constantes de ingreso k_c y k_d muestran un incremento moderado de la mutante WG8F a la 3F, las constantes de salida k_{-c} y k_{-d} experimentan un incremento grande (~100 veces). Este incremento se corresponde, de manera cualitativa, con los perfiles de energía

libre de la figura III.6. Una fuente adicional de modulación de velocidad, tomando como ejemplo a la constante k_{-1}, (trHb : CO)$_1$ $\rightarrow$ (trHb-CO), puede ser el tamaño de la cavidad distal. En la mutante Tf-trHbO 3F, la cavidad distal resulta ser la más voluminosa comparándola con las otras dos proteínas, lo que deriva en una mayor entropía del sistema. Esto puede también aportar a una leve estabilización del ligando en la cavidad, lo cual podría explicar la recombinación geminada más lenta en comparación con la mutante Tf-trHbO WG8F.

Estudios computacionales llevados a cabo por otros investigadores han propuesto la hipótesis de que las fenilalaninas podrían jugar un rol importante en la modulación de la migración del CO[31–33]. Utilizando un campo de fuerzas modificado, los investigadores han encontrado, por ejemplo, que las interacciones de *stacking* entre el CO y la cadena lateral de una fenilalanina hacen que el CO se mantenga cerca de los anillos de tres fenilalaninas de mioglobina (F33, F43 y F46)[34]. Aunque ésta puede no ser la única razón por la disminuida velocidad de la recombinación geminada para la k_{-1} en la Tf-trHbO 3F, la presencia de tres anillos fenilos en la cavidad distal refuerza la estabilización proveniente del aumento de tamaño en la cavidad.

Una cuestión clave conectada a la pérdida de cualquier interacción polar posible proveniente de la proteína en la cavidad distal es que las moléculas de solvente no están estabilizadas, por lo que es completamente razonable observar una marcada disminución en las barreras cinéticas entre cavidades en la proteína Tf-trHbO 3F.

III.3.3 G8: una posición estructural estratégica para la modulación de cavidades internas y moléculas de agua

Como se mencionó previamente en la sección Introducción (III.1), la Mt-trHbN presenta un sitio activo con dos aminoácidos polares, YB10 y EE11, y una valina en G8. Además, posee bloqueado el portal E7 y abierto los otros dos canales para la migración de ligandos pequeños. En esta sección se mostrarán resultados de simulaciones de DM de la proteína wt y de dos mutantes en la posición estructural G8, VG8F y VG8W. Se hará especial hincapié en el rol de dicha posición estructural sobre la modulación de los dos factores estudiados en profundidad en la sección anterior.

Colaboradores de este trabajo realizaron mediciones por flujo detenido de las constantes cinéticas de asociación de CO para las tres proteínas, la Mt-trHbN wt y las mutantes únicas VG8F y VG8W, a fin de obtener los valores de k_{on} correspondientes. Se observó una marcada disminución en el valor de k_{on} en la mutante Mt-trHbN VG8F, e incluso más acentuada en la

mutante de VG8W, modificándose de $4495\,mM^{-1}s^{-1}$ a $715\,mM^{-1}s^{-1}$ y $48\,mM^{-1}s^{-1}$, respectivamente.

Cálculos de ILS sobre la Mt-trHbN wt mostraron efectivamente ambos canales conocidos disponibles para la migración de CO, el canal largo y el corto G8 (figura III.7A). Se encontró que el canal largo conecta tres cavidades: $(trHb : CO)_1$, $(trHb : CO)_2$ y $(trHb : CO)_3$; mientras que el canal corto G8, por su parte, involucra a la cavidad $(trHb : CO)_1$, la cual está conectada con el solvente a través de la cavidad $(trHb : CO)_2$.

Los perfiles de energía libre calculados (figura III.7B y C) muestran que la mutante Mt-trHbN VG8F conserva ambos canales, aunque con una restricción moderada comparándolos con la wt, puesto que la FG8 aumenta la barrera energética para pasar a la cavidad $(trHb : CO)_1$. En el caso

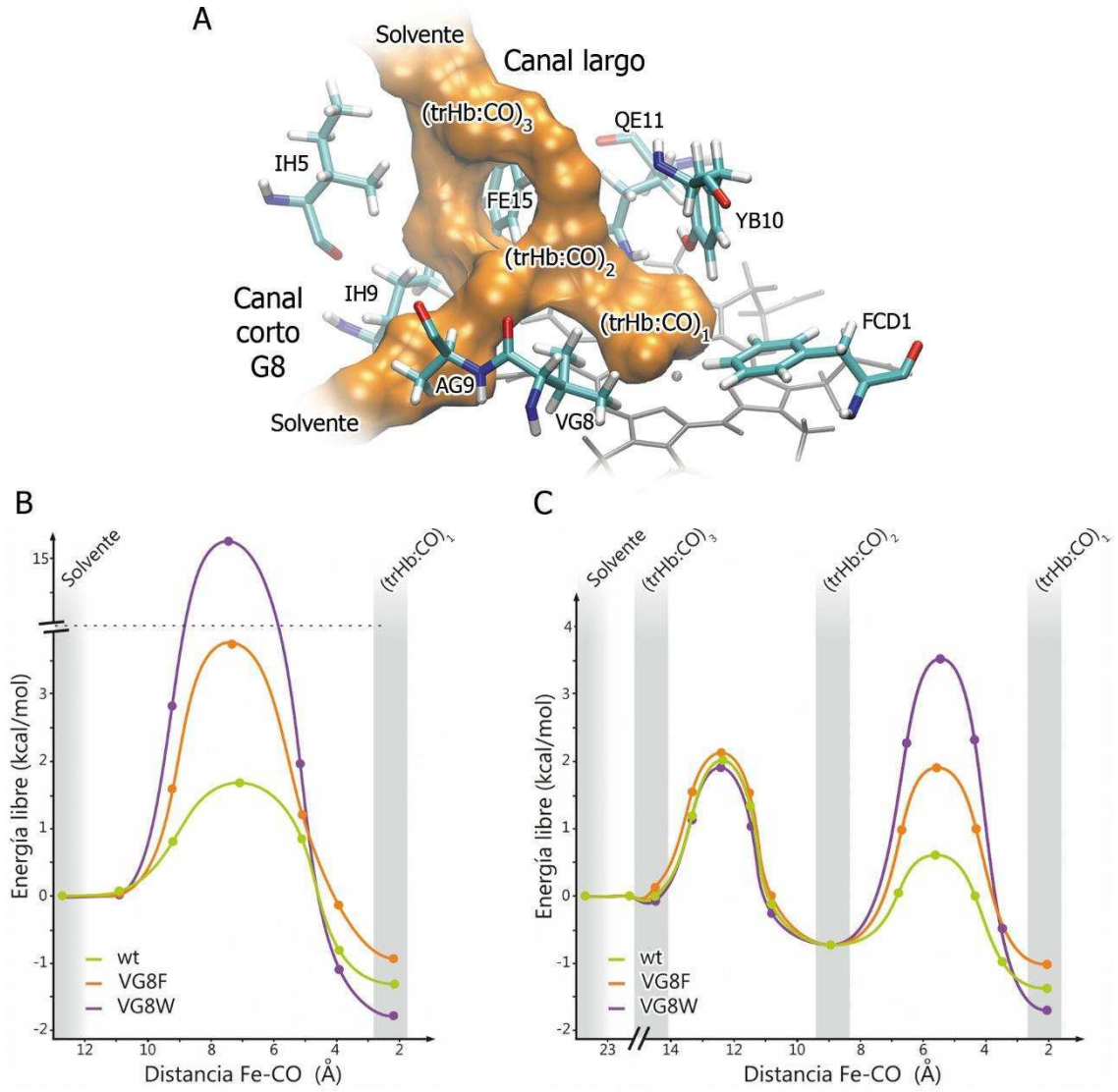


Figura III.7. Migración de CO a través de los dos posibles canales en la Mt-trHbN. (A) Representación esquemática de los aminoácidos involucrados en el sitio distal y en los túneles, los dos túneles y sus cavidades asociadas para la proteína wt. Se muestran perfiles de energía libre sobre el canal corto G8 (B) y canal largo (C) desde el solvente

conectando las cavidades en la wt (verde) y las mutantes VG8F (naranja) y VG8W (violeta). Los círculos representan valores calculados de energía libre con ILS y las líneas corresponden a un ajuste de dichos valores.

de la Mt-trHbN VG8W, sin embargo, los perfiles sugieren un canal corto G8 completamente bloqueado y un canal largo a través del cual la accesibilidad a la cavidad distal se encuentra notablemente reducida.

En términos de energía libre, la Mt-trHbN wt presenta pequeñas barreras (~2 $kcal\ mol^{-1}$) para la migración de CO desde el solvente al sitio activo a través de ambos canales. El incremento notable en estas barreras para ambas mutantes, del doble en la VG8F y de 7 veces más en la VG8W, es consistente con las tendencias observadas siguiendo los cambios en las mediciones experimentales. Sin embargo, aún resta conocer la estructura de solvatación de cada cavidad distal. En cada caso, una molécula de agua fue capaz de ingresar al sitio activo y ser estabilizada por aminoácidos polares. Específicamente, tanto en la proteína wt como en la mutante VG8F, dicha molécula de agua ocupó el sitio de unión a CO durante el 40% del tiempo de simulación (figura III.8A y B). Mientras que en el caso de la Mt-trHbN VG8W, dicho sitio fue ocupado durante un 80% del tiempo simulado, gracias al puente de hidrógeno adicional del WG8 (figura III.8C).

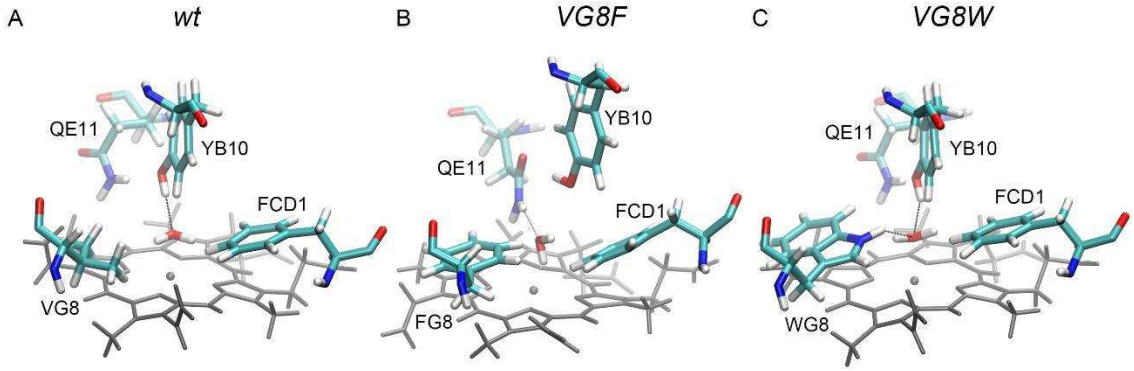


Figura III.8. Representación esquemática del sitio activo de las proteínas wt (A) y las mutantes VG8F (B) y VG8W (C) de la Mt-trHbN mostrando, sobre la base de las simulaciones de DM, la red de puentes de hidrógeno (líneas punteadas) estabilizando una molécula de agua sobre el Fe hémico.

Estos resultados teóricos demuestran el rol fundamental de la posición estructural G8 sobre la migración del CO y la estabilización de moléculas de aguas retenidas en el sitio activo. Notoriamente, la mutación estratégica le permite a la Mt-trHbN wt modificar su constante de asociación al orden de la observada en la trHb wt del grupo O (Mt-trHbO), dos órdenes de magnitud más lenta: $48\ mM^{-1}s^{-1}$ para la Mt-trHbN VG8W y $13\ mM^{-1}s^{-1}$ para la Mt-trHbO wt (siendo $13\ mM^{-1}s^{-1}$ la contribución mayor, 79%, de un ajuste bifásico, y $180\ mM^{-1}s^{-1}$ la menor, 21%[35]). Cabe notar que, al menos en términos estructurales, esto no quiere decir que se ha transformado la trHbN en trHbO, solo se ha obtenido un valor del mismo orden de magnitud

de k_{on}. Es importante recordar que el canal previamente reportado para la migración de CO en la trHbO es el portal E7, y no el canal largo[36], y, por otro lado, si bien la trHbO tiene un triptófano en la posición G8 y a la invariable YB10, el resto del sitio activo y canales poseen una identidad química distinta.

Todos los resultados teóricos y experimentales presentados en este capítulo fueron publicados de manera conjunta en los siguientes tres artículos:

1- A. Marcelli, S. Abbruzzetti, **JP. Bustamante**, A. Feis, A. Bonamore, A. Boffi, C. Gellini, P.R. Salvi, D.1. Estrin, S. Bruno, C. Viappiani, P. Foggi. *Following Ligand Migration Pathways from Picoseconds to Milliseconds in Type II Truncated Hemoglobin from Thermobifida fusca*. **PLoS ONE**, 2012;7(7):e39884. doi:10.1371/journal.pone.0039884.

2- **JP. Bustamante**, S. Abbruzzetti, A. Marcelli, D. Gauto, L. Boechi, A. Bonamore, A. Boffi, S. Bruno, A. Feis, P. Foggi, DA. Estrin, C. Viappiani. *Ligand Uptake Modulation by Internal Water Molecules and Hydrophobic Cavities in Hemoglobins*. **J. Phys. Chem. B**, 2014, 118 (5), pp 1234–1245.

3- I. Boron*, **JP. Bustamante***, KS. Davidge, S. Singh, F. Forti, LAH. Bowman, M. Tinajero-Trejo, R. Poole, K. Dikshit, DA. Estrin, MA. Marti, L. Boechi. Ligand uptake by *M. tuberculosis* truncated hemoglobin N is modulated by both two tunnels and retained water molecules. **F1000Research**, 2015, 4:22. doi: 10.12688/f1000research.5921.1. *Iguales contribuciones.

III.4 Conclusiones

El análisis de uno de los dos factores cruciales en el proceso de captación de ligandos, el sistema de cavidades permanentes y transitorias, ha permitido identificar al portal E7 como el canal responsable del acceso de ligandos en la Tf-trHbO. En base a los resultados presentados, ha sido posible proponer un esquema cinético asociado al proceso de captación. La complementariedad entre el enfoque teórico y los experimentos llevados a cabo por los grupos colaboradores ha permitido validar satisfactoriamente este esquema cinético propuesto. Partiendo de éste, se estudiaron las contribuciones moleculares de ambos factores claves en el proceso de captación de ligandos: las cavidades internas y el rol de moléculas de agua en el sitio activo. En este sentido, el intercambio del ligando entre el sitio de unión y el solvente se mostró regulado por un sistema dinámico de cavidades con un canal de ingreso / egreso que presenta cavidades transitorias. Por otro lado, se corroboró que la presencia de moléculas de agua

estabilizadas por aminoácidos del sitio activo modula la unión de ligandos, como lo han evidenciado las barreras de los pasos críticos en la cinética de re-unión en la Tf-trHbO. La mutación de aminoácidos involucrados en la estabilización de moléculas de agua en el sitio distal ha resultado en una reducción drástica de estas barreras energéticas.

Adicionalmente, el estudio sobre la Mt-trHbN permitió elucidar el rol clave de la posición estructural G8 tanto en la alteración de la migración de ligandos, como en la retención de moléculas de agua en el sitio activo. Si bien la k_{on} de la Mt-trHbN VG8W resultó ser del mismo orden de magnitud que la de su homóloga trHbO, los procesos de asociación ocurren a través de canales distintos y con sitios activos diferentes.

En resumen, los estudios aquí presentados han permitido no solo la caracterización, sino también la comprensión molecular del grado de importancia de dos factores cruciales que definen el proceso de captación de ligandos pequeños. Dichos análisis integradores realzan la utilidad de simulaciones de DM complementadas con cinéticas de unión de ligandos, lo cual ha facilitado una caracterización muy detallada, que de otra manera sería difícil de alcanzar.

Obtener una correspondencia cuantitativa entre los resultados experimentales y teóricos expuestos no es una tarea sencilla, ya que esto implicaría conocer el delicado balance entre las variables que definen el proceso de asociación y la relación exacta entre éstas, estudio que será abordado en profundidad en el siguiente capítulo.

III.6 Referencias

(1) Tilton, R.; Kuntz, I. *Biochemistry* **1982**, *21* (26), 6850–6857.

(2) Scott, E. E.; Gibson, Q. H.; Olson, J. S. *J. Biol. Chem.* **2001**, *276*, 5177–5188.

(3) Brunori, M.; Vallone, B.; Cutruzzola, F.; Travaglini-Allocatelli, C.; Berendzen, J.; Chu, K.; Sweet, R.; Schlichting, I. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97* (5), 2058–2063.

(4) Brunori, M. *Biophys. Chem.* **2000**, *86* (2-3), 221–230.

(5) Milani, M.; Pesce, A.; Ouellet, Y.; Dewilde, S.; Friedman, J.; Ascenzi, P.; Guertin, M.; Bolognesi, M. *J. Biol. Chem.* **2004**, *279* (20), 21520–21525.

(6) Li, Z.; Lazaridis, T. *J. Phys. Chem. B* **2005**, *109* (1), 662–670.

(7) Li, Z.; Lazaridis, T. *J. Am. Chem. Soc.* **2003**, *125*, 6636–6637.

(8) Michel, J.; Tirado-Rives, J.; Jorgensen, W. *J. Phys. Chem. B* **2009**, *113* (40), 13337–13346.

(9) Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. *J. Am. Chem. Soc.* **2008**, *130* (9), 2817–2831.

(10) Schmidtke, P.; Luque, F.J.; Murray, J. B.; Barril, X. *J. Chem. Theory Comput.* **2011**, *133* (46), 18903–18910.

(11) Hummer, G. *Nat. Chem.* **2010**, *2* (11), 906–907.

(12) Goldbeck, R. a; Bhaskaran, S.; Ortega, C.; Mendoza, J. L.; Olson, J. S.; Soman, J.; Kliger, D. S.; Esquerra, R. M. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (5), 1254–1259.

(13) Baron, R.; Setny, P.; McCammon, J. A. *J. Am. Chem. Soc.* **2010**, *132* (34), 12091–12097.

(14) Droghetti, E.; Nicoletti, F. P.; Bonamore, A.; Boechi, L.; Arroyo Mañez, P.; Estrin, D. a; Boffi, A.; Smulevich, G.; Feis, A. *Biochemistry* **2010**, *49* (49), 10394–10402.

(15) Eich, R. F.; Li, T.; Lemon, D. D.; Doherty, D. H.; Curry, S. R.; Aitken, J. F.; Mathews, A. J.; Johnson, K. A.; Smith, R. D.; Phillips, G. N.; Olson, J. S. *Biochemistry* **1996**, *35* (22), 6976–6983.

(16) Ouellet, Y. H.; Daigle, R.; Lagüe, P.; Dantsker, D.; Milani, M.; Bolognesi, M.; Friedman, J. M.; Guertin, M. *J. Biol. Chem.* **2008**, *283* (40), 27270–27278.

(17) Bonamore, A.; Ilari, A.; Giangiacomo, L.; Bellelli, A.; Morea, V.; Boffi, A. *FEBS J.* **2005**, *272* (16), 4189–4201.

(18) Nicoletti, F. P.; Comandini, A.; Bonamore, A.; Boechi, L.; Boubeta, F. M.; Feis, A.; Smulevich, G.; Boffi, A. *Biochemistry* **2010**, *49* (10), 2269–2278.

(19) Nicoletti, F. P.; Droghetti, E.; Boechi, L.; Bonamore, A.; Sciamanna, N.; Estrin, D. a; Feis, A.; Boffi, A.; Smulevich, G. *J. Am. Chem. Soc.* **2011**, *133* (51), 20970–20980.

(20) Milani, M.; Pesce, A.; Ouellet, Y.; Ascenzi, P.; Guertin, M.; Bolognesi, M. *EMBO J.* **2001**, *20* (15), 3902–3909.

(21) Di Lella, S.; Martí, M. A.; Alvarez, R. M. S.; Estrin, D. A.; Ricci, J. C. D. *J. Phys. Chem. B* **2007**, *111* (25), 7360–7366.

(22) Gauto, D. F.; Di Lella, S.; Guardia, C. M. A.; Estrin, D. A.; Martí, M. A. *J. Phys. Chem. B* **2009**, *113* (25), 8717–8724.

(23) Gauto, D. F.; Di Lella, S.; Estrin, D. A.; Martí, M. A. *Carbohydr. Res.* **2011**, *346* (7), 939–948.

(24) Feis, A.; Lapini, A.; Catacchio, B.; Brogioni, S.; Foggi, P.; Chiancone, E.; Boffi, A.; Smulevich, G. *Biochemistry* **2008**, *47* (3), 902–910.

(25) Boechi, L.; Mañez, P. A.; Luque, F. J.; Marti, M. A; Estrin, D. A. *Proteins* **2010**, *78* (4), 962–970.

(26) Nardini, M.; Pesce, A.; Labarre, M.; Richard, C.; Bolli, A.; Ascenzi, P.; Guertin, M.; Bolognesi, M. *J. Biol. Chem.* **2006**, *281* (49), 37803–37812.

(27) Boechi, L.; Martí, M. A.; Milani, M.; Bolognesi, M.; Luque, F. J.; Estrin, D. A. *Proteins Struct. Funct. Genet.* **2008**, *73* (2), 372–379.

(28) Guallar, V.; Lu, C.; Borrelli, K.; Egawa, T.; Yeh, S. *J. Biol. Chem.* **2009**, *284* (5), 3106–3116.

(29) Nicoletti, F. P.; Droghetti, E.; Howes, B. D.; Bustamante, J. P.; Bonamore, A.; Sciamanna, N.; Estrin, D. A.; Feis, A.; Boffi, A.; Smulevich, G. *Biochim. Biophys. Acta - Proteins Proteomics* **2013**, *1834*, 1901–1909.

(30) Durrant, J. D.; de Oliveira, C. A. F.; McCammon, J. A. *J. Mol. Graph. Model.* **2011**, *29* (5), 773–776.

(31) Borrelli, K. W.; Vitalis, A.; Alcantara, R.; Guallar, V. *J. Chem. Theory Comput.* **2005**, 1304–1311.

(32) Cohen, J.; Arkhipov, A.; Braun, R.; Schulten, K. *Biophys. J.* **2006**, *91* (5), 1844–1857.

(33) Spyrakis, F.; Luque, F. J.; Viappiani, C. *Plant Sci.* **2011**, *181* (1), 8–13.

(34) Lucas, M. F.; Guallar, V. *Biophys. J.* **2012**, *102* (4), 887–896.

(35) Ouellet, H.; Juszczak, L.; Dantsker, D.; Samuni, U.; Ouellet, Y. H.; Savard, P.; Wittenberg, J. B.; Wittenberg, B. A.; Friedman, J. M.; Guertin, M. *Biochemistry* **2003**, *42*, 5764–5774.

(36) Boechi, L.; Mañez, P. A.; Luque, F.J.; Marti, M. A.; Estrin, D. A. *Proteins Struct. Funct. Bioformatics* **2010**, *78* (4), 962–970.

IV. Un modelo cuantitativo para la captación y liberación de O_2

IV.1 Introducción

La combinación de simulaciones computacionales y de experimentos bioquímicos se ha convertido en una muy buena estrategia para describir los detalles moleculares subyacentes a las diferencias observadas en la actividad de proteínas. Con el incremento en la cantidad de datos de proteínas individuales, la construcción de modelos matemáticos que permita una comprensión más generalizada de estos detalles moleculares se ha vuelto un desafío de gran interés para muchos grupos alrededor del mundo[1–4].

Dentro de la familia de trHbs (con aproximadamente 1100 miembros), solo un pequeño grupo cercano a 10 proteínas wt y varias mutantes han sido caracterizadas extensamente, aunque en estudios separados de una o dos proteínas a la vez, utilizando técnicas experimentales y simulaciones teóricas. En todos los casos, solo se han encontrado tendencias cualitativas respecto a datos experimentales de k_{on} y k_{off}. En este capítulo se presentará un modelo fisicoquímico que, incluyendo relaciones cuantitativas entre los factores moleculares que definen a la captación y liberación de ligandos pequeños, permita predecir constantes de k_{on} y k_{off} de O_2. Dicho modelo usa datos obtenidos exclusivamente de simulaciones a nivel molecular, como las descriptas en detalle en el capítulo anterior.

A través de un riguroso análisis estadístico llevado a cabo en colaboración con las Dras. Mariela Sued y María Eugenia Szretter del Instituto de Cálculo – UBA, se demostrará que el modelo desarrollado recaptura satisfactoriamente todas las constantes cinéticas experimentales reportadas de k_{on} y k_{off} de O_2. Esto permite establecer cuáles y en qué proporción los factores claves determinan las tasas a las cuales estas hemoproteínas captan y liberan O_2. Basados en el modelo presentado, se propondrán además dos estrategias para predecir las constantes cinéticas para el resto de los miembros de la familia de las trHbs.

IV.2 Métodos computacionales

IV.2.1 Configuración de los sistemas y parámetros de simulación

Las estructuras cristalográficas correspondientes a todas las proteínas utilizadas corresponden a los siguientes identificadores de la base de datos PDB: 1IDR (Mt-trHbN, perteneciente a *Mycobacterium tuberculosis*), 1NGK (Mt-trHbO), 2BMM (Tf-trHbO, perteneciente a *Thermobifida fusca*), 4UUR (Ph-trHbO, perteneciente a *Pseudoalteromonas haloplanktis*), 2XYK (At-trHbO, perteneciente a *Agrobacterium tumefaciens*), 1UX8 (Bs-trHbO, perteneciente a *Bacillus subtilis*), 1DLY (Ce-trHbN, perteneciente a *Chlamydomonas eugametos*), 1DLW (Pc-trHbN, perteneciente a *Paramecium caudatum*), 2BKM (Gs-trHbO perteneciente a *Geobacillus stearothermophilus*), 1RTX (Syn-trHbN perteneciente a *Synechocystis*), 2IG3 (Cj-trHbP, perteneciente a *Campylobacter jejuni*) y 1VXG (perteneciente a la mioglobina de cachalote). Para todas ellas, así como también para algunas formas mutantes, se siguió el protocolo presente en la sección II.4.3 del capítulo de métodos. Las simulaciones de DM realizadas fueron de 50ns cada una.

IV.2.2 Topología y perfiles de energía libre de migración de O_2

Los cálculos de ILS sobre los sistemas proteicos se realizaron siguiendo el protocolo presentado en la sección II.4.3. La sonda utilizada, en todos los casos, fue una molécula de O_2.

La energía libre para los perfiles de energía fue configurada en $0\ kcal\ mol^{-1}$ para el O_2 en el seno del solvente, a 9, 13 y 24 Å del Fe hémico para el portal E7, el canal corto G8 y el canal largo, respectivamente.

IV.2.3 Energía de unión de O_2 (ΔE_{O_2} y $\Delta\Delta E_{O_2}$)

Los cálculos de energía de unión de O_2 se computaron de acuerdo al protocolo presentado en la sección II.4.4.

IV.3 Resultados y discusión

IV.3.1 Modelado de la constante de captación del ligando (k_{on})

La ecuación de Eyring es ampliamente usada para modelar reacciones bimoleculares en el contexto de la teoría del estado de transición:

$$k = \frac{k_B T}{h} \cdot e^{-\frac{\Delta G^{\ddagger}}{RT}} \quad (1)$$

donde $\Delta G^{\ddagger}$ es la energía de activación de Gibbs, T es la temperatura, k_B, h y R, son las constantes de Boltzmann, Planck y la constante universal de los gases, respectivamente.

Se construyó un modelo de Eyring que incluye la barrera de energía libre para el ingreso del ligando a través de los túneles de migración, así como también la penalidad en energía libre de remover una molécula de agua en el sitio activo. Es importante resaltar que una estimación más detallada de las tasas de asociación observadas debería incluir también una estimación de la barrera energética asociada con el proceso de formación del enlace de sexta coordinación entre el O_2 y el átomo de Fe del grupo hemo ($k_{unión}$). Sin embargo, dado que estudios teóricos previos han mostrado que parece haber solo una pequeña barrera energética debido a una transición de espín quinteto a singulete por parte del Fe, es esperable que su efecto no sea significativamente grande entre diferentes trHbs, todas conteniendo una HisF8 en el quinto enlace de coordinación[5,6].

En el modelo propuesto, los túneles de cada trHb se consideraron de manera independiente, y se seleccionó a la mayor barrera del túnel más accesible como la barrera de energía libre para el ingreso de O_2, denotada por $\Delta G_{in}^{\ddagger}$. Debido a que las moléculas de H_2O en el sitio activo están estabilizadas a través de puentes de hidrógeno, la energía libre para removerlas puede considerarse de manera aproximada como el producto entre la cantidad de puentes de hidrógeno (n_{pteH}) entre la molécula de H_2O y los aminoácidos del sitio activo y la energía libre de una interacción de pteH (ΔG_{pteH}). Las contribuciones individuales de $\Delta G_{in}^{\ddagger}$ y n_{pteH} se combinaron en el siguiente modelo fisicoquímico para el valor teórico de k_{on}:

$$k_{on} = \frac{k_B T}{h} \cdot e^{-\frac{\Delta G_{in}^{\ddagger}}{RT}} \cdot e^{-\frac{\Delta G_{pteH} \cdot n_{pteH}}{RT}} \quad (2)$$

donde T = 300K (27°C). Debido a que se reportaron valores diferentes de ΔG_{pteH} en la literatura (entre 2 y 5 $kcal\ mol^{-1}$)[7] dependiendo del ambiente inmediato, no se ha definido un valor específico para éste en el modelo propuesto.

Para validar estadísticamente el modelo presentado, se considerará la forma logarítmica del mismo, lo cual lo convertirá en un modelo lineal, y se asumirá que cada valor obtenido experimentalmente para cada proteína difiere del valor de k_{on} teórico en un error aleatorio aditivo (ε_i). Esto es, se asumirá que:

$$log\left(k_{on_i}\right) = log\left(\frac{k_B T}{h}\right) + \frac{(-1)}{RT} \cdot log(e) \cdot \Delta G_{in_i}^{\ddagger} + \frac{\left(-\Delta G_{pteH}\right)}{RT} \cdot log(e) \cdot n_{HB_i} + \varepsilon_i \quad (3)$$

donde el índice i corresponde a la $i^{ésima}$ proteína. Es importante notar que el error aleatorio (ε_i) considera errores de la medición, variabilidad biológica y posibles equivocaciones o incompletitud en el modelo.

Hay dos aspectos involucrados en la construcción del modelo 3. Por un lado, el modelo sugiere que $\Delta G_{in}^{\ddagger}$ y n_{pteH} son las variables que permiten describir de manera lineal el comportamiento (en escala logarítmica) de k_{on}. Por otro lado, esto establece de qué manera estas variables deben combinarse, determinando el valor del coeficiente que acompaña a $\Delta G_{in}^{\ddagger}$, así como la intersección, $log(k_B T/h)$. Para validar estadísticamente cada uno de estos dos aspectos, se considerará un modelo lineal en escala logarítmica en el cual serán usadas las mismas variables que en el modelo 3, $\Delta G_{in}^{\ddagger}$ y n_{pteH}, pero no se impondrá ningún tipo de restricciones a los coeficientes asociados:

$$log(k_{on_i}) = \beta_0 + \frac{\beta_1}{RT} \cdot log(e) \cdot \Delta G_{in_i}^{\ddagger} + \frac{\beta_2}{RT} \cdot log(e) \cdot n_{pteH_i} + \varepsilon_i \qquad (4)$$

Notar que si se reemplazan $\beta_0 = log(k_B T/h)$, $\beta_1 = -1$ y $\beta_2 = -\Delta G_{pteH}$ en el modelo 4, se recupera el modelo 3. Se ajustó el modelo 4 para obtener el mejor modelo lineal para los valores de k_{on} reportados para trHbs y mioglobina de cachalote (Mb) (n = 26, tabla A.IV.1). No se impuso ninguna restricción sobre los coeficientes β_0, β_1 y β_2, de manera tal que fueron dejados libres de asumir cualquier valor.

Una vez encontrado el mejor ajuste para el modelo 4 mediante regresión de cuadrados mínimos, se evaluó la precisión de la descripción lineal de $log(k_{on})$ con las variables consideradas $\Delta G_{in}^{\ddagger}$ y n_{pteH}, así como los valores propuestos para los coeficientes en el modelo 3. El R^2 hallado fue de 0,78, lo que significa que el 78% de la variabilidad de $log(k_{on})$ puede explicarse a nivel lineal en términos de ambas variables explicativas. La desviación estándar del error ε_i para el modelo 4 fue 0,56, la cual es realmente pequeña comparada con los valores observados para la variable respuesta ($log\ (k_{on})$).

Todos los coeficientes β_0, β_1 y β_2, resultaron ser estadísticamente significativos (los p-valores correspondientes son menores a 0.05), lo que significa que hay evidencia en los datos apoyando que cada coeficiente es diferente de cero en el modelo que incluye todas las otras variables explicativas. Esta afirmación asegura que ambas variables son relevantes para explicar $log\ (k_{on})$. Más aún, ambas variables mostraron estar muy poco correlacionadas (la correlación cuadrada entre las variables $\Delta G_{in}^{\ddagger}$ y n_{pteH} es 0.09), demostrando que cubren diferente información. Esto refuerza la importancia de incluirlas a ambas en el modelo. En la sección IV.2.2 se presenta la validación de las asunciones requeridas para los valores estadísticos (p-valores e intervalos de confianza) mostrados en la tabla IV.1.

Tabla IV.1. Valores estimados para los coeficientes con sus respectivos intervalos de confianza (IC) del 95% y p-valores para el modelo 4.

Ecuación	Coeficiente mejor ajustado	Coeficientes propuestos	Valor estimado ± IC del 95%	Valor p
4 (R^2 = 0.78)	β_0	$log\left(\frac{k_B T}{h}\right) = 12.796$	11.81 ± 1.15	$< 2 \cdot 10^{-16}$
	β_1	coeficiente $\Delta G_{in}^{\ddagger} = -1$	-1.25 ± 0.40	$1.65 \cdot 10^{-6}$
	β_2	$-\Delta G_{pteH} = [-5, -2]$	-1.99 ± 0.50	$3.14 \cdot 10^{-8}$

Otra manera de evaluar la importancia de ambas variables en el modelo consiste en modelar $log(k_{on})$ solo considerando $\Delta G_{in}^{\ddagger}$ o n_{pteH}. Se encontró una correlación muy pobre con los datos experimentales cuando $\Delta G_{in}^{\ddagger}$ (R^2 = 0.16) o n_{pteH} (R^2 = 0.40) fueron usados por separado. Los coeficientes (β_0 y β_1 o β_0 y β_2, dependiendo del modelo reducido), sin embargo, resultaron ser muy similares a los del modelo original (datos no mostrados). Esta evidencia refuerza el modelo lineal multivariado propuesto (modelo 4) que combina las dos variables independientes explicativas.

Los resultados anteriores permiten validar el modelo lineal y concluir que ambas variables no solo son necesarias para explicar la k_{on} en estas proteínas, sino que además, determinan los valores de k_{on} con una precisión alta.

En cuanto a los coeficientes propuestos, los resultados mostrados en la tabla IV.1 apoyan los valores teóricos de $log(k_B T/h)$ para β_0 y -1 para β_1, puesto que ambos, $log(k_B T/h)$ y -1 caen dentro de los respectivos intervalos de confianza del 95%. El intervalo de confianza del 95% para β_2 (-2.49, -1.48) también es compatible con el rango de valores para ΔG_{pteH} en la literatura[7].

Todos estos datos indican que el modelo de Eyring propuesto (modelo 2) por sí solo, junto con los coeficientes presentados (o su rango) es conceptualmente adecuado para describir el proceso de captación de O_2, lo que significa que la barrera energética impuesta por la matriz proteica, así como la estabilización de moléculas de H_2O en el sitio activo son los principales factores que controlan este proceso fisicoquímico.

IV.3.2 Modelado de la constante de liberación del ligando (k_{off})

Hay dos fenómenos que regulan el proceso de liberación: la ruptura de la unión Fe-O_2 (ΔE_{O_2}) y la posterior migración desde la cavidad distal hacia el solvente ($\Delta E_{out}^{\ddagger}$). Por consiguiente, se propone un modelo que incluye ambas variables, considerando $\Delta\Delta E_{O_2}$ en vez de ΔE_{O_2}(ver sección II.4.4 para una explicación detallada. Notar que la liberación de O_2 es una reacción unimolecular, por lo tanto el modelo propuesto es el siguiente:

$$k_{off} = e^{-\frac{\Delta\Delta E_{O_2}}{RT}} \cdot e^{-\frac{\Delta G^{\ddagger}_{out}}{RT}} \quad (5)$$

Este tipo de enfoque ya ha sido utilizado previamente de manera satisfactoria[8,9]. Como se llevó a cabo para el proceso de captación, se considerará un modelo lineal en escala logarítmica con las variables explicativas $\Delta\Delta E_{O_2}$ y $\Delta G^{\ddagger}_{out}$, dado por:

$$log(k_{off_i}) = \gamma_0 + \frac{\gamma_1}{\mathrm{RT}} \cdot log(e) \cdot \Delta\Delta E_{O_2 i} + \frac{\gamma_2}{\mathrm{RT}} \cdot log(e) \cdot \Delta G^{\ddagger}_{out_i} + \varepsilon_i \quad (6)$$

Asumiendo que las dos variables propuestas son suficientes para describir $log(k_{off})$, los valores teóricos para los coeficientes son $\gamma_0 = log(A)$, $\gamma_1 = -1$ y $\gamma_2 = -1$. Dejando a los coeficientes libres para que puedan tomar cualquier valor, puede ajustarse el modelo 6 para obtener el mejor modelo lineal para los datos disponibles (n = 13, tabla A.IV.2). Los valores ajustados se muestran en la tabla IV.2.

Tabla IV.2. Valores estimados para los coeficientes con sus respectivos intervalos de confianza (IC) del 95% y p-valores para los modelos 6 y 7.

Ecuación	Coeficiente mejor ajustado	Coeficientes propuestos	Valor estimado $\pm$ IC del 95%	Valor p
6 (R^2 = 0.787)	γ_0	Intersección $= log(A)$	-0.48 ± 2.78	0.71
	γ_1	coeficiente $\Delta\Delta E_{O_2} = -1$	-0.87 ± 0.45	$1.5 \cdot 10^{-3}$
	γ_2	coeficiente $\Delta G^{\ddagger}_{out} = -1$	0.17 ± 1.08	0.73
7 (R^2 = 0.784)	γ_0	Intersección $= log(A)$	-0.043 ± 0.64	0.89
	γ_1	coeficiente $\Delta\Delta E_{O_2} = -1$	-0.81 ± 0.28	$5.66 \cdot 10^{-5}$

Los datos muestran que si bien el valor de R^2 es alto (R^2 = 0.787), $\Delta G^{\ddagger}_{out}$ no contribuye para describir $log(k_{off})$ cuando $\Delta\Delta E_{O_2}$ está incluido en el modelo, puesto que el intervalo de confianza para γ_2 incluye al cero. Esto concuerda con el hecho de que las dos variables, $\Delta G^{\ddagger}_{out}$ y $\Delta\Delta E_{O_2}$, están altamente correlacionadas (la correlación cuadrada entre ambas es 0.55), sugiriendo que la consideración de ambas variables explicativas puede no ser necesaria. En tal sentido, se encontró que $\Delta\Delta E_{O_2}$ muestra ser relevante para describir $log(k_{off})$ debido a que su coeficiente, γ_1, es significativamente diferente de cero (su p-valor es menor que 0.05). Los datos también sugieren el valor 0 para γ_0. A continuación, se consideró entonces un modelo lineal sin la barrera energética $\Delta G^{\ddagger}_{out}$, dado por:

$$log(k_{off_i}) = \gamma_0 + \frac{\gamma_1}{\mathrm{RT}} \cdot log(e) \cdot \Delta\Delta E_{O_2 i} + \varepsilon_i \quad (7)$$

Los valores ajustados para los coeficientes γ_0 y γ_1 se muestran en la tabla IV.2. Una vez más, los datos sugieren el valor $\gamma_0 = 0$. La consideración única de $\Delta\Delta E_{O_2}$ como variable explicativa es apropiada para modelar $log(k_{off})$, puesto que este ajuste alcanza un R^2 = 0.784, basado en los datos disponibles (n = 13, A.IV.1). La desviación estándar de los errores ε_i en el modelo 7 fue estimada en 0.98.

Por todo lo presentado, se concluye que el modelo 6 es redundante, puesto que, dado $\Delta\Delta E_{O_2}$, $\Delta G^{\ddagger}_{out}$ no es relevante para describir el comportamiento lineal de $log(k_{off})$. El modelo 7, por otro lado, junto con los coeficientes propuestos, son conceptualmente adecuados para describir el proceso de liberación de O_2, dando cuenta de que solo $\Delta\Delta E_{O_2}$, y no el escape del O_2 a través de los túneles, es el factor dominante que controla este proceso fisicoquímico.

IV.3.3 Predicción de la constante de captación de O_2 (k_{on})

En las secciones anteriores se presentó un modelo fisicoquímico que fue validado comparándolo con constantes cinéticas experimentales, y que permitió hallar la relación entre los determinantes fisicoquímicos involucrados en la captación y liberación de O_2. En esta sección, se utilizarán dichos modelos para predecir constantes cinéticas desconocidas. La primera de las estrategias sienta sus bases en la fisicoquímica: toma los valores propuestos para β_0 y β_1 ($\beta_0 = log(\mathrm{k_B T/h})$ y $\beta_1 = -1$) y luego toma el valor de β_2 que da origen a los residuos más pequeños, una vez que se fijan β_0 y β_1. Se minimizó $\sum_{i=1}^{n}\left(log(k_{on_i}) - \left[log(\mathrm{k_B T/h}) - (1/\mathrm{RT}) \cdot log(e) \cdot \Delta G^{\ddagger}_{in_i} + (\beta_2/\mathrm{RT}) \cdot log(e) \cdot n_{pteH_i}\right]\right)^2$, y se encontró que $\beta_2 = -2.95$. Por consiguiente, el primer enfoque para predecir k_{on} es:

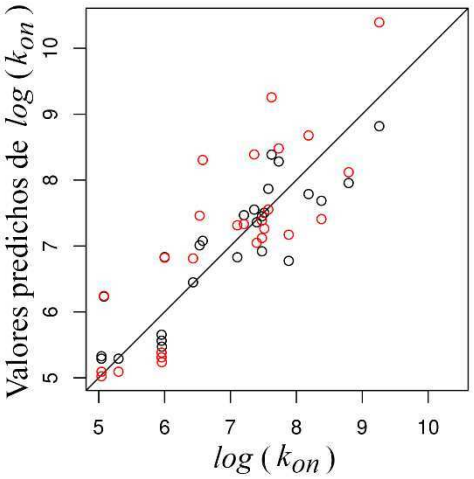
$$k_{on}chem = \frac{k_B T}{h} \cdot e^{-\frac{\Delta G^{\ddagger}_{in}}{RT}} \cdot e^{-\frac{2.95 \cdot n_{pteH}}{RT}} \quad (8)$$

Un enfoque diferente para la predicción es el que está completamente orientado por la naturaleza de los datos, ignorando un conocimiento previo de las constantes fisicoquímicas relevantes, y usando solo los coeficientes que mejor ajusten (tabla IV.2):

$$k_{on}stat = 10^{11.81} \cdot e^{-\frac{1.25 \cdot \Delta G^{\ddagger}_{in}}{RT}} \cdot e^{-\frac{1.99 \cdot n_{pteH}}{RT}} \quad (9)$$

En ambos enfoques, los valores propuestos para ΔG_{pteH}, 2.95 y 1.99, están en el rango de los valores publicados en la literatura. La figura IV.1 muestra la precisión de ambos enfoques predictivos, $log(k_{on}chem)$ y $log(k_{on}stat)$, mostrando su habilidad para asignar correctamente el orden de las constantes de captación experimentales. Ambos enfoques resultan en predicciones muy similares sumamente comparables, destacando que las constantes predichas no difieren de los valores reportados experimentalmente por más de un orden de magnitud en promedio.

Figura IV.1. Gráfica de valores experimentales de $log(k_{on})$ versus $log(k_{on}chem)$ y $log(k_{on}stat)$, en rojo y negro, respectivamente, con la línea identidad.



IV.3.4 Predicción de la constante de liberación de O_2 (k_{off})

De manera similar que para el caso de k_{on}, aquí se presentan dos enfoques para predecir valores de k_{off}. El primer enfoque, nombrado $k_{off}chem$, obedece al modelo fisicoquímico:

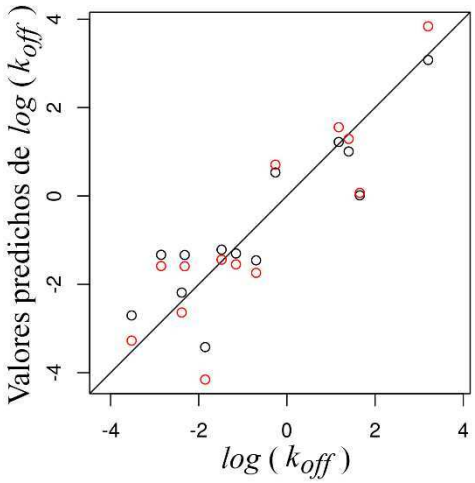
$$k_{off}chem = e^{-\frac{\Delta\Delta E_{O_2}}{RT}} \quad (10)$$

Mientras que el segundo enfoque, $k_{off}stat$, derivado del modelo ajustado desde la estadística, está dado por:

$$k_{off}stat = 10^{-0.04} e^{-\frac{0.81 \cdot \Delta\Delta E_{O_2}}{RT}} \quad (11)$$

En la figura IV.2 se grafican los valores k_{off} experimentales versus ambos $k_{off}chem$ y $k_{off}stat$ en escala logarítmica. Ambos modelos predictivos fueron capaces de asignar correctamente el orden de las constantes experimentales predichas.

Figura IV.2. Gráfica de valores experimentales de $log(k_{off})$ versus $log(k_{off}chem)$ y $log(k_{off}stat)$, en rojo y negro, respectivamente, con la línea identidad.



IV.3.5 Validación estadística

Todas las cuantificaciones involucradas en el modelo ajustado presentado en las tablas IV.1 y IV.2 (intervalos de confianza y p-valores) requieren que los errores ε_i satisfagan las condiciones usuales de un modelo de regresión lineal, es decir, distribución normal cero con desviación estándar constante (homocedasticidad). En colaboración con la Dra. Mariela Sued y Dra. María Eugenia Szretter del Instituto de Cálculo - UBA se demostró, de manera rigurosa, que los datos de entrada al modelo cumplen con estas asunciones. La figura IV.3A muestra gráficas de cuantiles de los residuos, donde cada valor de los residuos corresponde a restar el dato predicho al experimental. El hecho de que los puntos caigan sobre una línea en el gráfico es una clara indicación de que la distribución normal describe de manera apropiada los residuos del modelo. Además, el ensayo de ajuste Shapiro-Wilks para testear normalidad dio un valor p de 0,99 para k_{on} y de 0,73 para k_{off}, apoyando la asunción de normalidad de los errores en los modelos utilizados.

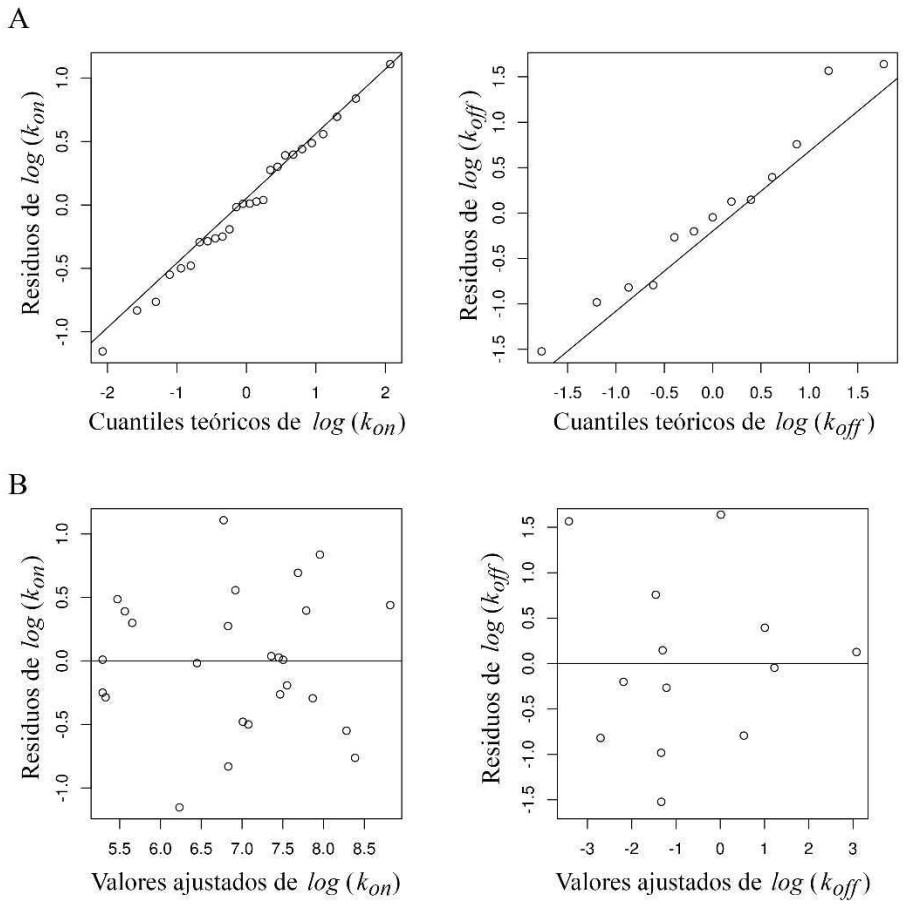
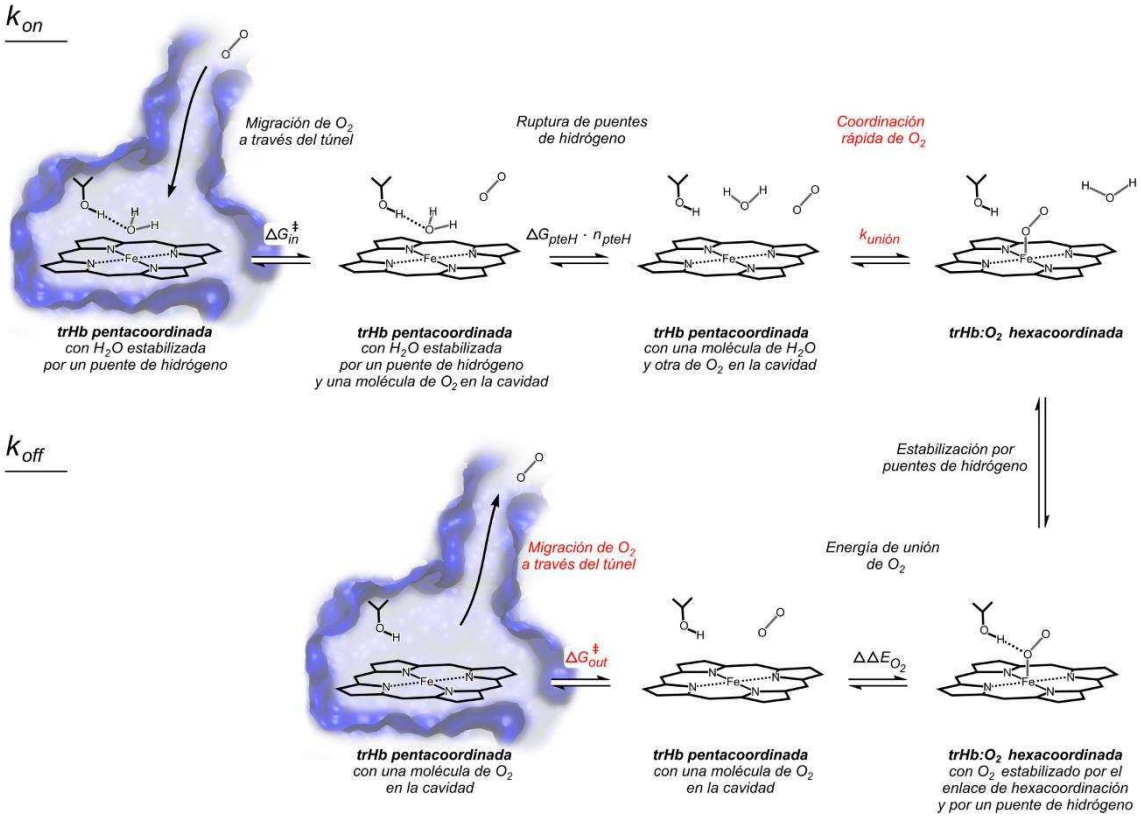


Figura IV.3. Gráficas de cuantiles de los residuos (A) y residuos versus valores ajustados (B) para el modelo 4 para $log(k_{on})$ (panel izquierdo) y para el modelo 7 para $log(k_{off})$ (panel derecho).

En la figura IV.3B se observan los residuos versus los valores ajustados. La ausencia de estructura en el gráfico garantiza la correcta asunción de homocedasticidad para los errores del modelo. Todos los análisis estadísticos se llevaron a cabo utilizando el programa R[10].

IV.4 Discusión

La capacidad predictiva de los modelos aquí presentados, junto con la evaluación estadística de los componentes individuales incluidos en cada modelo, proveen de un contexto cuantitativo que permite comprender los principales factores responsables de la captación y liberación de O_2 en la familia de hemoglobinas truncadas, como se resume en el esquema IV.1.



Esquema IV.1. Paso por paso de la captación y liberación de oxígeno en las hemoglobinas truncadas, resaltando los pasos considerados por los modelos finales. Se resaltan en rojo dos pasos no considerados en los modelos finales.

Mientras que los sistemas de túneles y cavidades internas de migración y las moléculas de agua en el sitio activo modulan la entrada del ligando, la naturaleza de los aminoácidos del sitio activo controla la cinética de liberación de O_2. Su variada capacidad para modular la captación y la liberación de ligandos se traduce en diversas funciones biológicas para estas hemoproteínas. Por ejemplo, trHbs con una captación rápida (alta k_{on}) y rápida liberación (alta k_{off}) generalmente son transportadoras de O_2 (como la hemoglobina y mioglobina); mientras que aquellas con una captación rápida (alta k_{on}) y baja liberación (baja k_{off}) normalmente muestran

una reactividad por múltiples ligandos, puesto que el oxígeno debe ingresar rápido y permanecer dentro hasta que un segundo ligando ingrese para reaccionar (por ejemplo, la trHbN de *Mycobacterium tuberculosis*). Es interesante notar que ambos procesos, captación y liberación, dependen de las interacciones en el sitio activo, donde aminoácidos con regiones de densidad de carga positiva pueden, por un lado, retrasar el ingreso de oxígeno reteniendo moléculas de H_2O y, por el otro, retrasar la liberación del O_2 coordinado. Los modelos presentados sugieren, por lo tanto, que la única manera en que una trHb puede controlar la captación sin modificar la liberación de O_2, es modificando la accesibilidad a la cavidad distal, a través de los túneles y cavidades internas de migración.

Un aspecto a considerar es que el modelo propuesto para k_{on} no considera el fenómeno de hexacoordinación interna, el cual involucra a un aminoácido del sitio activo ocupando la sexta posición de coordinación del átomo de Fe del grupo hemo. Por lo tanto, el modelo propuesto describe el proceso de captación de O_2 y estima el valor experimental de k_{on} de trHbs pentacoordinadas (con el Fe hémico en estado ferroso), o del estado pentacoordinado de trHbs hexacoordinadas. En este sentido, se encontraron dos trHbs, pertenecientes a los microorganismos *Synechococcus* y *Synechocystis*[11,12], con el hemo hexacordinado en estado ferroso[13–15]. El valor experimental de k_{on} de O_2 para la trHb de *Synechocystis* es $2{,}4 \cdot 10^8 M^{-1}s^{-1}$ [16]. Esta proteína ha sido incluida en el estudio presentado en este capítulo, y el valor predicho, utilizando un modelo pentacoordinado de esta proteína, ha sido de $2{,}0 \cdot 10^7 M^{-1}s^{-1}$. Como puede observarse, la diferencia entre ambos valores no dista del error promedio para todos los casos considerados en el estudio, lo que sugiere que para dicha trHb de *Synechocystis*, de poseer un estado hexacoordinado, la población correspondiente no parece ser significativa como para desplazar el equilibrio pentacordinado-hexacoordinado y afectar a la k_{on} de O_2.

Otro aspecto a tener en cuenta es que los valores de error (ε_i) en los modelos 4, 6 y 7, incluyen el error inherente en el modelo, así como también en las mediciones experimentales. Para ambos k_{on} (modelo 4) y k_{off} (modelos 6 y 7), el error encontrado resultó ser del mismo orden, así como la varianza entre constantes experimentales reportadas independientemente en la literatura por distintos grupos y técnicas utilizadas. Este resultado respalda la ausencia de cualquier equivocación teórica significativa en la construcción de los modelos.

Se debe tener en cuenta una consideración especial para el valor dado de R^2, el cual, aunque es un buen indicador del ajuste en un modelo lineal, éste no es suficiente para juzgar su precisión. Un claro ejemplo de este problema se encontró en el análisis presentado en este capítulo para el caso de la constante de liberación, k_{off}. Los valores de R^2 para los modelos 6 y G3 son ambos altos e indistinguibles (0.787 y 0.784, respectivamente), incluso cuando una de

las variables usadas en 6 no aporta una contribución significativa a $log(k_{off})$. Basado solamente en los valores de R^2, ambos modelos pueden ser interpretados como de una precisión equivalente. Sin embargo, con un análisis posterior de p-valores y de covarianza de variables, se ha podido concluir que el modelo reducido 7 es más preciso que el modelo 6.

Hacer extrapolaciones para el resto de los miembros de trHbs (aproximadamente 1100) es riesgoso considerando de que no hay evidencia sobre la variabilidad entre los diferentes miembros de la subfamilia. Sin embargo, puesto que el análisis presentado se realizó también sobre proteínas mutantes diseñadas para explorar amplios rangos de k_{on} y k_{off}, y considerando que esta familia de proteínas comparten un plegado común y aminoácidos relevantes, es posible predecir k_{on} y k_{off} para los otros miembros de la familia. En este sentido, se han presentado dos enfoques. El enfoque teórico usa conocimiento previo desde la fisicoquímica, y por lo tanto no está estrictamente limitado para un subconjunto de proteínas; aunque sus coeficientes no son los mejores para los datos disponibles. El enfoque basado en estadística, el cual usa los valores mejor ajustados para los datos disponibles, es altamente dependiente de este pequeño subconjunto de trHbs. De cualquier manera, ambas predicciones muestran valores muy similares dados los datos disponibles, difiriendo en aproximandamente un orden de magnitud.

Todos los resultados presentados en este capítulo están en proceso de publicación en el siguiente artículo:

JP. Bustamante, ME. Szretter, M. Sued, MA. Martí, DA. Estrin, L. Boechi. *A quantitative model for oxygen uptake and release in hemeproteins*. **Bajo revisión en Bioinformatics**.

IV.5 Conclusión

Llevando a cabo un análisis estadístico riguroso, se logró determinar los factores claves que controlan la captación y liberación de ligandos en hemoglobinas truncadas. Se encontró que el sistema de túneles y cavidades internas, así como las moléculas de agua en la cavidad distal controlan la captación del ligando, mientras que la estabilización de O_2 en el sitio distal por aminoácidos polares controlan su liberación. Se propuso, además, dos estrategias cuantitativas para predecir k_{on} y k_{off}, dentro de un orden de magnitud, para el resto de los miembros de hemoglobinas truncadas.

IV.6 Referencias

(1) Oliveira, A. S. F.; Damas, J. M.; Baptista, A. M.; Soares, C. M. *PLoS Comput. Biol.* **2014**, *10* (12).

(2) Pucci, F.; Rooman, M. *PLoS Comput. Biol.* **2014**, *10* (7).

(3) Silk, D.; Kirk, P. D. W.; Barnes, C. P.; Toni, T.; Stumpf, M. P. H. *PLoS Comput. Biol.* **2014**, *10* (6).

(4) Potapov, V.; Kaplan, J. B.; Keating, A. E. *PLoS Comput. Biol.* **2015**, *11* (2).

(5) 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. *Phys. Chem. Chem. Phys.* **2006**, *8* (48), 5611–5628.

(6) Franzen, S. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (26), 16754–16759.

(7) Markovitch, O.; Agmon, N. *J. Phys. Chem. A* **2007**, *111*, 2253–2256.

(8) Laverman, L.; Hoshino, M.; Ford, P. *J. Am. Chem. Soc.* **1997**, *119*, 12663–12664.

(9) Laverman, L.; Ford, P. *J. Am. Chem. Soc.* **2001**, *123*, 11614–11622.

(10) R Core Team. R Foundation for Statistical Computing: Vienna, Austria 2013.

(11) Scott, E.; Lecomte, J. *Protein Sci.* **2000**, *9* (3), 587–597.

(12) Scott, E.; Falzone, C.; Vuletich, D.; Zhao, J.; Bryant, D.; Lecomte, J. *Biochemistry* **2002**, *41* (22), 6902–6910.

(13) Hoy, J.; Kundu, S.; Trent III, J.; Ramaswamy, S.; Hargrove, M. *J. Biol. Chem.* **2004**, *276* (16), 16535–16542.

(14) Halder, P.; Trent III, J.; Hargrove, M. *Proteins Struct. Funct. Bioformatics* **2007**, *66*, 172–182.

(15) Couture, M.; Das, T.; Savard, P.-Y.; Ouellet, Y.; Wittenberg, J.; Wittenberg, B.; Rousseau, D.; Guertin, M. *Eur. J. Biochem.* **2000**, *267*, 4770–4780.

(16) Hvitved, A.; Trent III, J.; Premer, S.; Hargrove, M. *J. Biol. Chem.* **2001**, *276*, 34714–34721.

V. Relaciones evolutivas y funcionales en la familia de trHbs

V.1 Introducción

Determinar propiedades o aspectos funcionales específicos, como especificidad/afinidad por sustrato o la eficiencia catalítica de una proteína con precisión y detalle a nivel aminoacídico, no es una tarea sencilla. Sin embargo, si la función proteica está determinada por la estructura de la proteína y ésta por su secuencia de aminoácidos, debería ser posible, en principio, predecir la función de las mismas a partir de estructuras y secuencias proteicas de función conocida. Todo esto sería válido únicamente si se conoce la variabilidad de todas las variables relevantes en estas proteínas (las calculadas y las no calculadas). En este caso, considerando que los miembros de la familia de trHbs comparten un mismo plegado y aminoácidos críticos conservados, se espera poder predecir adecuadamente sus posibles funciones.

Debido a que la mayoría de trHbs anotadas en base de datos no poseen una función asignada y ni siquiera una estructura tridimensional conocida, en este capítulo se presenta un estudio evolutivo-molecular de toda la familia de trHbs. Además de las funciones más conocidas, almacenamiento y transporte de O_2, las hemoglobinas (Hbs) pueden tener funciones enzimáticas y de sensado[1]. En general, éstas y otras hemoproteínas con alta afinidad por O_2, tal como las trHbs micobacterianas, usualmente funcionan como enzimas redox de O_2 (y otras especies reactivas de oxígeno y nitrógeno)[2,3]. Globinas con afinidad moderada por O_2, como la mioglobina y hemoglobina de mamíferos, actúan como transportadoras y almacenadoras de O_2[4,5], mientras que las de baja afinidad por O_2, tales como la guanilato ciclasa soluble o los sensores acoplados a globinas (GCS), son sensores de NO, CO o redox[6,7].

Para llegar a una descripción funcional de cada miembro de la familia, a cada uno de ellos se le asignará una posible función en base a su capacidad para captar y liberar O_2. Para esto, se utilizarán los modelos presentados en el capítulo anterior, con una pequeña modificación sobre el modelo de captación que le otorgará un nivel de detalle mayor. Además, dado que al momento no hay ninguna caracterización de toda la familia, ya sea de las secuencias aminoacídicas o un enfoque estructural-funcional, esta asignación funcional servirá para caracterizar a cada grupo filogenético, identificando particularidades de cada uno.

Partiendo de la hipótesis de trabajo de que es posible predecir la función de trHbs partiendo solamente desde información de su secuencia, en este capítulo se presentará una combinación de métodos *in silico*. Dicha estrategia está basada en modelado por homología y cálculos energéticos, complementados con un análisis filogenético, cuyo último objetivo es predecir y analizar la función de trHbs en un contexto evolutivo. Un meta análisis de los resultados no solo permitirá profundizar sobre la comprensión de las bases estructurales que gobiernan la afinidad por ligandos, sino que también resaltarán algunas tendencias evolutivas interesantes en la función de trHbs.

Las predicciones realizadas en este capítulo están basadas en la suposición de que las trHbs están en el estado pentacoordinado en su estado ferroso. En caso de que también exista una población de trHbs ferrosas hexacoordinadas, la polación correspondiente no debería ser la mayoritaria. Si la población hexacoordinada ferrosa fuese la mayoritaria, las predicciones tendrían un error por exceso en la constante cinética de captación de O_2.

V.2 Métodos computacionales

V.2.1 Análisis filogenético basado en secuencias proteicas

V.2.1.1 Fuente de datos e identificación de secuencias de trHbs

El conjunto de secuencias de entrada estuvo compuesto por 111 casos asignados por Vuletich y Lecomte en el año 2006, a los grupos N, O y P[8], más 200 secuencias adicionales derivadas y manualmente chequeadas de las bases de datos de Pfam y PDB[9,10]. Se construyeron perfiles separados de HMMER para cada grupo de trHbs, utilizando el programa *hmmbuild*, usando las configuraciones por defecto (HMMER 3.0[11]). Con estos perfiles, y usando las configuraciones por defecto, con *hmmsearch* se exploraron las bases de datos completas de SwissProt, Uniprot y PDB para acceder a todas las posibles secuencias de trHbs disponibles. Se consideraron como trHbs todas las secuencias identificadas con un e-value menor que el umbral de exclusión establecido por HMMER. Se descartaron las secuencias redundantes con el programa CD-Hit, usando un 90% de identidad como umbral mayor[12].

V.2.1.2 Alineamientos múltiples de secuencias (AMS)

Se realizaron AMS proteicas con el programa Promals3D, con configuraciones por defecto e incluyendo información estructural considerando los siguientes diecisiete identificadores de PDBs: 2BKM, 1UX8, 3AQ5, 2BMM, 1NGK, 3AQ9, 1DLY, 1UVX, 2HZ2, 1S69, 1MWB, 1IDR, 2KSC,

2XYK, 2GKN, 1DLW, 2IG3. La inclusión de estructuras de rayos X mejora la calidad del AMS considerando propiedades claves del plegado proteico. Se optimizó manualmente el AMS usando JalView2.8[13] de acuerdo a: i) retener solo secuencias menores a 160 aminoácidos, ii) descartar secuencias sin la presencia típica de la conservada HisF8, iii) mejora manual del alineamiento. Se chequeó el AMS final con las estructuras de rayos X anteriormente citadas. Finalmente, se identificó un total de 1107 secuencias como trHbs, consistente con el trabajo de Vinogradov y otros investigadores en el año 2013[1].

V.2.1.3 Análisis filogenético

Puesto que las 1107 secuencias poseen regiones divergentes, especialmente los terminales debido a diferentes historias evolutivas, el AMS contiene bloques de subsecuencias pobremente alineadas. Estos bloques se removieron con el algoritmo y programa *Block Mapping and Gathering with Entropy* (BMGE), el cual permite seleccionar bloques del alineamiento que son adecuados para una inferencia filogenética apropiada. El recorte para la filogenia se realizó con una matriz de sustitución Blosum62, una frecuencia de gaps a 0.2 y entropía a 0.9, resultando en un recorte de 356 a 143 columnas. El AMS recortado se usó para construir filogenias por máxima verosimilitud (ML, del inglés, *maximum likelihood*) y bayesianas, usando PhyML 3.0[14] y MrBayes 3, respectivamente.

Específicamente, el análisis con PhyML se realizó luego de la selección de un modelo WAG usando ProtTest[15], con proporción estimada de sitios invariantes y un árbol de inicio de BIONJ optimizado, con medidas de soporte SH-aLRT. El análisis bayesiano se inició con árboles ML utilizando 10 perturbaciones. Se chequeó la convergencia usando *Awty* (http://www.ncbi.nlm.nih.gov/pubmed/17766271). Los árboles resultantes se observaron y editaron con iTol v2.2.2[16], GIMP e Inkscape (ambos con licencias GNU, www.gimp.org y www.inkscape.org).

Como caso control, se construyó un árbol filogenético usando las secuencias disponibles en 2006. Se encontró la misma topología reportada previamente[8] (ver figura A.V.1).

V.2.1.4 Análisis adicionales

Se llevó a cabo un análisis de posiciones determinantes de grupos y de especificidad (CDP y SDPs, por sus siglas en inglés, *Cluster and Specificity Determining Positions*), para: i) analizar si ciertos aspectos fisicoquímicos podrían ser atribuidos a aminoácidos en posiciones claves en la estructura; ii) incluir la mayor cantidad de posiciones claves inesperadas. Las CDPs son

posiciones en la estructura proteica, con columnas correspondientes en el AMS, que contribuyen de manera significativa a un grupo observado. Las SDPs son CDPs que afectan a la función. Las CDPs pueden determinarse por métodos estadísticos, y las SDPs cruzando análisis de datos obtenidos con información mutua (IM). Las CDPs se determinaron usando SDPfox[17], y se utilizó Mistic[18] para determinar IM. En la tabla A.V.1 se presentan los resultados de los análisis de SDPs e IM.

V.2.2 Modelado y simulación

V.2.2.1 Modelos por homología. Configuración de los sistemas y parámetros de simulación clásica

Se construyeron todas las variantes de trHbs armando modelos simplificados, partiendo de la estructura de rayos X según el grupo filogenético al que pertenezcan: Mt-trHbN, Mt-trHbO y Cj-trHbP (correspondientes a las entradas de PDB 1IDR, 1NGK y 2IG3). Se mutaron los aminoácidos correspondientes a los sitios activos y a las topologías de túneles, de manera *in silico*. Para todas las estructuras modeladas se siguió el protocolo presente en la sección II.4.3 del capítulo de métodos. Las simulaciones de DM realizadas fueron de 50ns para cada sistema generado.

V.2.2.2 Topología y perfiles de energía libre de migración de O_2

Los cálculos de ILS sobre los sistemas proteicos se realizaron siguiendo el protocolo presentado en la sección II.4.3. La sonda utilizada, en todos los casos, fue una molécula de O_2.

La energía libre para los perfiles de energía fue configurada en $0\ kcal\ mol^{-1}$ para el O_2 en el seno del solvente, a 9, 13 y 24 Å del Fe hémico para el portal E7, el canal corto G8 y para el canal largo, respectivamente.

V.2.2.3 Energía de unión de O_2 (ΔE_{O_2} y $\Delta\Delta E_{O_2}$)

Los cálculos de energía de unión de O_2 se computaron de acuerdo al protocolo presentado en la sección II.4.4.

V.2.2.4 Determinación de las constantes de captación y liberación de O_2 (k_{on} y k_{off})

Como se esquematizó en el esquema IV.1, el proceso de captación de ligandos pequeños involucra principalmente tres fenómenos, migración del ligando a través del túnel y cavidades internas ($\Delta G_{in}^{\ddagger}$), desplazamiento de moléculas del solvente en el sitio activo ($\Delta G_{pteH} \cdot n_{pteH}$) y coordinación al Fe ($k_{unión}$), siendo este último fenómeno no limitante de la velocidad del proceso global. Por otro lado, la ruptura de la unión Fe-O_2 ($\Delta\Delta E_{O_2}$) y la posterior migración desde la cavidad distal hacia el solvente ($\Delta E_{out}^{\ddagger}$) caracterizan a la liberación de ligandos. En el capítulo anterior se concluyó que al momento de modelar este proceso solo es necesario considerar el $\Delta\Delta E_{O_2}$, puesto que el $\Delta E_{out}^{\ddagger}$ no aporta información para mejorar la predicción. Aquí se utilizó el enfoque teórico del capítulo IV para predecir los valores de las constantes de k_{on} y k_{off}, según ecuaciones IV.8 y IV.10 ($k_{on}chem$ y $k_{off}chem$, respectivamente), puesto que este enfoque no está sesgado a un subconjunto de proteínas, como sí lo está el basado en estadística (ecuaciones IV.9 y IV.11, $k_{on}stat$ y $k_{off}stat$).

$$\mathrm{Hb} + \mathrm{O_2} \underset{kb_{\mathrm{n}}}{\overset{kf_{\mathrm{n+1}}}{\rightleftharpoons}} (\mathrm{Hb} : \mathrm{O_2})_{\mathrm{n}} \underset{kb_{\dots}}{\overset{kf_{\dots}}{\rightleftharpoons}} \cdots \underset{kb_{2\cdot 1}}{\overset{kf_{2\cdot 1}}{\rightleftharpoons}} (\mathrm{Hb} : \mathrm{O_2})_{1} \overset{\mathrm{k_{bond}}}{\rightleftharpoons} \mathrm{Hb} - \mathrm{O_2}$$

Esquema V.1. Esquema cinético genérico para el modelo propuesto de cinética de O_2. $(Hb : O_2)_n$ indica sitios de unión secundarios dentro de la proteína a lo largo del canal de salida (entrada) hacia (desde) el solvente.

En el capítulo anterior, para modelar k_{on}, se consideraron los túneles de cada trHb de manera independiente, y se seleccionó la mayor barrera del túnel más accesible como la barrera de energía libre para el ingreso de O_2 (denotada por $\Delta G_{in}^{\ddagger}$). En este caso, para calcular la barrera de energía libre para el ingreso de O_2 se consideraron los tres túneles de manera explícita, y no solo al de la barrera más accesible. Cabe aclarar que ambas estrategias (considerar solo el canal más accesible o los tres posibles) arrojaron valores muy similares. Esto muestra que ambas estrategias son igualmente válidas para las estimaciones de k_{on} que se desean realizar. Esta última estrategia parte de un modelo cinético genérico (ver esquema V.1) que considera la presencia de varios sitios de unión secundarios y sus barreras energéticas asociadas[19,20].

Asumiendo un equilibrio rápido entre el seno del solvente y los sitios de unión de la proteína, la velocidad de migración del ligando desde el solvente al sitio activo a través de un túnel dado, está dado por la siguiente relación:

$$k_t = \frac{kf_t}{kf_t + kb_t} \qquad (1)$$

donde los subíndices t corresponden a cada túnel posible (LT, E7, G8), kf y kb corresponden, de acuerdo al esquema V.1, a las velocidades con las que el ligando migra hacia el sitio activo y hacia el solvente. Usando este esquema, se calcularon las tasas de ingreso por cada túnel de todos los modelos de trHbs diseñados utilizando ILS. La lista completa de las constantes cinéticas estimadas para todas las combinaciones de aminoácidos definiendo cada túnel se presenta en la A.V.2.

Una vez obtenidas las constantes correspondientes a los tres túneles en una trHb, se combinaron para arribar a una constante de captación global, de acuerdo a:

$$k_{túneles} = k_{LT} + k_{STG8} + k_{E7G} \quad (3)$$

Por otro lado, como se explicó en el capítulo anterior, debido a que las moléculas de H_2O en el sitio activo están estabilizadas a través de puentes de hidrógeno, la energía libre para removerlas puede considerarse de manera aproximada como el producto entre la cantidad de puentes de hidrógeno (n_{pteH}) entre la molécula de H_2O y los aminoácidos del sitio activo y la energía libre de una interacción de pteH (ΔG_{pteH}). Dicho de otra forma, este efecto puede considerarse asumiendo que cada trHb existe en un equilibrio entre la molécula de agua bloqueada y el estado reactivo del grupo hemo libre, el cual se caracteriza por la constante de equilibrio K_{H_2O}. De esta manera, la expresión para calcular k_{on} pasa a ser:

$$k_{on}chem = \frac{k_B T}{h} \cdot k_{túneles} \cdot K_{H_2O} \quad (4)$$

donde $K_{H_2O} = e^{-\frac{\Delta G_{pteH} \cdot n_{pteH}}{RT}}$.

V.3 Resultados

V.3.1 Identificación de aminoácidos claves que definen la afinidad por ligandos pequeños

Ser capaces de predecir la afinidad por ligandos partiendo solamente de la secuencia proteica implica poder desarrollar una estrategia para obtener cálculos fiables de las distintas variables involucradas y volcarlas en un modelo. Dicha estrategia para computar las constantes de captación y liberación de O_2 fue desarrollada y validada estadísticamente en el capítulo IV. Pero por otro lado, si el objetivo es hacer predicciones partiendo solo de la secuencia aminoacídica, se debe conocer qué define a cada variable a nivel molecular. Esto implica poder identificar a aminoácidos claves cuya identidad esté estrechamente relacionada con la modulación de las variables de interés.

Dentro del proceso de liberación de O_2, se conoce que la variable $\Delta\Delta E_{O_2}$ está modulada por la dinámica del quinteto del sitio activo (posiciones B10, CD1, E7, E11 y G8). Dicho quinteto también es el encargado de la retención de moléculas de agua para el proceso de captación de O_2. Sin embargo, al momento de considerar las barreras energéticas impuestas por los canales de migración, una etapa crucial para el proceso de captación de ligandos, no hay una identificación directa de aminoácidos claves. En este sentido, se analizaron las simulaciones de DM llevadas a cabo para las trHbs presentadas en el capítulo IV. La superposición estructural entre las distintas proteínas permitió identificar aminoácidos claves, determinantes de las topologías de los túneles observados. Esto también posibilitó definir la cantidad de pozos y barreras energéticas para cada canal de migración. En el caso del canal largo, como fue previamente descripto para la Mt-trHbN, se observó con tres pozos y dos barreras, siendo seis las posiciones de aminoácidos claves que lo definen, estableciendo cuán alta será la barrera energética y la profundidad de los pozos: H5, B2, H9, E15, E11 y G8 (ver figura V.1). Para el caso del canal corto G8, los aminoácidos claves son H9, G8 y G9, y como muestra también la Mt-trHbN, posee dos pozos y una barrera. En cuanto al último túnel, el denominado portal E7,

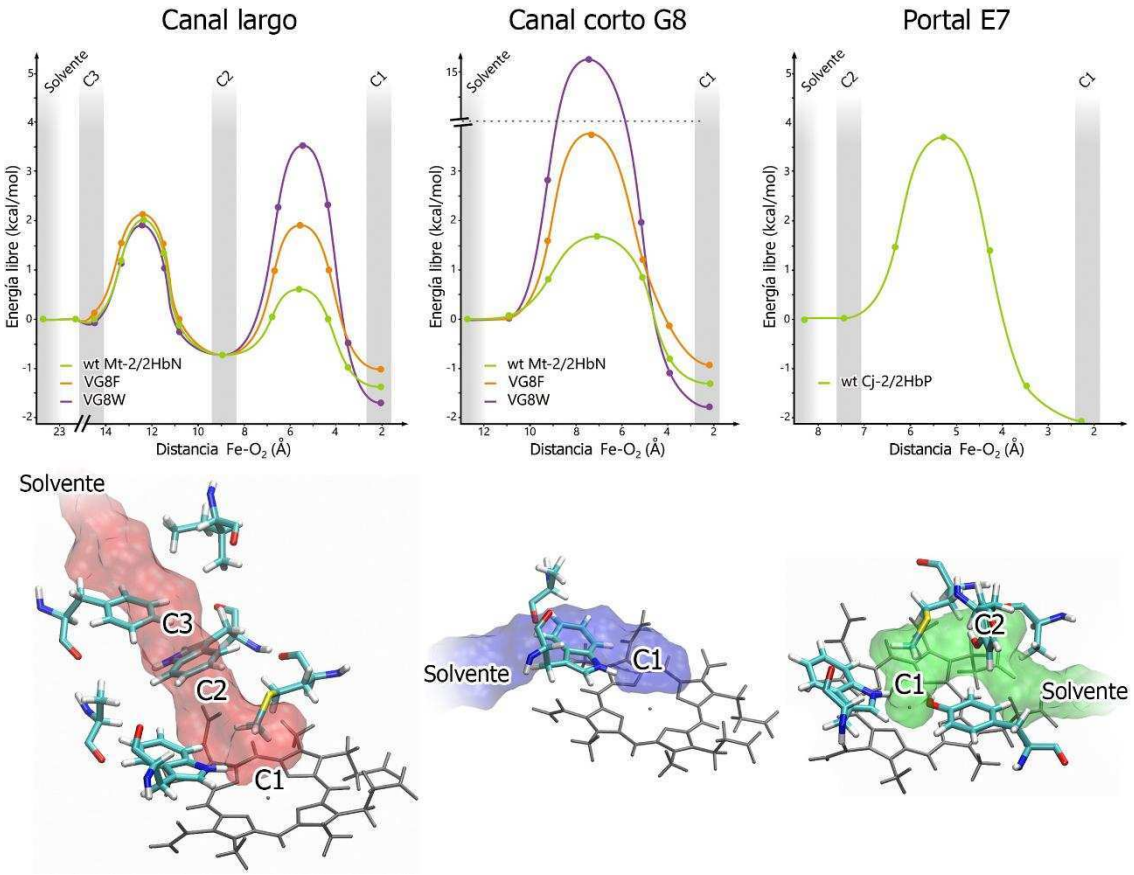


Figura V.1. Arriba: ejemplos de perfiles de energía libre para cada túnel. Abajo: representación esquemática de la combinación de aminoácidos que define tanto la topología como los perfiles de energía libre asociados a cada túnel. Pueden observarse tanto el túnel como el sistema de cavidades internas estimadas. C1, C2 y C3 corresponden a las cavidades $(trHb{:}O_2)_1$, $(trHb{:}O_2)_2$ y $(trHb{:}O_2)_3$, respectivamente.

descripto en esta tesis en Tf-trHbO está delimitado por los aminoácidos ocupando las posiciones B10, CD1, E7 y E11, y presenta dos pozos separados por una barrera. Como se muestra en la figura V.1, el tamaño, la forma y la hidrofobicidad de los aminoácidos que delimitan los túneles determinan tanto la barrera como los pozos energéticos a lo largo de los perfiles de energía libre. Generalmente, barreras en el rango de $1 - 3\ kcal\ mol^{-1}$ se consideran como muy bajas y resultan en un canal de migración completamente abierto, mientras que barreras de entre $10 - 20\ kcal\ mol^{-1}$ repercuten en un canal bloqueado o completamente cerrado.

En esta sección se ha logrado identificar a los aminoácidos constituyentes de los túneles de migración de ligandos pequeños en trHbs, lo que ha permitido hacer una caracterización molecular de la modulación de los mismos. Con esto, se ha completado la identificación de aminoácidos determinantes de las variables fisicoquímicas que permiten calcular las constantes de k_{on} y k_{off}. En la próxima sección se presentará una revisión del árbol filogenético de trHbs, junto a un estudio de las secuencias de cada grupo filogenético. La integración de todos estos resultados permitirá aplicar una estrategia de modelado que permita la inferencia de posibles funciones moleculares para los miembros de la familia de trHbs.

V.3.2 Revisión del árbol filogenético de las trHbs y análisis de composición aminoacídica

La filogenia de las trHbs ya ha sido previamente descripta a través de la conocida clasificación en N, O y P (o I, II y III), derivados de un análisis de 111 secuencias disponibles hace 10 años atrás[8], como se explicitó en el capítulo I. Muchas secuencias nuevas se han vuelto de acceso público en los últimos años, permitiendo elaborar un estudio y análisis más actualizados.

Se utilizaron perfiles construidos con HMMER para extraer de las bases de datos de Uniprot y PDB todas las secuencias correspondientes a miembros de trHbs. Se alineó utilizando Promals3D (el cual incorpora información estructural) una selección no redundante de las secuencias obtenidas. Este alineamiento múltiple de secuencias (AMS) fue utilizado para obtener un nuevo árbol de la familia de trHbs, en este caso, un árbol bayesiano que contiene 1107 secuencias (figura V.2A). Éste árbol corrobora ampliamente la clasificación previa, conteniendo al grupo N con un 24% de las secuencias, al O con un 45% y al P con un 27% de ellas. Sin embargo, la topología también sugiere la existencia de un nuevo grupo de miembros, el cual alberga un 4% de secuencias, aquí denominado grupo Q (o IV).

La distribución taxonómica a lo largo y ancho del árbol muestra un amplio rango de filos dentro de los super reinos de eucariotas, bacterias e incluso arqueas para el grupo N, mientras que P y Q contienen solo secuencias bacterianas y el grupo O la mayoría bacterianas.

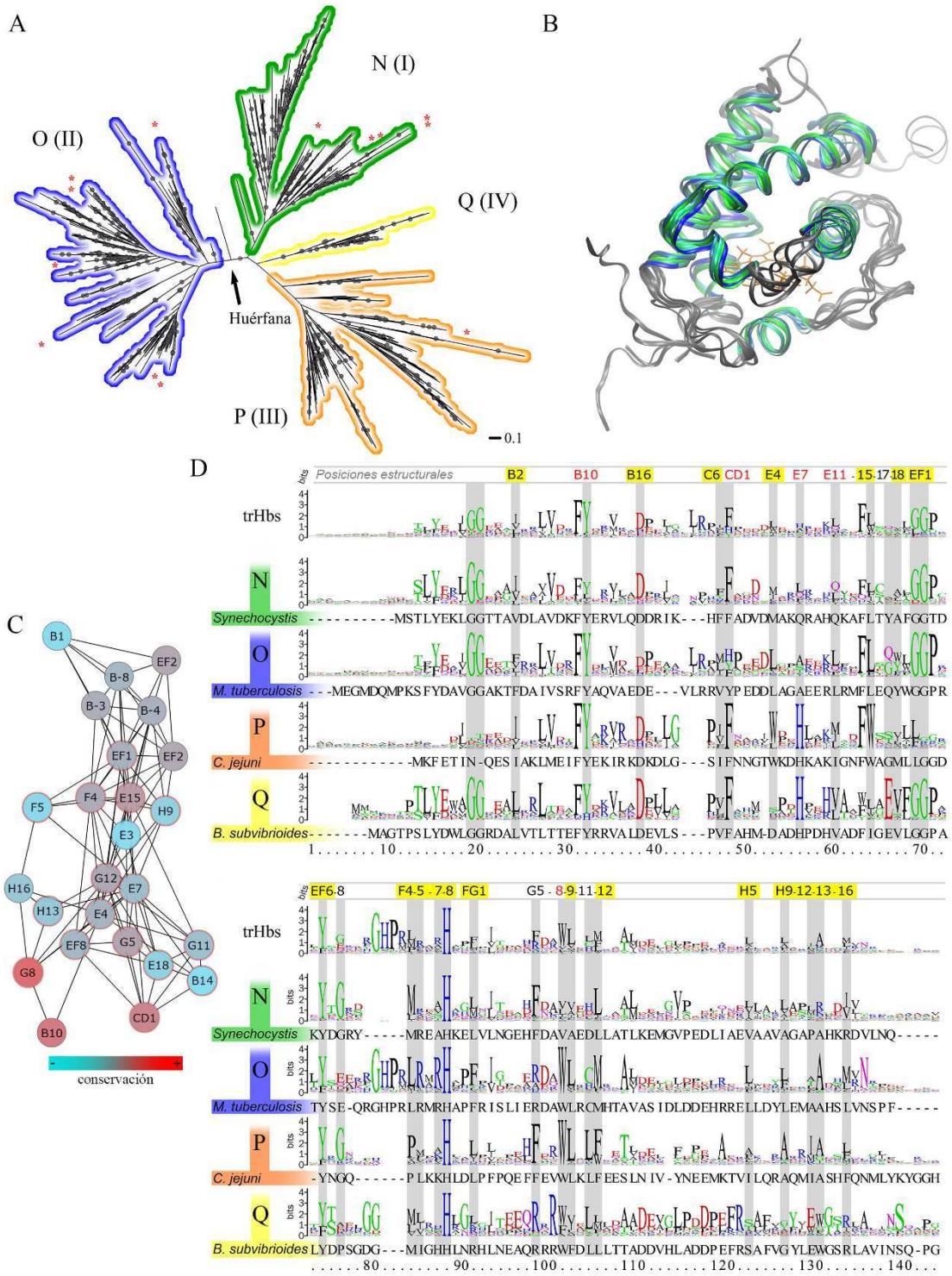


Figura V.2. (A) Árbol bayesiano de la familia de trHbs, presentado como un filograma radial. Los puntos grises indican una cobertura bayesiana mayor o igual al 80%. La barra de escala representa una distancia de 0.1 aminoácido de una mutación aceptada por sitio. Los asteriscos rojos indican secuencias correspondientes a estructuras 3D que se utilizaron para el alineamiento estructural presentado en (B). (B) Alineamiento estructural de 17 trHbs (ver métodos). Las regiones coloreadas corresponden a las regiones más conservadas en el AMS. (C) Red de información mutua (IM) de posiciones que poseen un valor de IM acumulado (IMa) > 65. El color del nodo es un indicativo de la conservación, información de Kullback-Leibler. Los nodos delimitados con rojo corresponden al SDP central, E7, y sus vecinos directos. (D) Secuencias logo recortadas con el método de BMGE, para las 1107 trHbs y para cada grupo por separado.

Se puede observar también un miembro de cada grupo para facilitar la localización de las posiciones en la secuencia. Los aminoácidos distintivos y otras posiciones importantes están sombreadas con gris, con sus respectivas posiciones estructurales indicadas arriba, en rojo para aminoácidos del sitio activo y con sombreado amarillo para las SDPs identificadas.

Aunque hay miembros de las trHbs mostrando no más de un 15% de identidad de secuencia, la estructura global de todas ellas está altamente conservada (figura V.2B). Esta conservación estructural hace que sea posible desarrollar una estrategia para llegar a una caracterización de la familia completa basada en relaciones de secuencia – estructura – función.

Se analizó la información derivada del AMS para determinar los aminoácidos claves que definen las propiedades fisicoquímicas de las trHbs y las características de cada grupo. Esto queda evidenciado por las secuencias logo, así como por los análisis de posiciones determinantes de grupos y de especificidad (CDPs y SDPs, respectivamente) y de IM (ver sección de métodos V.2.1.4 para mayor detalle). Estos resultados, presentados en las figuras V.2C, D y en la tabla S3, han generado una gran cantidad de información acerca de la relevancia de cada posición estructural. La figura V.2C, por ejemplo, muestra la red de todas las posiciones con un IM acumulado (IMa) mayor a 65, resaltando la SDP E7 y sus vecinos de IM directos (la mayoría de los cuales también son SDPs), la cual integra posiciones estructurales del sitio activo, de estabilización del grupo hemo y otras posiciones estructurales claves. De esta manera, el análisis de SDPs sugiere que los aminoácidos del sitio activo y los involucrados en la estabilización del hemo son la principal fuerza impulsora de la diversificación funcional en trHbs.

De manera concluyente, los datos muestran la estrictamente conservada HF8 que coordina al grupo hemo, cuya estabilización es reforzada por las posiciones estructurales de identidad fisicoquímica básica, correspondientes a E10, [YF]EF6 y F7, con E4, F4 y H16 construyendo el entorno hémico hidrofóbico. Hay dos motivos GG, el primero entre las regiones pre-A y hélice B, y el segundo al final de la hélice E (empezando en EF1) que están conservados en los grupos N, O y Q, pero son altamente variables en el grupo P. Un DB16 altamente conservado marca el final de la hélice B y es, como C6 para la hélice C, importante para el plegado típico de las trHbs.

En cuanto a los aminoácidos que permiten la caracterización específica de cada grupo, la posición E7, con una H conservada definiendo el portal E7 (ver más abajo), y la E15, gobernada por W en el grupo P, permiten diferenciar al grupo filogenético P de O y N, y por lo tanto, debe ser considerada como una característica distintiva de P. El grupo O muestra un tramo básico característico (H-P-R-L-R-X-R) en el loop F. Las características claves del nuevo grupo Q identificado son los anteriormente mencionados motivos GG, W o F en G8 y casi un 100% de conservación de HE7, junto a características nuevas únicas, como lo es la E estrictamente conservada en E17, un motivo R-X-R cercano a la posición G8 y un extremo N terminal con una

S altamente conservada en la posición H22. Estas peculiaridades, por lo tanto, son consideradas como características específicas y distintivas del grupo Q respecto a los demás.

Los aminoácidos del sitio activo muestran una conservación global intermedia, y también contribuyen a la separación entre subgrupos. Los más conservados son los comúnmente encontrados FB9-YB10, clave para determinar la afinidad por ligandos. G8 posee el típico W en los grupos O, P y Q. En CD1 hay predominantemente una F, pero dentro del grupo O también hay una frecuencia alta de Y o H. E7 y E11 muestran altos niveles de variación (figura V.2D).

Es interesante resaltar que hay cuatro posiciones estructurales adicionales con valores significativos de IM, E17, EF8, G5 y G11, lejos del sitio activo y de los túneles, para los cuales su función molecular es actualmente desconocida (resaltados en rojo en la figura V.2D). Posiblemente, estas posiciones estructurales estén relacionadas con estabilidad del plegado, interacciones proteína-proteína y/o modificaciones post-traduccionales.

Hay un último punto a considerar relacionado con el largo de las secuencias, puesto que muchas proteínas muestran extensos extremos C o N terminal. Los datos estructurales sugieren que, si bien éstos pueden adoptar estructuras secundarias, éstas no alteran el plegado global de la proteína.

V.3.3 Modelado de toda la familia proteica

Sobre las 1107 secuencias de trHbs identificadas, un análisis de todas las posibles combinaciones de aminoácidos en canales, usando el AMS, arrojó 460, 156 y 137 combinaciones que definen al canal largo, al corto G8 y al portal E7, respectivamente. Combinando aminoácidos en función de sus propiedades fisicoquímicas, la selección de las combinaciones más representativas resultó en 41, 36 y 17, cubriendo más del 87% de todas las trHbs.

Por otro lado, el análisis de todas las posibles combinaciones de aminoácidos en el sitio activo que interactúan con el O_2 unido, o, eventualmente, con una molécula de agua, mostró un total de 158 combinaciones diferentes de aminoácidos claves. Siguiendo la misma estrategia que para los canales, dicha cantidad ha sido reducida a 28 combinaciones, cubriendo más del 75% de la diversidad de sitios activos en trHbs.

Se construyeron modelos simplificados de la mayoría de trHbs tomando a los aminoácidos claves de canales y del sitio activo, e intercambiándolos en las estructuras de referencia de cada grupo filogenético (código de PDB 1IDR para el grupo N, 2BMM para el grupo O, 2IG3 para los grupos P y Q, ver sección II.2.2.1 para más detalles).

Se realizaron simulaciones de DM para todos los modelos, seguidas de análisis de ILS, cálculos de energías de unión para el O_2 unido y para una molécula de agua. De estas simulaciones y

posteriores análisis se obtuvieron los valores de las variables fisicoquímicas $\Delta G_{in}^{\ddagger}$, n_{pteH} y $\Delta\Delta E_{O_2}$, requeridas por los modelos matemáticos desarrollados para predecir constantes de k_{on} y k_{off} de O_2.

En la tabla A.V.2 se presentan los valores para k_{LT}, k_{STG8}, k_{E7G}, K_{H_2O}, k_{on} y k_{off} para cada combinación de aminoácidos y, por consiguiente, para cada trHb.

Como control, también se construyeron modelos simplificados para todas las trHbs con estructura de rayos X disponibles, y se compararon resultados partiendo de estos modelos y de las estructuras cristalográficas. Los resultados para los parámetros obtenidos y las constantes cinéticas predichas resultaron equivalentes, sugiriendo que la estrategia adoptada es apropiada para la inferencia de constantes cinéticas.

V.3.4 Análisis global de las constantes cinéticas y la afinidad por O_2

Un análisis sobre la distribución de valores de constantes de captación para todas las trHbs calculadas (ver figura V.3) muestra que, si bien se observa un amplio rango de valores, la mayoría de las trHbs presenta valores en el rango de 10^5 – $10^8\ M^{-1}s^{-1}$. Este rango de constantes corresponde a tener al menos un túnel completamente abierto, cuyo sitio activo está bloqueado parcialmente por una o dos moléculas de agua. También hay un grupo de proteínas con valores mayores, llegando a 10^9 – $10^{10}\ M^{-1}s^{-1}$, debido a la pérdida de moléculas de aguas retenidas. Finalmente hay un pequeño grupo con constantes bajas, 10^3 – $10^4\ M^{-1}s^{-1}$, que corresponden a aquellas proteínas cuyos túneles se encuentran mayoritariamente bloqueados y con moléculas de aguas fuertemente estabilizadas en el sitio activo.

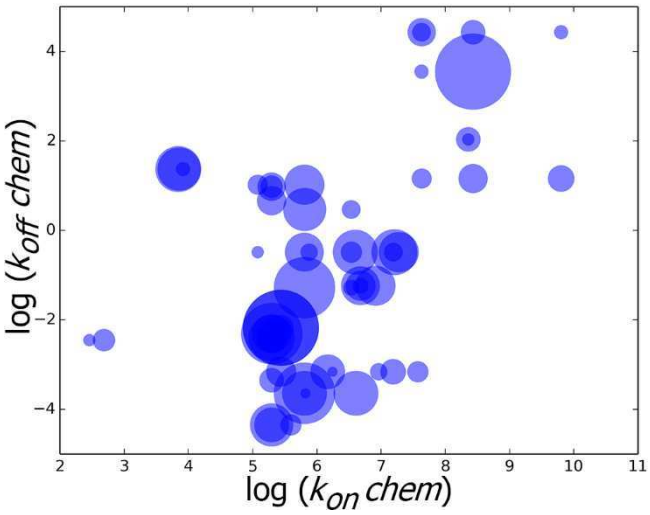


Figura V.3. Gráfica de valores de $log(k_{off}chem)$ versus $log(k_{on}chem)$. El tamaño del círculo es un indicativo de la cantidad de proteínas con el mismo valor predicho.

La distribución del logaritmo de las constantes de liberación de O_2 (ver figura V.3) resulta ser más homogénea, pero con una predominancia de valores por debajo de cero. El rango se extiende desde $10^{-4}\ s^{-1}$, que corresponde a aquellas proteínas que unen fuertemente al O_2, estabilizándolo con varios puentes de hidrógeno, hasta $10^4\ s^{-1}$ para aquellas proteínas con una cavidad altamente hidrofóbica.

La determinación de ambas constantes cinéticas ha permitido calcular la afinidad por O_2, que habitualmente se calcula como el cociente k_{on}/k_{off} y se expresa como p50, siendo ésta la presión de oxígeno a la cual la mitad de la concentración de proteína se encuentra con O_2 unido. Los resultados mostrados en la figura V.4 muestran que la mayoría de trHbs muestran valores bajos o muy bajos de p50 (< 1mmHg), sugiriendo que la proteína se encuentra oxigenada incluso en ambientes microaeróbicos. Estas proteínas generalmente muestran valores moderados de k_{on} y algunas variaciones en k_{off}, con valores inferiores a $1\ s^{-1}$. Un segundo grupo con valores moderados de p50 (1-5mmHg), posiblemente está relacionado con proteínas involucradas en transporte de O_2, el cual está caracterizado por valores moderados de k_{on} ($\sim 10^5\ M^{-1}s^{-1}$) y k_{off} ($1 - 100\ s^{-1}$). Y finalmente, hay algunos miembros con valores altos de p50 (con valores elevados de k_{on} y k_{off}), sugiriendo que son incapaces de unir oxígeno. Un análisis de correlación de p50 con k_{on} y k_{off} sugiere que la p50 está predominantemente controlada por k_{off} (R^2 = 0.60), con un pequeño impacto de k_{on} (R^2 = 0.05).

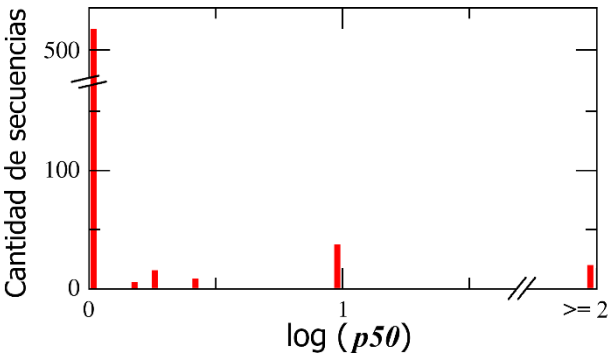


Figura V.4. Histograma de valores de p50 para todas las trHbs cuyos valores fisicoquímicos han sido calculados.

V.3.5 Combinación de resultados fisicoquímicos y filogenéticos

Se combinaron los parámetros fisicoquímicos calculados para todas las trHbs en un contexto filogenético, para analizar cómo las propiedades de unión de O_2 se relacionan con el proceso evolutivo que resulta en la diversificación funcional de las trHbs. Para comprender el patrón resultante, primero se analizará cómo la filogenia resulta en una agrupación jerárquica de secuencias de trHbs (a nivel de grupo y subgrupos) y segundo, qué propiedades se agrupan con cada clado. La figura V.5 muestra el árbol filogenético de toda la familia proteica, junto con un mapeo de la distribución de estabilización de O_2 y el grado de apertura de cada canal. También se llevaron a cabo análisis similares para los otros parámetros calculados (ver figura A.V.2). La imagen emergente no solo permite caracterizar mejor a cada grupo, sino también identificar varios subgrupos (monofiléticos y parafiléticos), que comparten propiedades claves relacionadas con sus propiedades de unión de O_2, y por ende con sus posibles funciones, como se discutirá en la sección "V.4 Discusión".

Los cladogramas muestran la topología derivada de la figura V.2A, con bloques de colores indicando la clasificación de los clados. Las propiedades fisicoquímicas se han mapeado como

círculos concéntricos, con el primero presentando valores de $\Delta\Delta E_{O_2}$ en un gradiente de color violeta (a mayor estabilización, violeta más oscuro). También se mapeó la apertura de cada uno de los tres túneles, en círculos concéntricos rojo, celeste y verde, para el canal largo, corto G8 y portal E7, respectivamente, cuanto mayor es el tamaño del círculo, más abierto se encuentra el túnel.

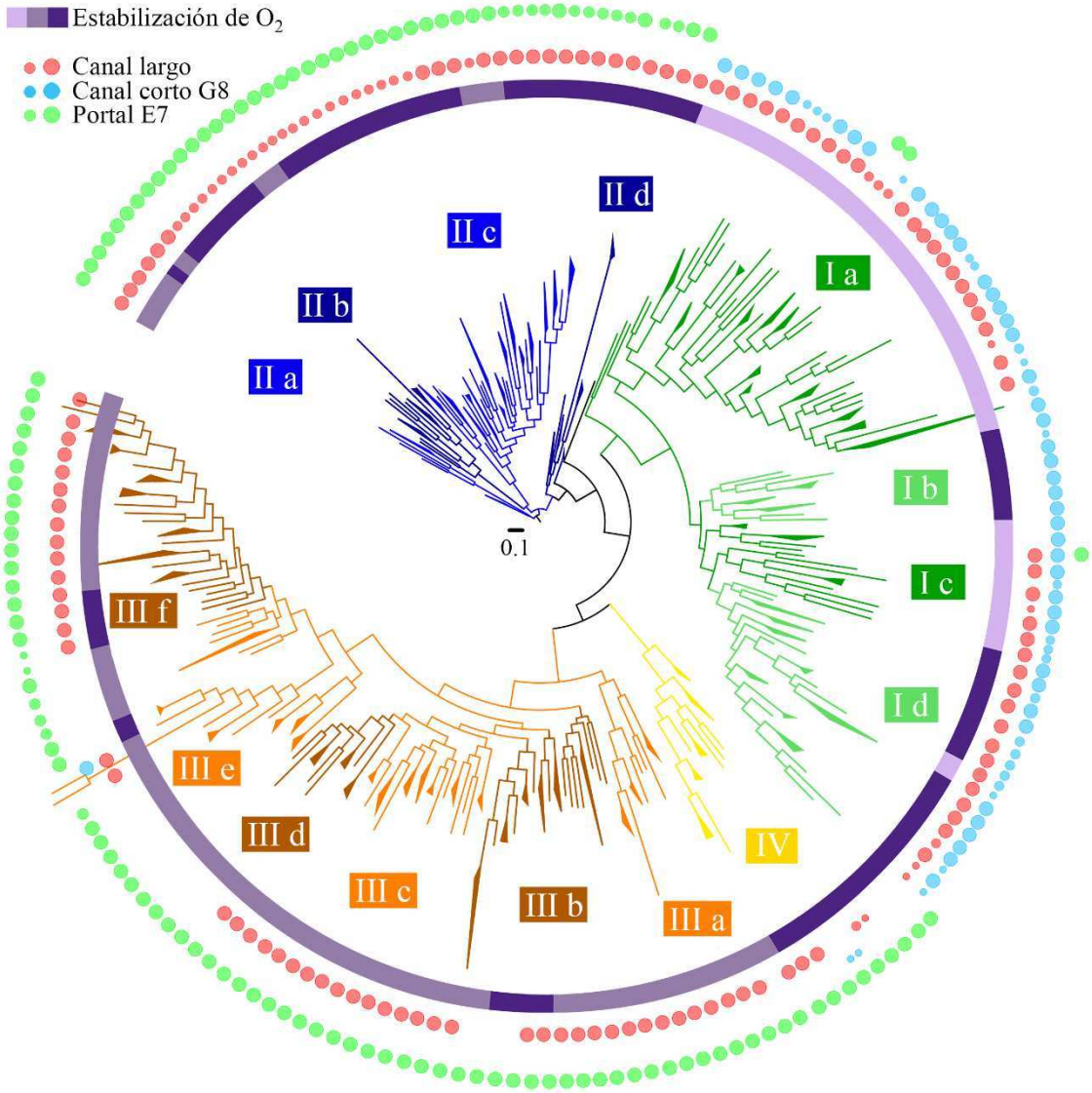


Figura V.5. Árbol filogenético colapsado de las 1107 trHbs. Se muestra como un filograma circular con características fisicoquímicas calculadas para cada trHb mapeadas sobre él, en círculos concéntricos.

Analizando esta figura, la estabilización de oxígeno (asignada como alta, moderada y baja) muestra una clara distribución de subgrupos, resaltando la presencia de aminoácidos claves que en determinados clados. Por otro lado, la apertura de túneles también muestra preferencias de grupos. El canal corto G8, por ejemplo, solo está presente (o abierto) en el grupo N, el cual parece también estar caracterizado por no presentar al portal E7 abierto (excepto para el

pequeño linaje cercano al grupo O en la figura V.5). Sin embargo, el portal E7 se muestra siempre abierto para trHbs del grupo O y P. La razón molecular del bloqueo del canal corto G8 en estos dos grupos de trHbs (incluso en el clado Ia), se debe a la presencia unívoca de WG8 y de FG8 en algunas trHbs del clado Ia.

El clado I (N) tiene la particularidad de presentar altas o bajas afinidades por O_2, pero no intermedias, causadas en mayor medida por valores muy bajos de k_{off}. La baja afinidad se correlaciona con un componente hidrofóbico en el sitio activo, una leucina en las posiciones estructurales E7, E11 y G8. Este clado puede subdividirse en los 4 subclados Ia, Ib, Ic y Id, de los cuales Ib y Id son clados monofiléticos con alta afinidad por O_2, correlacionado con Q en E7 y en E11. Una única mutación a nivel nucleotídico (CTA → CAA o CTG → CAG) podría explicar el gran cambio en k_{off}, pasando de leucina a glutamina. El subclado Ib difiere del resto de los subclados en que éste posee solo uno, y no dos, canales abiertos.

El clado II (O), con un portal E7 completamente abierto y un canal largo abierto / semi abierto como característica distintiva, puede agruparse en 4 grupos, IIa y IId siendo parafiléticos y IIb y IIc monofiléticos. Este clado presenta trHbs con moderadas y muy bajas k_{off}. La diferencia principal entre las trHbs está en su constante de liberación de O_2, junto con la identidad de la posición estructural CD1, que puede agruparse por H, F o Y.

El clado III (P), el cual también presenta un portal E7 abierto, puede dividirse en 6 subgrupos, teniendo 3 monofiléticos (a, d y f) y 3 parafiléticos (b, c y e). En general, la mayoría de los subgrupos presentan constantes cinéticas moderadas, excluyendo al subgrupo IIIb, con k_{off} moderadas pero también muy bajas.

Finalmente, el grupo IV (Q) presenta solo un grupo monofilético caracterizado por valores moderados de k_{on} y muy bajos de k_{off}.

V.4 Discusión

En este capítulo se ha presentado un análisis filogenético y estructural actualizado y detallado de las trHbs disponibles al momento, buscando razones estructurales que gobiernen las posibles funciones. Utilizando modelos simplificados de las trHbs se calcularon valores de constantes de captación y liberación de O_2, para más del 80% de las trHbs encontradas. En los próximos segmentos se analizará toda la información presentada de manera integrada, a fin de arribar a un escenario general sobre la evolución y función de trHbs.

V.4.1 Razones estructurales subyacentes de la afinidad por ligandos pequeños

El análisis global muestra que las trHbs presentan, en general, constantes de liberación de O_2 moderadas a muy bajas, y por ende afinidades moderadas a altas, debido a la presencia de al menos una, y usualmente varias interacciones de puentes de hidrógeno entre el ligando y la proteína, y cuyos ejemplos más comunes son la YB10, el WG8, H o Y en CD1 y QE11.

En cuanto a los túneles, tanto el tamaño como la hidrofobicidad de los aminoácidos que los delimitan definen la profundidad de los pozos energéticos (o cavidades) y las barreras de energía a lo largo del perfil de energía libre. Prácticamente todas las trHbs muestran al menos 1 canal abierto, y muchas de ellas, 2. Resulta importante resaltar que, en concordancia con observaciones experimentales previas sobre mutantes directas[20,21], la presencia de más de un túnel es redundante en términos de constantes de captación, puesto que el ligando accederá al grupo hemo a través del túnel con la vía más accesible. Dentro de este contexto, la pregunta sobre cuál es la relación entre la presencia de múltiples túneles y la función de trHbs, debe ir más allá de la simple determinación de k_{on}, e involucra también otros aspectos o posibilidades (ver siguiente subsección).

V.4.2 De la estructura a la función

Habiendo determinado y analizado las propiedades de unión de O_2 y sus relaciones filogenéticas, la pregunta ahora es si es posible inferir o al menos sugerir una posible función para las trHbs. El hecho de hablar de función de globinas es algo controversial, puesto que incluso para la globina estrella mioglobina, se han propuesto y validado como posibles funciones la de transportar O_2, actuar como reductasa de nitrito o como dioxigenasa de NO[19,22,23]. Esta cuestión se debe a la vasta reactividad hémica, la cual permite llevar a cabo diferentes tareas bajo diversas condiciones del entorno. Sin embargo, no todas las tareas pueden desarrollarse con la misma eficiencia debido a que la reactividad hémica es diferencial, y entonces hay algunas funciones que podrían ser más probables que otras. Más aún, como se mencionó en la sección "V.1 Introducción", la presencia ubicua de oxígeno en el ambiente y la gran variación observada en la afinidad de las trHbs por O_2 (en contraposición con CO o NO cuya afinidad es menos sensible al entorno), permiten concluir algunas tendencias generales basadas en este fenómeno clave. Aunque no hay una trHb para la cual su función haya sido indudablemente establecida, y muchas de ellas permanecen pobremente caracterizadas más allá de medidas cinéticas, para algunas de ellas (que serán las elegidas como casos modelos), se pueden hacer asignaciones

tentativas, basándose en propiedades funcionales.

Posiblemente la trHb más estudiada es la Mt-trHbN. La función ampliamente conocida de esta proteína es detoxificar NO a través de su oxidación a nitrato, una reacción mediada por el oxígeno hémico. Para lograr esta tarea se requiere una alta estabilización de O_2 y, la presencia de múltiples canales, aparentemente es un factor importante[24–28]. La mayoría de las trHbs del subgrupo Id comparten estas propiedades, por lo que la detoxificación de NO parece ser una función muy probable. De manera interesante, los otros subgrupos de N (Ia y Ic) muestran una k_{off} grande, que reduce la afinidad por O_2, haciéndola más similar a casos como la mioglobina o la Pc-trHbN[19,29]. Para estas proteínas, así como para otras trHbs compartiendo grandes valores de k_{off} y la presencia de uno o dos túneles abiertos, un rol involucrando almacenamiento y transporte de O_2 parece lo más probable, puesto que una k_{on} moderada y una k_{off} grande es un prerrequisito para permitir una captación y liberación de O_2 eficientes.

Otro caso está representado por la Tf-trHbO o la Mt-trHbO, ambas cuyas funciones parecen estar relacionadas con especies reactivas de O_2, probablemente como catalasas-peroxidasas[30,31]. Para actuar como catalasas-peroxidasas, las propiedades claves que poseen estas proteínas son: la presencia de una red de puentes de hidrógeno fuertes, revelada por valores bajos o muy bajos de k_{off}, y el portal E7 abierto. La disponibilidad del portal E7 es el acceso al Fe hémico más corto y más polar, y puede entonces ser particularmente adecuado para la entrada y salida de ligandos cargados o polares tales como el superóxido. Otra característica importante es que habitualmente las hemoproteínas que realizan estas tareas disponen de aminoácidos polares y aromáticos (Trp-Tyr-Arg-His) en sus sitios activos que pueden participar en reacciones redox estabilizando especies de radicales libres. Estas funciones emergen como probables candidatas para muchas (casi la mayoría) de trHbs del grupo O, las cuales comparten las propiedades mencionadas.

En cuanto a los miembros del grupo P, es menos lo que se conoce al respecto, y el miembro mejor caracterizado es la trHbP de *Campylobacter jejuni* (Cj-trHbP). Aunque su función no está clara, muestra propiedades estructurales y de unión de ligandos que revelan una fuerte red de puentes de hidrógeno y la presencia del portal E7 abierto (como en el caso descripto de las trHbs O). Estas propiedades, sin embargo, no son compartidas por todos los miembros del grupo, y se observa una alta variabilidad en término de interacciones con el ligando, impidiendo extraer una inferencia general acerca de sus posibles funciones. Aquí, la razón por la cual las trHbs del grupo P forman un grupo distinto, puede explicarse por la estrictamente conservada HE7.

V.4.3 Distribución funcional basada en organismos

Habiendo hallado para la mayoría de las trHbs sus grupos filogenéticos de pertenencia, junto con sus características funcionales claves, aquí se analizará cómo se distribuyen las trHbs en los organismos a los que pertenecen. Las secuencias de 1107 trHbs provienen de más de 600 especies diferentes, con la mayoría de ellas (73%) conteniendo solo un tipo de trHb, 23% teniendo dos trHbs de diferentes grupos y algunos organismos conteniendo más de dos. Análisis de la filogenia mostraron que, para aquellos organismos que albergan dos tipos de trHbs, casi la mitad de ellas son tipos O y N, cerca de un 40% O y P, y solo un 15% de N y P.

Desde un punto de vista funcional, y aunque se puede hacer un análisis de co-presencia del organismo y distribución de cada propiedad calculada, ninguna de ellas analizada de manera independiente ha mostrado una tendencia clara, y parece solo reflejar la distribución de la propiedad entre las trHbs. Sin embargo, y basado en la descripción previa, se puede hacer una asignación funcional definida por: una química de multiligandos NO/O_2 (tipo Mt-trHbN), almacenamiento y transporte de O_2 (tipo mioglobina) y catalasa-peroxidasa (tipo Tf-trHbN o Mt-trHbO). Los organismos que contienen solo una trHb predominantemente poseen una del tipo funcional catalasa-peroxidasa (64%), seguidas de almacenamiento y transporte de O_2 y química de multiligandos NO/O_2 (18% cada una), lo cual es la distribución esperada basada en la abundancia relativa de cada tipo funcional. Aquellos organismos con dos tipos de trHbs también reflejan las distribuciones esperadas. Así, los datos disponibles no muestran evidencias claras de diversificación funcional para las trHbs coexistentes en un mismo organismo.

En lo que respecta a las relaciones filogenéticas – funcionales, la función propuesta de catalasa-peroxidasa emerge como el candidato más probable para la mayoría (86%) del grupo O, todas compartiendo las propiedades mencionadas. Para el grupo N, de manera interesante, mientras que el 57% se infiere como de tipo química de multiligandos NO/O_2, el 43% restante se predice involucrado en transporte y almacenamiento de O_2. En cuanto al grupo P, la mayoría (78%) comparten propiedades relacionadas con el tipo catalasa-peroxidasa, mientras que el otro 22% se ha inferido como transportadoras de O_2, debido a sus altas constantes de liberación de O_2. Finalmente, todos los miembros del grupo Q han sido asignados como catalasas-peroxidasas. Es interesante resaltar que se encontraron tipos funcionales diferentes dentro del mismo grupo filogenético, y por lo tanto, se debe tener cuidado en asignar propiedades funcionales partiendo solo del análisis de la filogenia resultante.

Todos los resultados presentados en este capítulo están en proceso de publicación en el siguiente artículo:

JP. Bustamante, L. Boechi, L. Radusky, DA. Estrin, A. ten Have, MA. Martí. *Evolutionary and Functional relationships in the truncated hemoglobin family*. **Bajo revisión en PLoS Comp. Biol**.

V.5 Conclusiones

Los resultados aquí presentados muestran que el plegado de las trHbs parece ser lo suficientemente flexible como para permitir cambios importantes en la afinidad por pequeños ligandos. Habiendo determinado y analizado las propiedades de unión de O_2 y sus relaciones filogenéticas, se han inferido posibles funciones a más del 80% de las 1107 trHbs identificadas, partiendo de casos modelos, que pueden variar entre almacenar y transportar O_2, funcionar como catalasas-peroxidasas o intervenir en reacciones que involucran multiligandos. Esta asignación funcional ha evidenciado que dentro de un mismo grupo filogenético puede haber diferentes funciones posibles, por lo que no es posible hacer asignaciones funcionales partiendo solo de análisis evolutivos. Por el contrario, estos estudios demuestran que es necesario recurrir a un enfoque integrador, donde la sinergia entre la fisicoquímica, la bioinformática y la filogenia pueden dar respuestas a interrogantes de este tipo.

V.6 Referencias

(1) Vinogradov, S. N.; Tinajero-Trejo, M.; Poole, R. K.; Hoogewijs, D. *BBA - Proteins Proteomics* **2013**, *1834* (9), 1789–1800.

(2) Milani, M.; Pesce, A.; Ouellet, Y.; Ascenzi, P.; Guertin, M.; Bolognesi, M. *EMBO J.* **2001**, *20* (15), 3902–3909.

(3) Ouellet, H.; Ouellet, Y.; Richard, C.; Labarre, M.; Wittenberg, B.; Wittenberg, J.; Guertin, M. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (9), 5902–5907.

(4) Kendrew, J. C.; Dickerson, R. E.; Strandberg, B. E.; Hart, R. G.; Davies, D. R.; Phillips, D. C.; Shore, V. C. *Nature* **1960**, *185* (4711), 422–427.

(5) Muirhead, H.; Perutz, M. F. *Nature* **1963**, *199* (4894), 633–638.

(6) Ignarro, L. *Semin Hematol* **1989**, *26* (1), 63–76.

(7) Hou, S.; Freitas, T. A. K.; Larsen, R.; Piatibratov, M.; Sivozhelezov, V.; Yamamoto, A.; Meleshkevitch, E.; Zimmer, M.; Ordal, G.; Alam, M. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (16), 9353–9358.

(8) Vuletich, D. a; Lecomte, J. T. J. *J. Mol. Evol.* **2006**, *62* (2), 196–210.

(9) Bateman, A.; Birney, E.; Durbin, R.; Eddy, S.; Howe, K.; Sonnhammer, E. *Nucleic Acids Res.* **2000**, *28* (1), 263–266.

(10) Bernstein, F.; Koetzle, T.; Williams, G.; Meyer, E.; Brice, M.; Rodgers, J.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112* (3), 535–542.

(11) Eddy, S. *PLoS Comput. Biol.* **2011**, *7* (10).

(12) Huang, Y.; Niu, B.; Gao, Y.; Fu, L.; Li, W. *Bioinformatics* **2010**, *26* (5), 680–682.

(13) Waterhouse, A. M.; Procter, J. B.; Martin, D. M. A.; Clamp, M.; Barton, G. J. *Bioinformatics* **2009**, *25* (9), 1189–1191.

(14) Guindon, S.; Dufayard, J.-F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O. *Syst. Biol.* **2010**, *59* (3), 307–321.

(15) Abascal, F.; Zardoya, R.; Posada, D. *Bioinformatics* **2005**, *21* (9), 2104–2105.

(16) Letunic, I.; Bork, P. *Nucleic Acids Res.* **2011**.

(17) Mazin, P.; Gelfand, M.; Mironov, A.; Rakhmaninova, A.; Rubinov, A.; Russell, R.; Kalinina, O. *Algorithms Mol. Biol.* **2010**, *5* (29).

(18) Simonetti, F.; Teppa, E.; Chernomoretz, A.; Nielsen, M.; Marino Buslje, C. *Nucleic Acids Res.* **2013**.

(19) Olson, J.; Phillips, G. *J. Biol. Inorg. Chem.* **1997**, *2*, 544–552.

(20) Scott, E. E.; Gibson, Q. H.; Olson, J. S. *J. Biol. Chem.* **2001**, *276*, 5177–5188.

(21) Salter, M. D.; Blouin, G. C.; Soman, J.; Singleton, E. W.; Dewilde, S.; Moens, L.; Pesce, A.; Nardini, M.; Bolognesi, M.; Olson, J. S. *J. Biol. Chem.* **2012**, *64*.

(22) Gardner, P. *J. Inorg. Biochem.* **2005**, *99*, 247–266.

(23) Kamga, C.; Krishnamurthy, S.; Shiva, S. *Nitric Oxide* **2012**, *26* (4), 251–258.

(24) Bidon-Chanal, A.; Martí, M. A.; Estrin, D. A.; Luque, F. J. *J. Am. Chem. Soc.* **2007**, *129* (21), 6782–6788.

(25) Bidon-Chanal, A.; Martı, M. A.; Crespo, A.; Milani, M.; Orozco, M.; Bolognesi, M.; Luque, F. J. *Proteins* **2007**, *464*, 457–464.

(26) Crespo, A.; Martí, M. A.; Kalko, S. G.; Morreale, A.; Orozco, M.; Gelpi, J. L.; Luque, F. J.; Estrin, D. A. *J. Am. Chem. Soc.* **2005**, *127* (12), 4433–4444.

(27) Lama, A.; Pawaria, S.; Bidon-Chanal, A.; Anand, A.; Gelpi, J. L.; Arya, S.; Marti, M.; Estrin, D. A.; Luque, F. J.; Dikshit, K. L. *J. Biol. Chem.* **2009**, *284* (21), 14457–14468.

(28) Ouellet, Y. H.; Daigle, R.; Lagüe, P.; Dantsker, D.; Milani, M.; Bolognesi, M.; Friedman, J. M.; Guertin, M. *J. Biol. Chem.* **2008**, *283* (40), 27270–27278.

(29) Das, T.; Weber, R.; Dewilde, S.; Wittenberg, J.; Wittenberg, B.; Yamauchi, K.; Van Hauwaert, M.; Moens, L.; Rousseau, D. *Biochemistry* **2000**, *39*, 14330–14340.

(30) Ouellet, H.; Ranguelova, K.; Labarre, M.; Wittenberg, J. B.; Wittenberg, B. A.; Magliozzo, R. S.; Guertin, M. *J. Biol. Chem.* **2007**, *282* (10), 7491–7503.

(31) Torge, R.; Comandini, A.; Catacchio, B.; Bonamore, A.; Botta, B.; Boffi, A. *J. Mol. Catal. B Enzym.* **2009**, *61* (3-4), 303–308.

VI. Bases moleculares de la adaptación a altas temperaturas en la trHbO de *T. fusca*

VI.1 Introducción

Uno de los modelos representativos de trHbs ampliamente estudiados en esta tesis ha sido la Tf-trHbO, una proteína termoestable perteneciente a una actinobacteria termófila que habita en zonas de entre 55-60°C[1]. Dado este escenario, sería interesante evaluar cómo se comporta esta proteína a la temperatura fisiológica del organismo, o mejor dicho, a la temperatura a la cual esta proteína podría llevar a cabo su función. Es por esto que aquí se presentará un primer estudio sobre el comportamiento de esta proteína a altas temperaturas.

Entre los miembros de las trHb existen homólogos pertenecientes a organismos adaptados a distintas temperaturas que comparten una alta identidad de secuencia y un plegado de su estructura proteica muy similar. Esto hace que la adaptación diferencial a altas temperaturas no sea fácil de explicar a nivel molecular. En tal sentido, un análisis estructural comparativo entre la Tf-trHbO con algunas otras trHbs (presentado en el capítulo V), ha revelado que el plegado global de sus estructuras está altamente conservado. En particular, con la Mt-trHbO poseen una identidad de secuencia del 58% y, además, comparten muchos de los aminoácidos claves, tanto de la composición de los canales como del sitio activo. Sin embargo, poseen temperaturas de melting (T_m) muy diferentes entre sí, 57°C grados para Mt-trHbO y 75°C grados para Tf-trHbO[1]. Las similitudes estructurales y la notable diferencia en el valor de temperatura a la cual dichas proteínas pierden su plegado nativo, han propulsado el estudio de la estabilidad térmica diferencial de la Tf-trHbO.

Se ha propuesto que la termoestabilidad está estrictamente relacionada con la flexibilidad de una proteína[3,4]. Asumiendo que tanto enzimas homólogas termoestables como no termoestables tienen la misma función a diferentes temperaturas, y que la función proteica está relacionada con cierto grado de flexibilidad, la proteína termoestable debería tener menos flexibilidad a temperatura ambiente. Un incremento en la rigidez del estado plegado a temperatura ambiente podría estar relacionado con una estabilidad térmica mayor. La

disminución de la flexibilidad podría también explicar por qué muchas de las proteínas termoestables e hipertermoestables son inactivas a bajas temperaturas. Aunque muchos enfoques teóricos y experimentales han apoyado la hipótesis de la presencia de un entorno más rígido en proteínas termoestables a temperatura ambiente[3–7], otros resultados provenientes de experimentos de intercambios de amidas se han opuesto a esta hipótesis[8], abriendo así la discusión acerca de la relación entre flexibilidad y estabilidad térmica. En particular, se ha encontrado que en algunos casos la cantidad de puentes de hidrógeno y salinos correlaciona con la estabilidad térmica[9].

De acuerdo a la teoría del plegado a través del paisaje energético, la T_m está relacionada con la diferencia de energía entre la energía libre o efectiva (promediada sobre todos los grados de libertad del solvente) del ensamble de conformaciones del estado plegado versus el ensamble de conformaciones del estado desplegado[10]. Por lo tanto, cuanto mayor sea la cantidad de contactos hidrofílicos e hidrofóbicos en el estado plegado, en contraposición con aquellos observados para el estado desplegado, más se favorecerá a una T_m elevada.

Dentro del mismo marco teórico, puesto que el espacio configuracional de las proteínas es enorme en el estado desplegado y muy pequeño en el plegado, las proteínas deben perder entropía configuracional para alcanzar el estado nativo. Cualquier proceso que ayude a disminuir la entropía configuracional del sistema desplegado, también disminuirá la brecha de energía libre entre los ensambles plegado y desplegado, incrementando la T_m[11]. Considerando que cada aminoácido es capaz de explorar diferentes espacios configuracionales dependiendo de su identidad química, se ha propuesto que la presencia de aminoácidos más rígidos, como la prolina, son responsables de un incremento de la termoestabilidad gracias a que reducen la entropía configuracional[11]. Sin embargo, esta hipótesis también se encuentra altamente discutida en la literatura[12].

Como el lector habrá podido observar, se han propuesto muchas hipótesis, pero ninguna de ellas resulta ser conclusiva o absoluta para explicar la termoestabilidad desde un punto de vista racional.

Para comprender el mecanismo molecular a través del cual estas proteínas poseen distinta estabilidad térmica, se realizaron simulaciones de DM a diferentes temperaturas, a 27 y 87°C, así como mediciones experimentales de perfiles de fusión térmica (donde se evalúa el desplegado proteico en función de la temperatura) de ambas proteínas wt y de dos mutantes diseñadas para intercambiar un aminoácido encontrado por el estudio teórico como fuertemente implicado en la estabilidad térmica de la Tf-trHbO. Las mediciones experimentales han sido llevadas a cabo por el grupo de investigación liderado por el Prof. Alberto Boffi (Universitá La Sapienza, Roma, Italia).

VI.2 Métodos computacionales

VI.2.1 Configuración de los sistemas y parámetros de simulación

Las estructuras de inicio corresponden a las estructuras obtenidas por cristalografía de rayos X de la Tf-trHbO (código de PDB 2BMM[1]) y de la Mt-trHbO (código de PDB 1NGK[13]). Tanto la protonación de los aminoácidos, como la construcción de la caja de aguas y el resto de las condiciones de los sistemas se establecieron según el protocolo presentado en la sección II.4.3 del capítulo Métodos. Siguiendo el mismo protocolo, se construyeron las mutantes *in silico* ProE3Gly de la Tf-trHbO y la GlyE3Pro de la Mt-trHbO. El tiempo de simulación de cada una de las DM de producción fue de 100ns para todos los sistemas, que se simularon a 27 y a 87°C.

VI.3 Resultados y discusión

VI.3.1 Dinámica de las proteínas frente al cambio de temperatura

Para poder estudiar y analizar las diferencias en la estabilidad térmica entre la Tf-trHbO y la Mt-trHbO, se simularon 100ns de DM para cada proteína a temperatura ambiente (27°C) y a altas temperaturas (87°C). Los análisis de RMSD y RMSF, así como un monitoreo exhaustivo de las simulaciones, no revelaron una diferencia significativa de movimientos estructurales entre ambas proteínas, termoestable y no termoestable a temperatura ambiente. El valor de ΔRMSF promedio entre las dos simulaciones fue de 0.08 ± 0.22 Å, lo que invalida, al menos con estos resultados, la hipótesis de que las estructuras termoestables son menos flexibles que sus contrapartes mesoestables a temperatura ambiente.

El análisis de RMSD confirma que la Tf-trHbO conserva su estructura global durante toda la simulación, incluso a temperaturas elevadas (ver tabla VI.1).

Además, estos resultados muestran que la Tf-trHbO wt concentra un alto grado de fluctuaciones en el loop CD (figuras VI.1A y VI.2A). Por otro lado, la Mt-trHbO wt presenta un loop CD muy rígido y concentra muchas fluctuaciones en ambos extremos de la estructura proteica a altas temperaturas (figuras VI.1B y VI.2B). El aumento significativo en la movilidad de los aminoácidos terminales de la Mt-trHbO sugieren que ésta se encuentra en el comienzo de un proceso de desplegado.

Tabla VI.1. Valores de RMSD promedio y desvío estándar (Å) para las proteínas wt y mutantes, es decir, ProE3Gly de la Tf-trHbO y GlyE3Pro de la Mt-trHbO a diferentes temperaturas, 27°C y 87°C. Los valores de RMSD fueron calculados considerando la proteína completa, sin los extremos, solo considerando los extremos, o teniendo en cuenta extremos y el loop CD. Los extremos fueron considerados como los primeros y últimos 12 aminoácidos.

		Tf-trHbO		Mt-trHbO	
		27°C	87°C	27°C	87°C
Wild type	Completa	1.10 ± 0.13	1.98 ± 0.33	1.08 ± 0.14	3.03 ± 0.62
	Sin extremos	1.00 ± 0.11	1.48 ± 0.24	0.87 ± 0.14	1.35 ± 0.23
	Solo extremos	1.25 ± 0.19	2.18 ± 0.44	1.37 ± 0.24	5.21 ± 1.30
	Extremos y loop CD	1.31 ± 0.20	2.58 ± 0.47	1.35 ± 0.22	4.87 ± 1.17
Mutante	Completa	1.21 ± 0.24	1.91 ± 0.41	1.51 ± 0.32	2.03 ± 0.37
	Sin extremos	1.08 ± 0.23	1.35 ± 0.18	1.15 ± 0.14	1.46 ± 0.22
	Solo extremos	1.37 ± 0.30	2.65 ± 0.90	1.63 ± 0.31	2.84 ± 0.76
	Extremos y loop CD	1.40 ± 0.31	2.46 ± 0.75	1.59 ± 0.27	2.69 ± 0.65

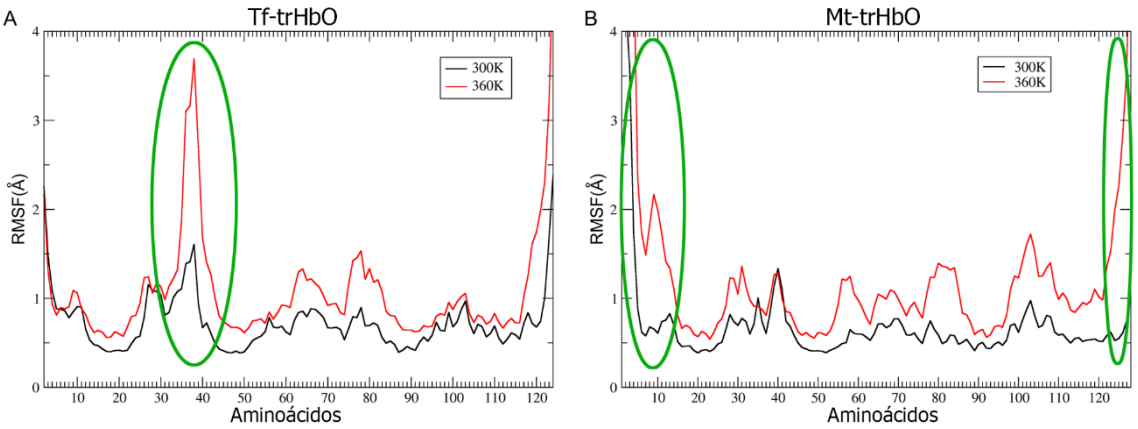


Figura VI.1. Análisis de RMSF para las proteínas Tf-trHbO (A) y Mt-trHbO (B) wt a 27°C (en negro) y 87°C (en rojo). Las diferencias significativas están resaltadas en verde. Las regiones entre los aminoácidos 35 a 42 en la Tf-trHbO (A), y 37 a 44 en la Mt-trHbO (B), corresponden al loop CD.

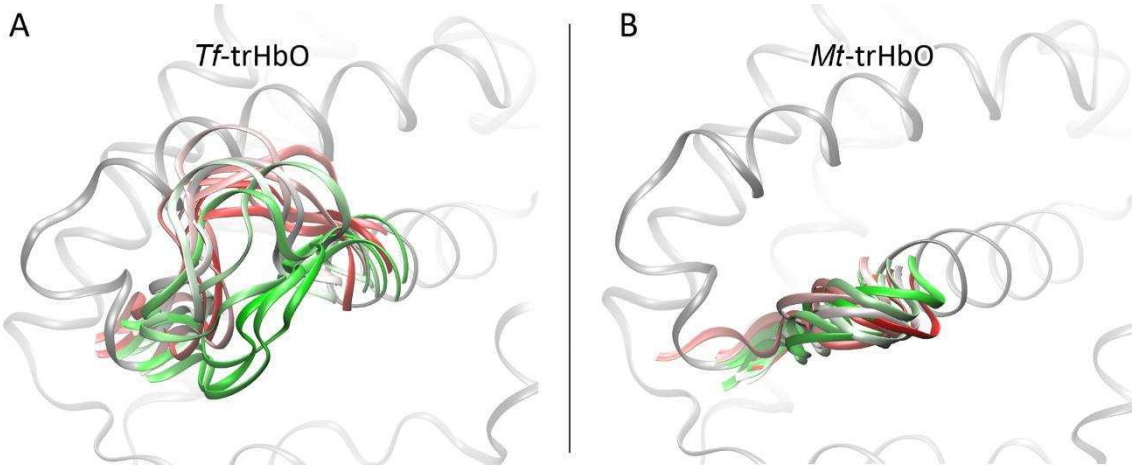


Figura VI.2. Conformaciones del loop CD durante las simulaciones de DM para las proteínas Tf-trHbO (A) y Mt-trHbO (B) wt a 87°C. Distintas fotos de la dinámica molecular han sido coloreadas de acuerdo al paso del tiempo, en una escala de color de verde-blanco-rojo.

La distorsión del loop CD observada en la Tf-trHbO a 87°C, está correlacionada con una incipiente ruptura de la hélice α E, contigua al loop CD (figura VI.3). Como es ampliamente conocido, debido a restricciones estéricas y de su cadena lateral, la prolina es incapaz de satisfacer los requisitos de ángulos Ramachandran para la formación de hélices α. En este sentido, por un lado, su cadena lateral está comprimida en el espacio del esqueleto de la hélice α y, por otro lado, el grupo metileno está en el espacio que normalmente es ocupado por un protón perteneciente a una amida que es capaz de formar un puente de hidrógeno. De esta

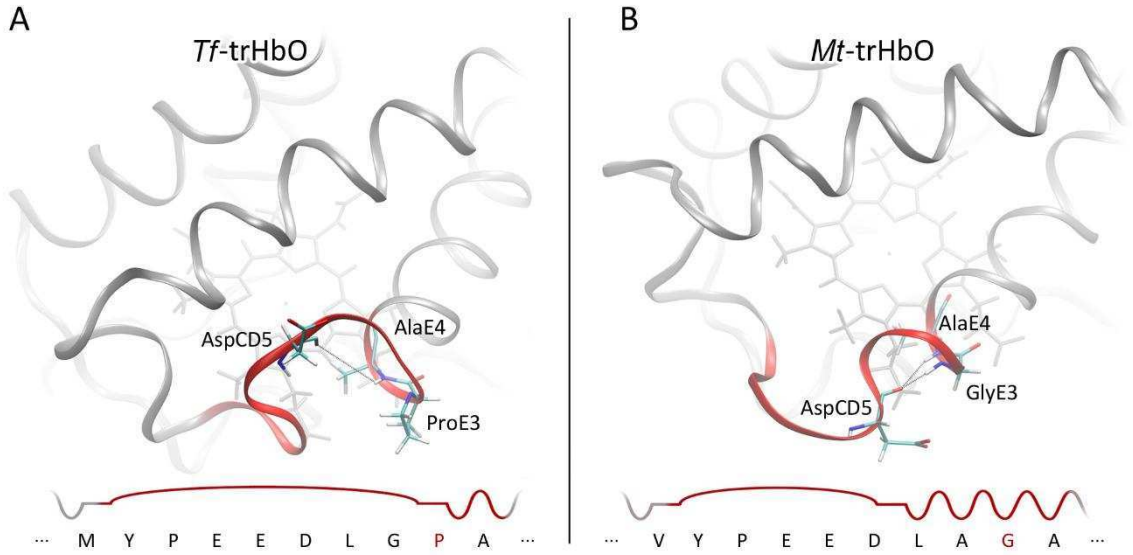


Figura VI.3. Representación esquemática de la ruptura y formación de la hélice α E (en rojo) para Tf-trHbO (A) y Mt-trHbO (B) wt, respectivamente, a 87°C. La interacción AspCD5 - AlaE4 en Mt-trHbO reduce el largo del loop CD y extiende la hélice α E. Debajo se encuentra la secuencia de aminoácidos acompañada por la estructura secundaria para el loop CD y el comienzo de la hélice α E. Los aminoácidos prolina y glicina en la posición E3 se encuentran resaltados en rojo.

manera, la prolina termina irrumpiendo en la red de puentes de hidrógeno que estabilizan a la hélice. Como consecuencia, la PE3 parece ser la responsable de ambas, desestabilizar la formación temprana de la hélice α E rompiendo el puente de hidrógeno entre AspCD5 y AlaE4 (figura VI.3), e incrementar la flexibilidad del loop CD (figura VI.2). La Mt-trHbO posee una glicina en E3 (GlyE3) que no interfiere en el puente de hidrógeno entre AspCD5 y AlaE4, favoreciendo así a la estabilización de la hélice E.

Es importante resaltar que ambas proteínas mantienen una estructura muy estable, similar a las encontradas en las estructuras cristalinas cuando las simulaciones fueron realizadas a 27°C. Ambos loops CD y hélices α se mantuvieron conservados.

VI.3.2 Comportamiento de proteínas mutantes *in silico*

Teniendo en cuenta que el aminoácido en la posición E3 parece ser el responsable de la gran flexibilidad del loop CD a altas temperaturas, y que por consiguiente surge la hipótesis de que estaría involucrado con la estabilidad térmica de la proteína, se realizaron simulaciones de DM de dos mutantes intercambiadas: una mutante de la Tf-trHbO en la cual la ProE3 fue reemplazada por Gly (ProE3Gly de la Tf-trHbO), y una mutante de la Mt-trHbO en la cual la GlyE3 fue sustituida por Pro (GlyE3Pro de la Mt-trHbO).

Los resultados muestran que la presencia de la GlyE3 en la Tf-trHbO no solo restringe significativamente la flexibilidad del loop CD, sino que incrementa la movilidad del extremo amino (figura VI.4A) y, a su vez, dispone un reacomodamiento estructural local razonable para la formación temprana de la hélice α E. Por otro lado, la presencia de la ProE3 en Mt-trHbO disminuye el grado de flexibilidad del extremo C terminal e incrementa ligeramente el movimiento de su loop CD (figura VI.4B).

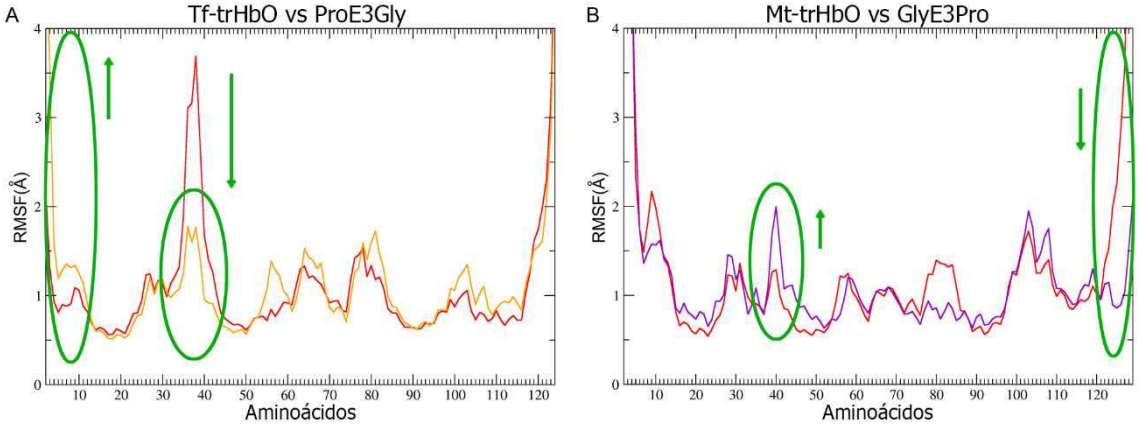


Figura VI.4. Análisis de RMSF para las proteínas wt (rojo) y ProE3Gly (naranja) de Tf-trHbO (A) y wt (rojo) y GlyE3Pro (violeta) de Mt-trHbO (B), todas a 87°C. Las diferencias significativas en ambos loops CD y en los extremos están resaltadas en verde. Las flechas indican los cambios en la fluctuación luego de la mutación *in silico*.

En el caso de la mutante GlyE3Pro de la Mt-trHbO, no se detectó una ruptura clara de la hélice α E, tal vez debido a la escala temporal de la simulación y/o a limitaciones del campo de fuerzas utilizado. A pesar de que es importante notar que las simulaciones de DM típicamente en la escala temporal de nanosegundos pueden no capturar cambios conformacionales grandes, los movimientos estructurales observados fueron suficientes para encontrar al menos una diferencia significativa entre las proteínas estudiadas. Estas observaciones permitieron arribar a conclusiones que se encuentran en sintonía con las mediciones experimentales mencionadas más adelante, en la sección VI.3.4.

Los resultados aquí presentados revelan que las simulaciones de las mutantes *in silico* apoyan la hipótesis de un rol clave del aminoácido E3 en la dinámica del loop CD, y sugieren una interesante conexión con la estabilidad térmica global del plegado de estas trHbs.

VI.3.3 Impacto en interacciones polares

Un análisis de secuencias permitió destacar un enriquecimiento general en aminoácidos cargados por parte de la proteína termoestable, respecto a su contraparte mesoestable (54 contra 41). A pesar de esta diferencia, se notó que la cantidad de interacciones polares (puentes salinos y puentes de hidrógeno) observada a lo largo de las simulaciones de DM, se mantenían similares a temperatura ambiente (tabla 2). Con el aumento de la temperatura a 87°C, ambas proteínas experimentaron una pérdida de interacciones polares, principalmente las de tipo proteína-solvente (tabla 2). La tendencia observada referida a las interacciones salinas es, sin embargo, notoriamente distinta: la Tf-trHbO wt muestra un incremento de ~6 puentes salinos a 87°C debido a la ganancia del aumento en la flexibilidad del loop CD, lo que facilita la formación de interacciones de puente salinos entre aminoácidos del loop CD y la hélice α B (figura VI.5).

La mutante GlyE3Pro de la Mt-trHbO sigue la misma tendencia y confirma que la mutación de este único aminoácido promueve el incremento en la flexibilidad del loop CD y, consecuentemente, del establecimiento de nuevos puentes salinos (tabla 2).

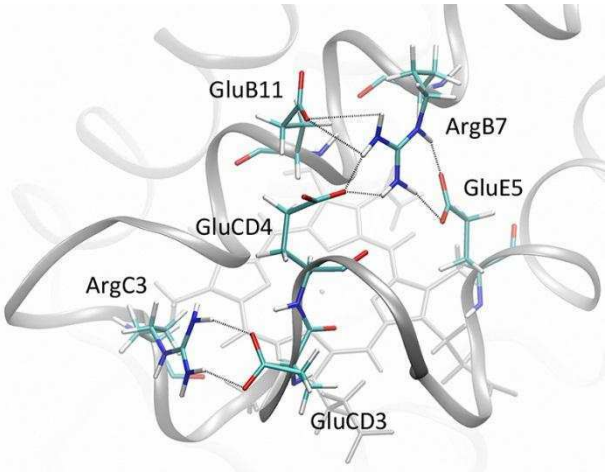


Figura VI.5. Red de puentes salinos entre aminoácidos del loop CD y la α-hélice B.

En síntesis, ambas proteínas muestran cantidades comparables de interacciones polares en el estado plegado, y las interacciones de puentes de hidrógeno parecen estar afectadas de manera similar por el incremento en la temperatura, a pensar de su termoestabilidad intrínseca. Por otro lado, se detectó una clara diferencia en la cantidad de puentes salinos que aumentan en presencia de ProE3 y a altas temperaturas.

Tabla 2. Cantidad de interacciones polares, puentes salinos y puentes de hidrógeno (pteH) calculados para las proteínas wt y mutantes a 27 y 87°C durante los 100ns de DM. Los puentes salinos y los pteH están definidos considerando una distancia de corte de menos de 3.3 Å entre dos átomos cargados o electronegativos, respectivamente.

		Tf-trHbO		Mt-trHbO	
		27°C	87°C	27°C	87°C
Wild type	Puentes salinos	17 ± 1	23 ± 2	17 ± 2	16 ± 2
	pteH - intra	53 ± 5	50 ± 5	57 ± 5	46 ± 4
	pteH - totales	230 ± 12	189 ± 11	241 ± 13	201 ± 12
Mutante	Puentes salinos	17 ± 1	18 ± 2	17 ± 2	23 ± 1
	pteH - intra	54 ± 5	47 ± 5	55 ± 5	50 ± 5
	pteH - totales	233 ± 14	190 ± 12	238 ± 14	201 ± 12

VI.3.4 Mediciones de desnaturalización de proteínas *wild type* y mutantes intercambiadas

Los experimentos presentes en esta sección de resultados fueron todos llevados a cabo por el grupo de investigación previamente citado en la sección "VI. Introducción" de este capítulo.

Con el propósito de determinar y caracterizar las propiedades termodinámicas que pueden estar afectadas por el único cambio del aminoácido correspondiente a la posición estructural E3, se examinó la señal de dicroísmo circular a 225nm en función de la temperatura para las proteínas wt y mutantes para distintas concentraciones de cloruro de guanidinio (GdmCl). Se logró medir reversibilidad en el proceso de desplegado a la concentración de GdmCl más alta. Los perfiles observados, ajustados a un único proceso de desplegado de dos estados bajo la suposición de Δ*Cp* independiente de la temperatura (ver capítulo Anexo, figura A.VI.1), permiten obtener un claro panorama sobre los determinantes termodinámicos de la termoestabilidad en ambas proteínas wt y sus correspondientes mutantes en E3. Como se esperaba, el análisis de los datos de los perfiles de fusión térmica indican que la Tf-trHbO es significativamente más termoestable que Mt-trHbO bajo todas las condiciones examinadas, mostrando una T_m de 75°C contra 57°C. Es también interesante resaltar que las cuatro proteínas se caracterizan por presentar valores similares de Δ*Cp* y estabilidades de solvente comparables, como lo indican los experimentos de desnaturalización isotérmica por GdmCl (figura A.VI.2). En cuanto al comportamiento térmico de la Tf-trHbO respecto a su mutante ProE3Gly, se observó una notable disminución en valores de T_m y de Δ*Cp* en la proteína mutante. En particular, los valores de T_m de las proteínas wt y mutante, 75°C y 71°C respectivamente, son fuertemente divergentes en una solución *buffer*, pero se asemejan al incrementar la concentración de GdmCl.

En otras palabras, el GdmCl parece bloquear el efecto de la mutación en el perfil de fusión térmica, esto es, a concentraciones mayores a 2 M, las curvas de desplegado son prácticamente superponibles en la proteína nativa y en la mutante. En términos estructurales, las nuevas interacciones de puentes salinos que estabilizan al estado de la proteína nativa, ahora la desestabilizan por la presencia de un cosolvente altamente polar, donde las moléculas de GdmCl tienen un completo acceso desde el solvente (figura VI.5). De esta manera, se obtiene el mismo resultado que para la mutación de ProE3Gly, donde las interacciones de puentes salinos entre el loop CD y la hélice α B son desestabilizadas por otra razón estructural, la reducción del largo del loop CD (figura VI.3B).

En este contexto, la interpretación termodinámica es consistente con el rol de las interacciones polares mencionadas y su dependencia con la extensión del loop CD, como fue previamente presentado y explicado. Además, el balance energético hallado, $\Delta\Delta G_{desplegado}$ (wt menos mutante) de cerca de 1 kcal/mol, es también consistente con el escenario propuesto.

Los datos del desplegado químico y térmico para Mt-trHbO y su mutante GlyE3Pro indican un escenario contrario respecto a Tf-trHbO y su mutante ProE3Gly. De hecho, la sustitución de glicina por prolina lleva a un incremento muy significativo en la T_m, de 57°C a 65°C, con un incremento de $\Delta\Delta G_{desplegado}$ (wt menos mutante) de cerca de 0,5 kcal/mol. En cuanto al efecto inhibitorio del GdmCl, se observa en menor medida respecto a la Tf-trHbO. Sobre estas bases, los determinantes estructurales para aumentar la termoestabilidad dada la mutación GlyE3Pro en Mt-trHbO, son mayoritariamente explicados en términos de un incremento en la flexibilidad del loop CD, lo cual confiere una estructura adicional con interacciones polares estabilizantes para la estructura proteica (como se muestra en la figura VI.5 para la Tf-trHbO wt).

Todos los resultados teóricos y experimentales presentados en este capítulo han sido publicados en el siguiente artículo:

JP. Bustamante, A. Bonamore, AD. Nadra, N. Sciamanna, A. Boffi, DA. Estrin and L. Boechi. *Molecular Basis of Thermal Stability in Truncated (2/2) Hemoglobins*. **BBA General Subjects**, 2014, 1840, pp 2281-2288

VI.4 Conclusiones

VI.4.1 Flexibilidad frente al cambio de temperatura

Los resultados teóricos y experimentales aquí presentados demuestran que hay importantes diferencias estructurales y dinámicas entre las trHbs termoestable y no termoestable al

someterlas a altas temperaturas. Las simulaciones de DM permitieron analizar si la termoestabilidad correlacionaba con la flexibilidad / rigidez de la estructura proteica. Comparando trHbs wt, no se encontraron diferencias significativas en la flexibilidad de los aminoácidos observados a temperatura ambiente. Previamente, se encontraron resultados similares con experimentos de intercambios de amidas[8], aunque otras evidencias teóricas y experimentales han mostrado diferentes grados de flexibilidad para una proteína termoestable respecto a su homóloga no termoestable[3–7]. Por otro lado, se observó un claro incremento en la flexibilidad del loop CD en la forma wt de la Tf-trHbO respecto a la Mt-trHbO wt cuando se las sometió a altas temperaturas.

VI.4.2 Interacciones polares frente al cambio de temperatura

No se observaron diferencias sustanciales en la cantidad total de interacciones polares (puentes de hidrógeno y salinos) entre la trHb termoestable y la no termoestable a temperatura ambiente. Cuando la temperatura se incrementó a 87°C, ambas proteínas perdieron varios puentes de hidrógeno, principalmente las interacciones aminoácidos – moléculas de agua. Esto significa que ambas proteínas se vieron afectadas de manera similar por el cambio en la perturbación térmica considerando los puentes de hidrógeno. Este hecho puede estar relacionado con una observación previa acerca de la pérdida de la primera capa de solvatación a altas temperaturas[14,15].

Ya se mencionó que la T_m está relacionada con la diferencia entre la energía de los estados colapsados plegado versus desplegado. A pesar de que no hay datos sobre interacciones polares en el estado colapsado desplegado, se demostró que la Tf-trHbO wt incrementa su cantidad de puentes salinos con el aumento de la temperatura, lo cual es posible gracias a reacomodamientos estructurales que permiten que aminoácidos de superficie cargados queden disponibles para formar interacciones salinas, tal como se muestra en la figura VI.5. Este fenómeno involucra una disminución en la magnitud de la energía libre de solvatación de los aminoácidos cargados a altas temperaturas, lo cual está en un balance energético con un aumento de interacciones coulómbicas y electroestáticas de cadenas laterales de estos aminoácidos, produciendo un aumento en la estabilidad del plegado global frente a la perturbación de la temperatura. Sin embargo, la cantidad de interacciones de puentes salinos no aumenta en el caso de la Mt-trHbO wt, debido a la inhabilidad estructural de formar estos puentes salinos dado su pequeño loop CD. Puesto que en la mutante GlyE4Pro de la Mt-trHbO el ambiente del loop CD es prácticamente el mismo que en la Tf-trHbO wt, con un loop CD largo, se observó el mismo incremento de puentes salinos, ganando interacciones que se suman al

estado plegado. De manera interesante, un único aminoácido, una prolina, no solo brinda mayor flexibilidad al loop CD, sino que además permite la formación de nuevas interacciones que estabilizan al estado nativo de la trHb.

VI.4.3 Influencia de la identidad de la posición E3 sobre el cambio de temperatura

Se ha propuesto que las prolinas aumentan la T_m reduciendo la entropía configuracional de las proteínas, fomentando estructuras más rígidas[11]. Esta interesante hipótesis considera solo la estructura primaria de una proteína, y no los cambios en las estructuras secundarias y terciarias que estos aminoácidos puedan causar. A través de las simulaciones de DM realizadas, se encontró que la presencia de una prolina específica, la PE3, permite al loop CD ganar flexibilidad en el estado nativo. Las mediciones de T_m de las proteínas wt y de sus respectivas mutantes confirman que un aminoácido específico es capaz de aumentar (en 8°C) o disminuir (en 4°C) la T_m en la proteína correspondiente. Por lo tanto, aquí se propone que el incremento de la T_m está relacionado con una distorsión terciaria producida solo por la ProE3. Específicamente, este aminoácido desbloquea al loop CD, haciendo que pueda fluctuar más activamente e interactuar con los aminoácidos de la hélice α B, estabilizando así al estado plegado.

En concordancia con los resultados presentados, todos los estudios previos sobre globinas no termoestables provenientes de microorganismos mesófilos como *M. tuberculosis* (Mt-trHbN), *B. subtilis* (Bs-trHbO) y *P. catodon* (mioglobina de cachalote), muestran la misma tendencia que la Mt-trHbO, sus loops (y particularmente el loop CD) se observan casi rígidos[16–18], sin mostrar el cambio conformacional observado en la Tf-trHbO a altas temperaturas. Por otro lado, con un loop CD largo y muy flexible[16], la neuroglobina humana hexacoordinada muestra una T_m de 100°C[19], donde el loop CD parece ser una región crítica en su estabilidad térmica, un hecho probablemente relacionado a hexacoordinación interna. De esta manera, este estudio elucida la importancia de la flexibilidad del loop CD en la termoestabilidad de trHbs y otros plegados globulares y abre las puertas hacia posteriores estudios que profundicen sobre las bases moleculares de la daptación a condiciones extremas.

VI.5 Referencias

(1) Bonamore, A.; Ilari, A.; Giangiacomo, L.; Bellelli, A.; Morea, V.; Boffi, A. *FEBS J.* **2005**, *272* (16), 4189–4201.

(2) Hicks, P. M.; Adams, M. W. W.; Kelly, R. M. *M. C. Flickinger S. W. Drew (ed.), Encycl. bioprocess Technol. Ferment. Biocatal. Biosep. John Wiley Sons, Inc., New York.* **1999**, 987–1004.

(3) Lazaridis, T.; Lee, I.; Karplus, M. *Protein Sci.* **1997**, *6* (12), 2589–2605.

(4) Manco, G.; Giosuè, E.; D'Auria, S.; Herman, P.; Carrea, G.; Rossi, M. *Arch. Biochem. Biophys.* **2000**, *373* (1), 182–192.

(5) Gershenson, A.; Schauerte, J. A.; Giver, L.; Arnold, F. H. *Biochemistry* **2000**, *39* (16), 4658–4665.

(6) Závodszky, P.; Kardos, J.; Svingor; Petsko, G. A. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (13), 7406–7411.

(7) Mamonova, T.; Glyakina, A.; Galzitskaya, O.; Kurnikova, M. *Biochim. Biophys. Acta.* **2013**, *1834* (5), 854–866.

(8) Hernández, G.; Jenney Jr., F. E.; Adams, M. W. W.; LeMaster, D. M. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97* (7), 3166–3170.

(9) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins Struct. Funct. Genet.* **1995**, *21* (3), 167–195.

(10) Vogt, G.; Woell, S.; Argos, P. *J. Mol. Biol.* **1997**, *269* (4), 631–643.

(11) Matthews, B. W.; Nicholson, H.; Becktel, W. J. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84* (19), 6663–6667.

(12) Bohm, G.; Jaenicke, R. *Int. J. Pept. Protein Res.* **1994**, *43* (1), 97–106.

(13) Milani, M.; Savard, P.-Y.; Ouellet, H.; Ascenzi, P.; Guertin, M.; Bolognesi, M. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (10), 5766–5771.

(14) De Bakker, P. I. W.; Hünenberger, P. H.; McCammon, J. A. *J. Mol. Biol.* **1999**, *285* (4), 1811–1830.

(15) Makhatadze, G. I.; Loladze, V. V; Ermolenko, D. N.; Chen, X.; Thomas, S. T. *J. Mol. Biol.* **2003**, *327* (5), 1135–1148.

(16) Capece, L.; Marti, M. A.; Bidon-Chanal, A.; Nadra, A.; Luque, F. J.; Estrin, D. A. *Proteins* **2009**, *75* (4), 885–894.

(17) Giangiacomo, L.; Ilari, A.; Boffi, A.; Morea, V.; Chiancone, E. *J. Biol. Chem.* **2005**, *280* (10), 9192–9202.

(18) Crespo, A.; Martí, M. A.; Kalko, S. G.; Morreale, A.; Orozco, M.; Gelpi, J. L.; Luque, F. J.; Estrin, D. A. *J. Am. Chem. Soc.* **2005**, *127* (12), 4433–4444.

(19) Hamdane, D.; Kiger, L.; Dewilde, S.; Uzan, J.; Burmester, T.; Hankeln, T.; Moens, L.; Marden, M. C. *FEBS J.* **2005**, *272* (8), 2076–2084.

VII. Conclusiones y perspectivas generales

Mediante la combinación de herramientas brindadas por la fisicoquímica, el modelado y la simulación computacional y, la bioinformática, en la presente tesis doctoral se pretendió profundizar en la comprensión sobre cuáles son y cómo se modulan los determinantes moleculares de la captación y liberación de ligandos en hemoglobinas truncadas. El estudio de dos factores cruciales, las cavidades internas y el rol de las moléculas de agua presentes en el sitio activo de estas proteínas, ha permitido mejorar la comprensión sobre el proceso de captación de ligandos pequeños. Dichos análisis integradores presentados realzan la utilidad de simulaciones de DM complementadas con experimentos cinéticos de unión de ligandos, brindando una caracterización muy detallada que, de otra manera, sería difícil de obtener.

Hallar una correspondencia cuantitativa entre resultados experimentales y teóricos no es una tarea sencilla, puesto que implica conocer el delicado balance entre las variables que determinan un proceso y la relación exacta entre éstas. Adentrándose en esta problemática, un segundo objetivo fue diseñar un modelo fisicoquímico sencillo que permita estimar y posteriormente predecir constantes k_{on} y k_{off} de O_2. Específicamente, se modelaron los fenómenos de captación y liberación de O_2, identificando la mínima información requerida para modelar estos fenómenos de manera aceptable. Luego, se utilizaron dichos modelos para predecir constantes k_{on} y k_{off} de O_2. Los resultados alcanzados mostraron una muy buena capacidad para estimar dichas constantes, con diferencias máximas entre valores predichos de 1,5 órdenes de magnitud. Este error obtenido resulta ser del mismo orden que la varianza entre constantes experimentales reportadas independientemente en la literatura por distintos grupos de investigación y técnicas utilizadas, lo que respalda la ausencia de cualquier equivocación teórica significativa en la construcción de los modelos.

Con este poder predictivo, y dado que la mayoría de trHbs no poseen una función asignada, el tercer objetivo se centró en desarrollar un estudio evolutivo-molecular de toda la familia de trHbs, asignando a cada miembro una posible función en base a su capacidad para captar y liberar O_2. Se asignaron posibles funciones moleculares y se analizó a esta familia de proteínas en un contexto evolutivo-funcional, caracterizando a cada grupo filogenético. En particular, los aportes de este estudio se resumen a continuación:

- ***Se realizó una revisión del árbol filogenético de trHbs, considerando 1107 secuencias halladas luego de una exahustiva búsqueda y posterior filtrado.*** La distribución de las secuencias en los grupos N, O y P en el árbol bayesiano generado fue la misma obtenida por el primer árbol realizado en el año 2006 por Vuletich y Lecomte, con 111 secuencias. En el año 2013, en un estudio liderado por Vinogradov sobre globinas se mencionó que la cantidad de trHbs anotadas en bases de datos había aumentado a aproximadamente 1200 y se presentó un árbol filogenético con un conjunto seleccionado de secuencias.
- ***Se halló un cuarto grupo de trHbs, denominado Q (o IV).*** Además de la divergencia evolutiva observada por la ubicación en el árbol, este grupo mostró características a nivel de secuencia distintivas que lo diferencian del resto de los grupos, cada uno de ellos con sus aminoácidos claves.
- ***Se construyeron modelos de trHbs representativos con combinaciones de aminoácidos claves definiendo los procesos de captación y liberación de O_2.*** Esto permitió, utilizando, en principio, los modelos fisicoquímicos desarrollados en el capítulo IV, la asignación a cada miembro de valores cinéticos predichos de k_{on} y k_{off}.
- ***Se mapearon características fisicoquímicas sobre el árbol filogenético.*** Este contexto evolutivo-fisicoquímico permitió estudiar la diversificación funcional presente en la familia de trHbs, además de la identificación de algunas diferencias entre distintos subgrupos dentro de los cuatro grupos N, O, P y Q respecto a sus capacidades para captar y liberar O_2.
- ***Se estudió la posible diversificación funcional.*** En base a casos estudiados previamente por otros autores e incluso por el grupo de investigación en el que se desarrolló esta tesis doctoral, se asignaron tres posibles funciones a las trHbs, considerando la reactividad predicha frente a O_2. Las funciones posibles son: una química de multiligandos NO/O_2 (tipo Mt-trHbN), almacenamiento y transporte de O_2 (tipo mioglobina) y catalasa-peroxidasa (tipo Tf-trHbN o Mt-trHbO). Esta asignación funcional evidenció que dentro de un mismo grupo filogenético puede haber diferentes funciones posibles, por lo que no es posible hacer asignaciones funcionales partiendo solo de análisis evolutivos. Por el contrario, estos estudios demuestran que es necesario recurrir a un enfoque integrador, donde la sinergia entre la fisicoquímica, la bioinformática y la filogenia pueden dar respuestas a interrogantes de este tipo.

Cabe destacarse que este último punto es, posiblemente, la contribución más novedosa de esta tesis, dado que no existía hasta el momento un estudio evolutivo-funcional de toda la familia de las hemoglobinas truncadas.

En el último capítulo de esta tesis se presentó un primer estudio sobre termoestabilidad de una trHb, siendo éste un punto de partida para posteriores estudios que profundicen sobre las bases moleculares de la adaptación a condiciones extremas en estas proteínas.

VIII. Otras publicaciones

Durante el transcurso de esta tesis doctoral, además de las publicaciones mencionadas en los capítulos correspondientes, se ha participado de los siguientes trabajos científicos:

1. A. Pesce, **JP. Bustamante**, A. Bidon-Chanal, L. Boechi, D. Estrin, FJ. Luque, A. Sebilo, M. Guertin, M. Bolognesi, A. Martino, P. Ascenzi, N. Marco. The N-terminal pre-A region of Mycobacterium tuberculosis 2/2HbN promotes NO-dioxygenase activity. **FEBS Journal**, 2015. Aceptado.

2. D. Giordano, A. Pesce, L. Boechi, **JP. Bustamante**, E. Caldelli, BD. Howes, A. Riccio, G. di Prisco, M. Nardini, D. Estrin, G. Smulevich, M. Bolognesi, C. Verde. *Structural flexibility of the heme cavity in the cold-adapted truncated hemoglobin from the Antarctic marine bacterium Pseudoalteromonas haloplanktis TAC125*. **FEBS Journal**, 2015, 282(15), pp 2948-2965.

3. FP. Nicoletti*, **JP. Bustamante***, E. Droghetti, BD. Howes, M. Fittipaldi, A. Bonamore, P. Baiocco, A. Feis, A. Boffi, DA. Estrin, G. Smulevich. *Interplay of the H-bond donor-acceptor role by distal residues in the hydroxyl ligand stabilization of Thermobifida fusca truncated hemoglobin*. **Biochemistry**, 2014, 53, pp 8021−8030.

4. FP. Nicoletti, E. Droghetti, BD. Howes, **JP. Bustamante**, A. Bonamore, N. Sciamanna, DA. Estrin, A. Feis, A. Boffi, G. Smulevich. *H-bonding Networks of the Distal Residues and Water Molecules in the Active Site of Thermobifida fusca Hemoglobin*. **Biochimica et Biophysica Acta**, 2013, 1834(9), pp 1901-1909.

5. DF. Gauto, C. Modenutti, VG. Dumas, L. Alvarez, **JP. Bustamante**, AG. Turjanski and MA. Marti. *Determining Free Energies of protein-ligand binding, and association/dissociation processes using computer simulations*. **World Research Journal of Peptide and Protein**, 2012, 1(1), pp 21-32.

IX. Anexo

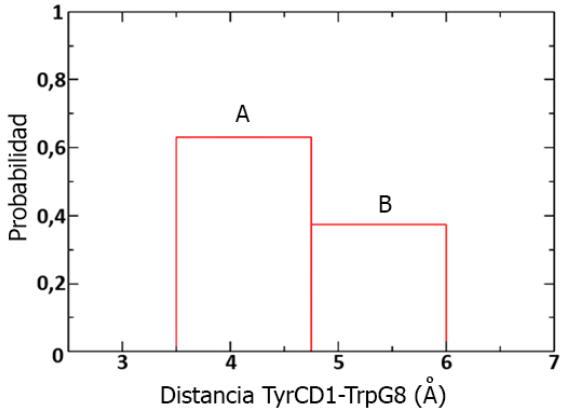


Figura A.III.1. Histograma de la distancia TyrCD1-TrpG8 en la Tf-trHbO, muestreando ambas conformaciones espaciales (A y B) de la YCD1 durante la simulación de DM.

Tabla A.IV.1. Valores de $log(k_{on})$ y los hallados utilizando los modelos de $log(k_{on}chem)$ y $log(k_{on}stat)$ para todas las proteínas estudiadas. Los valores de k_{on} poseen unidades de $M^{-1}s^{-1}$.

Organismo **trHb – *wild type* o mutante**	$\boldsymbol{log(k_{on})}$	$\boldsymbol{log(k_{on}chem)}$	$\boldsymbol{log(k_{on}stat)}$
Mycobacterium tuberculosis trHbO – wt	5,04	5,02	5,33
Mycobacterium tuberculosis trHbO - YCD1F	6,43	6,81	6,45
Mycobacterium tuberculosis trHbO - WG8F	7,10	7,31	6,83
Mycobacterium tuberculosis trHbO - WG8F - YCD1F	7,36	8,39	7,55
Mycobacterium leprae trHbO - wt	5,04	5,09	5,29
Thermobifida fusca trHbO - wt	5,95	5,38	5,65
Thermobifida fusca trHbO - WG8F	6,53	7,46	7,01
Thermobifida fusca trHbO - WG8F - YB10F - YCD1F	6,58	8,30	7,08
Pseudoalteromonas haloplanktis trHbO - wt	5,95	5,31	5,56
Agrobacterium tumefaciens trHbO - wt	5,30	5,09	5,29

Bacillus Subtilis trHbO - wt	7,48	7,39	6,92
Geobacillus Subtilis trHbO - wt	7,88	7,17	6,77
Campylobacter jejuni trHbP - wt	5,96	5,24	5,47
Campylobacter jejuni trHbP - YB10F	5,08	6,24	6,23
Campylobacter jejuni trHbP - YB10F - HE7L	7,62	9,26	8,39
Synechocystis trHbN - wt	8,38	7,41	7,69
Mycobacterium tuberculosis trHbN - wt	7,40	7,05	7,36
Mycobacterium tuberculosis trHbN - YB10F	7,73	8,48	8,28
Mycobacterium tuberculosis trHbN - YB10L	8,79	8,12	7,96
Mycobacterium tuberculosis trHbN - QE11V	7,51	7,26	7,50
Mycobacterium tuberculosis trHbN - QE11A	7,57	7,55	7,87
Mycobacterium tuberculosis trHbN - YB10L - QE11V	9,26	10,39	8,82
Chlamydomonas eugametos trHbN - wt	6,00	6,82	6,83
Paramecium caudatum trHbN - wt	7,48	7,12	7,45
Mioglobina de cachalote Mb - wt	7,20	7,33	7,47
Mioglobina de cachalote Mb - HE7G	8,18	8,68	7,79

Tabla A.IV.2. Valores de $log(k_{off})$ y los hallados utilizando los modelos de $log(k_{off}chem)$ y $log(k_{off}stat)$ para todas las proteínas estudiadas. Los valores de k_{off} poseen unidades de s^{-1}.

Organismo **trHb – *wild type* o mutante**	$log(k_{off})$	$log(k_{off}chem)$	$log(k_{off}stat)$
Mycobacterium tuberculosis trHbN - wt	-0,70	-1,74	-1,46
Mycobacterium tuberculosis trHbN - YB10A	1,65	0,07	0,01
Mycobacterium tuberculosis trHbO - wt	-2,85	-1,59	-1,33
Mycobacterium tuberculosis trHbO - WG8F	-0,26	0,71	0,53
Chlamydomonas eugametos trHbN - wt	-1,85	-4,15	-3,42
Campylobacter jejuni trHbP - wt	-2,39	-2,64	-2,19
Campylobacter jejuni trHbP - WG8F	-1,48	-1,44	-1,22
Campylobacter jejuni trHbP - HE7L	-3,52	-3,27	-2,70
Bacillus Subtilis trHbO - wt	-2,32	-1,59	-1,34
Thermobifida fusca trHbO - wt	-1,15	-1,55	-1,30
Paramecium caudatum trHbN - wt	1,40	1,29	1,01
Mioglobina de cachalote Mb - wt	1,18	1,56	1,22
Mioglobina de cachalote Mb - HE7G	3,20	3,84	3,08

Tabla A.V.1. Análisis de posiciones determinantes de especificidad (SDPs), información mutua (IM) y IM acumulada (IMa) con sus posiciones estructurales correspondientes. Se muestran los datos obtenidos con SDPfox y Mistic. Las SDPs se ordenaron de acuerdo a su valor de IM con la SDP E7.

Orden	SDPfox		Mistic		Posición
	Puntaje Z	p-valor	IM con E7	IMa	estructural
1	86,90	-3775,22	NA	223,82	E7
2	56,57	-6400,31	19,67	208,72	E15
3	81,76	-6683,97	15,34	214,79	G12
4	33,81	-8009,17	15,24	119,32	H9
5	35,42	-6903,53	14,59	189,96	G8
6	29,97	-8995,46	14,10	96,49	G11
7	59,80	-5363,92	12,77	164,15	F4
8	41,87	-5257,86	11,84	154,70	E4
9	30,50	-8845,49	10,71	64,19	C6
10	51,02	-6506,17	10,56	124,83	G5
11	33,07	-8208,31	8,80	225,91	E20
12	38,29	-5865,07	8,66	130,37	E18
13	34,63	-7200,55	7,68	127,17	EF8
14	36,63	-6039,21	7,45	61,20	F7
15	30,64	-8460,71	7,42	72,42	F5
16	31,74	-8066,83	6,55	85,80	H13
	31,61	-8502,18	5,53	77,03	H16
	35,94	-6459,36	5,19	63,01	FG1
	33,85	-7452,90	4,46	64,40	H12
	40,95	-5868,73	4,04	36,43	E17

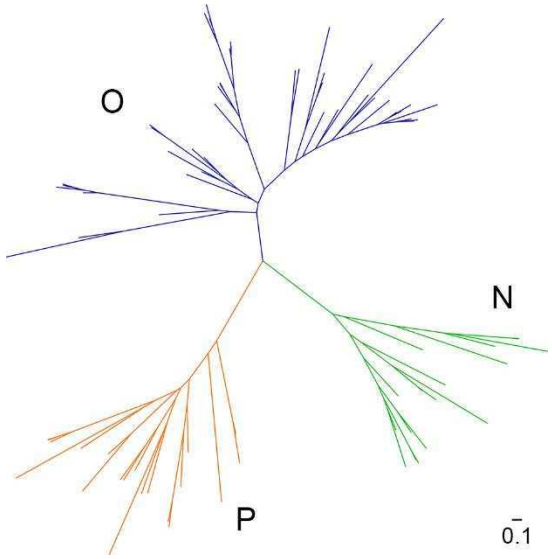


Figura A.V.1. Árbol de máxima verosimilitud con las secuencias tomadas del trabajo de Vuletich y Lecomte del año 2006. Se obtuvo la misma divergencia filogenética. Puede observarse la clasificación de trHbs en los tres grupos conocidos hasta ese momento, N, O y P.

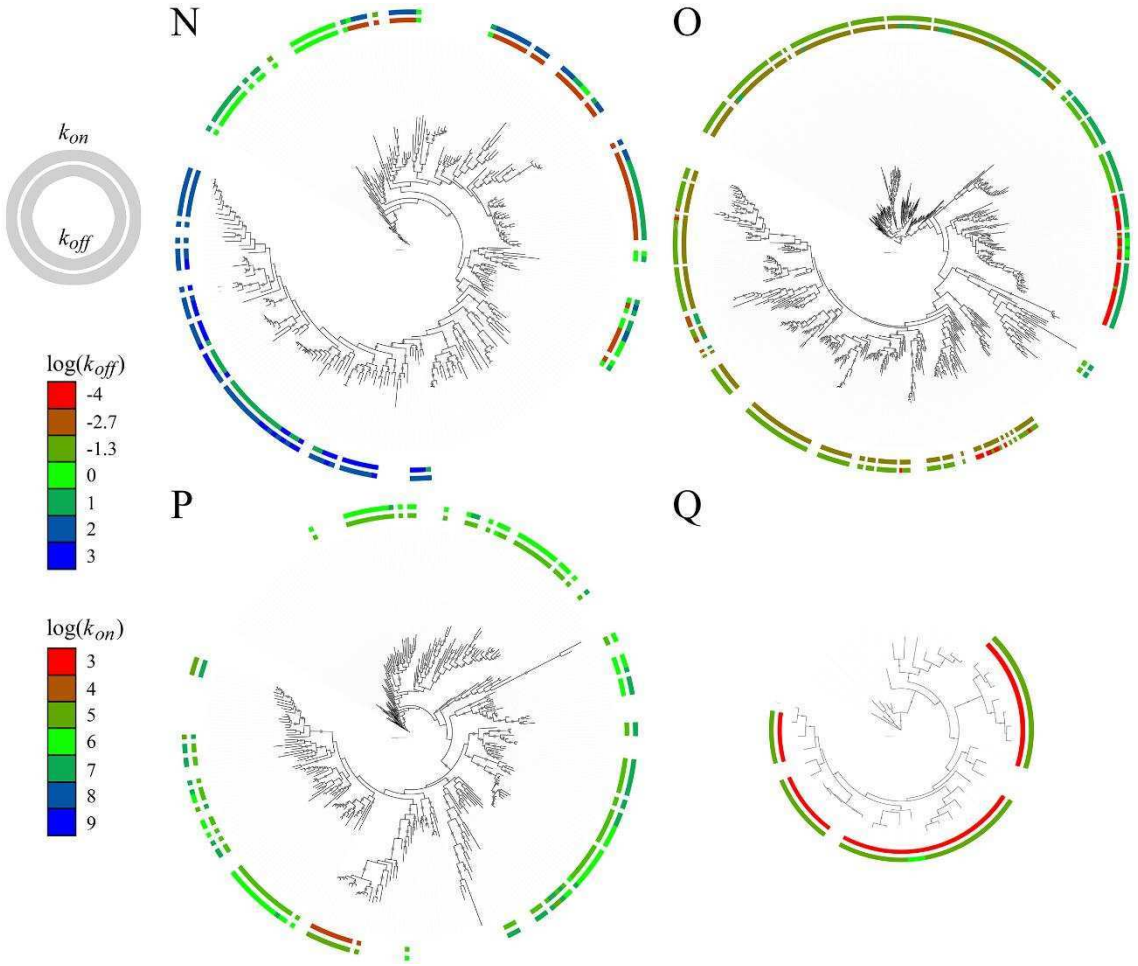


Figura A.V.2. Árboles filogenéticos bayesianos de cada uno de los cuatro grupos de trHbs, mostrados como filogramas circulares con valores mapeados de $log(k_{on}chem)$ y $log(k_{off}chem)$. Los cladogramas muestran la topología derivada de la figura V.2A. Los valores de k_{on} y k_{off} poseen unidades de $M^{-1}s^{-1}$ y s^{-1}, respectivamente.

Aminoácidos del sitio activo B10-CD1-E7-E11-G8	$k_{off}chem$ $[s^{-1}]$	K_{H_2O}	Especies
YHTLW	3,52E-03	3,04E-05	O: Xanthobacter autotrophicus Py2, marine gamma proteobacterium HTCC2143
YFMKW	4,47E-04	3,04E-05	P: Rhizobium leguminosarum bv trifolii, Agrobacterium sp ATCC 31749, Agrobacterium tumefaciens F2, Sinorhizobium medicae WSM419, Agrobacterium tumefaciens CCNWGS0286, Sinorhizobium fredii NGR234, Rhizobium sp PDO1-076, Rhizobium etli CFN 42, Rhizobium etli CIAT 652, Rhizobium leguminosarum bv viciae, Sinorhizobium meliloti CCNWSX0020
VFLLL	2,69E+04	1,87E-01	N: Acidithiobacillus caldus SM-1, Limnobacter sp MED105, Halothiobacillus neapolitanus c2, Oxalobacteraceae bacterium IMCC9480, Burkholderiales bacterium JOSHI_001, Cupriavidus necator N-1, Candidatus Nitrospira defluvii, Candidatus Nitrospira defluvii, Candidatus Nitrospira defluvii, Candidatus Nitrospira defluvii, Acidovorax sp NO-1
YHMLW	2,35E+01	3,04E-05	O: Sphingobium sp SYK-6, marine gamma proteobacterium HTCC2148
LFFFV	1,08E+02	1,87E-01	N: Halomonas sp TD01, Congregibacter litoralis KT71
YFKVW	2,93E+00	5,27E-04	P: Leadbetterella byssophila DSM 17132, Leadbetterella byssophila DSM 17132

YYALW	6,65E-03	3,04E-05	O: Mycobacterium tuberculosis NCGM2209, Clavibacter michiganensis subsp sepedonicus, Micrococcus luteus SK58, Corynebacterium efficiens YS-314, Streptomyces sviceus ATCC 29083, Microbacterium laevaniformans OR221, Mycobacterium rhodesiae NBB3, Leifsonia xyli subsp xyli, Actinoplanes sp SE50110, Mycobacterium intracellulare MOTT-64, Cellulomonas flavigena DSM 20109, Corynebacterium glutamicum ATCC 14067, Streptomyces venezuelae ATCC 10712, Blastococcus saxobsidens DD2, Xylanimonas cellulosilytica DSM 15894, Gordonia effusa NBRC 100432, Janibacter sp HTCC2649, Streptomyces zinciresistens K42, Beutenbergia cavernae DSM 12333, Frankia alni ACN14a, Segniliparus rotundus DSM 44985, Mycobacterium tusciae JS617, Corynebacterium amycolatum SK46, Corynebacterium variabile DSM 44702, Corynebacterium accolens ATCC 49725, Arthrobacter arilaitensis Re117, Saccharomonospora marina XMU15, Mycobacterium parascrofulaceum ATCC BAA-614, Mycobacterium avium 104, Streptomyces ghanaensis ATCC 14672, Mycobacterium vanbaalenii PYR-1, Nocardia farcinica IFM 10152, Corynebacterium diphtheriae NCTC 13129, Saccharomonospora cyanea NA-134, Brachybacterium faecium DSM 4810, Mycobacterium colombiense CECT 3035, Corynebacterium lipophiloflavum DSM 44291, Corynebacterium resistens DSM 45100, Frankia sp CN3, Streptomyces griseoaurantiacus M045, Mycobacterium abscessus subsp bolletii, Nakamurella multipartita DSM 44233, marine actinobacterium PHSC20C1, Isoptericola variabilis 225, Jonesia denitrificans DSM 20603, Frankia sp EAN1pec, Thermobifida fusca, Corynebacterium ammoniagenes DSM 20306, Dietzia cinnamea P4, Algoriphagus machipongonensis, Corynebacterium casei UCMA 3821, Mycobacterium gilvum PYR-GCK, Streptomyces cattleya NRRL 8057, Gordonia araii NBRC 100433, Saccharomonospora paurometabolica YIM 90007, Streptomyces griseoflavus Tu4000, Streptomyces viridochromogenes DSM 40736, Corynebacterium pseudogenitalium ATCC 33035, Micromonospora sp ATCC 39149, Corynebacterium glutamicum ATCC 13032, Stackebrandtia nassauensis DSM 44728, Saccharomonospora glauca K62, Frankia sp CcI3, Mycobacterium rhodesiae JS60, Streptomyces hygroscopicus subsp jinggangensis, Mycobacterium thermoresistibile ATCC 19527, Gordonia sputi NBRC 100414, Mycobacterium leprae, Gordonia otitidis NBRC 100426, Nocardia brasiliensis ATCC 700358, Mycobacterium sp MCS, Kocuria rhizophila DC2201, Corynebacterium striatum ATCC 6940, Corynebacterium jeikeium K411, Corynebacterium glucuronolyticum ATCC 51866, Micromonospora aurantiaca ATCC 27029, Streptomyces violaceusniger Tu 4113, Microbacterium testaceum StLB037, Streptomyces sp SPB78, Mycobacterium marinum M, Frankia sp EuI1c, Saccharopolyspora erythraea NRRL 2338, Saccharomonospora viridis DSM 43017, Streptomyces himastatinicus ATCC 53653, Gordonia alkanivorans NBRC 16433, Corynebacterium aurimucosum ATCC 700975, Salinispora tropica CNB-440, Mycobacterium smegmatis str MC2, Streptomyces bingchenggensis BCW-1, Saccharomonospora azurea NA-128, Mycobacterium sp JDM601, Streptomyces sp e14, Frankia sp EUN1f, Gordonia neofelifaecis NRRL B-59395, Kytococcus sedentarius DSM 20547, Gordonia bronchialis DSM 43247, Corynebacterium ulcerans 809, Salinispora arenicola CNS-205, Cellulomonas fimi ATCC 484, Nocardiopsis dassonvillei subsp dassonvillei, Catenulispora acidiphila DSM 44928, Nocardia cyriacigeorgica GUH-2, Gordonia polyisoprenivorans VH2, Thermomonospora curvata DSM 43183, Geodermatophilus obscurus DSM 43160, Mobilicoccus pelagius NBRC 104925, Corynebacterium pseudotuberculosis 106-A, Planctomyces limnophilus DSM 3776, Frankia symbiont of Datisca, Corynebacterium genitalium ATCC 33030, Brevibacterium mcbrellneri ATCC 49030, Streptomyces scabiei 8722, Acidothermus cellulolyticus 11B
YHALW	4,92E-03	3,04E-05	O: Methylotenera versatilis 301, Methylotenera versatilis 301, Bordetella petrii DSM 12804, Methylobacter tundripaludum SV96, Alicycliphilus denitrificans K601, Methylomonas methanica MC09, Novosphingobium aromaticivorans DSM 12444, Novosphingobium aromaticivorans DSM 12444, Marinomonas mediterranea MMB-1, Methylovorus sp MP688, Beijerinckia indica subsp indica, Acidovorax ebreus TPSY, Novosphingobium nitrogenifigens DSM 19370, Novosphingobium nitrogenifigens DSM 19370, Acidovorax delafieldii 2AN, Methylotenera mobilis JLW8, Curvibacter putative symbiont of, Bacteriovorax marinus SJ, Methylomonas sp 16a
YYAFW	9,49E+00	3,04E-05	O: Arthrobacter aurescens TC1, Streptomyces sp AA4, Arthrobacter sp FB24, Streptomyces clavuligerus ATCC 27064, Arthrobacter chlorophenolicus A6, Kribbella flavida DSM 17836, Streptomyces sp Mg1, Dermacoccus sp Ellin185, Streptomyces albus J1074, Kitasatospora setae KM-6054, Streptomyces sp SirexAA-E, Streptomyces griseus XylebKG-1, Streptosporangium roseum DSM 43021, Amycolicicoccus subflavus DQS3-9A1, Streptomyces avermitilis MA-4680 =, Actinosynnema mirum DSM 43827, Arthrobacter phenanthrenivorans Sphe3, Streptomyces sp SPB74, Microlunatus phosphovorus NM-1, Streptomyces pratensis ATCC 33331, Arthrobacter globiformis NBRC 12137, Streptomyces coelicolor A3(2), Streptomyces sp C, Amycolatopsis mediterranei S699

YFTQW	2,29E-04	5,27E-04	O: Bacillus clausii KSM-K16, Bacillus megaterium QM B1551, Bacillus cytotoxicus NVH 391-98, Bacillus cereus Rock3-44, Geobacillus sp WCH70, Bacillus thuringiensis serovar huazhongensis
YFQQV	6,91E-04	2,28E-04	N: Plesiocystis pacifica SIR-1, Plesiocystis pacifica SIR-1, Cyanothece sp PCC 8802
YFLKW	4,62E+00	3,04E-05	P: Novosphingobium aromaticivorans DSM 12444, Sphingomonas wittichii RW1, Sphingomonas wittichii RW1, Afipia sp 1NLS2
LFFLL	3,54E+03	1,87E-01	N: Shewanella denitrificans OS217, Marinobacter manganoxydans MnI7-9, Marinobacter adhaerens HP15, Marinobacter sp ELB17, Marinobacter algicola DG893, Halomonas sp HAL1
YFLVW	5,24E-02	5,27E-04	P: Alicycliphilus denitrificans K601, Acidovorax avenae subsp avenae, Polaromonas naphthalenivorans CJ2, Polaromonas naphthalenivorans CJ2, Burkholderia cenocepacia PC184, Lautropia mirabilis ATCC 51599, Roseibium sp TrichSKD4, Burkholderia gladioli BSR3, Burkholderia ambifaria MEX-5, Mucilaginibacter paludis DSM 18603, Burkholderia multivorans ATCC 17616, Sphingobacterium sp 21, Bordetella bronchiseptica RB50, Nitrosomonas eutropha C91, Nitrosomonas eutropha C91, Acidovorax citrulli AAC00-1, Hahella chejuensis KCTC 2396, Azoarcus sp BH72, Burkholderia glumae BGR1, Burkholderia vietnamiensis G4, Burkholderia sp, Acidovorax sp NO-1
LFLLL	1,44E+01	1,87E-01	N: Pseudomonas psychrotolerans L19, Haliscomenobacter hydrossis DSM 1100, Pseudomonas syringae Cit 7, Pseudomonas syringae pv oryzae, Pseudomonas syringae pv maculicola, Idiomarina sp A28L, Parvularcula bermudensis HTCC2503, Pseudomonas fulva 12-X, Pseudomonas savastanoi pv savastanoi, Pseudomonas fluorescens SBW25, Xanthomonas oryzae pv oryzae, Xanthomonas albilineans GPE PC73, Pseudomonas mendocina ymp, Salinisphaera shabanensis E1L3A, Xanthomonas gardneri ATCC 19865, Xanthomonas campestris pv campestris, Pseudomonas syringae pv aptata, Pseudomonas syringae pv morsprunorum, Pseudomonas syringae pv tomato, Xanthomonas vesicatoria ATCC 35937, Ricinus communis, Xanthomonas citri pv mangiferaeindicae
YFLQV	3,26E-01	9,85E-05	N: Mycobacterium intracellulare MOTT-64, Mycobacterium parascrofulaceum ATCC BAA-614, Mycobacterium bovis BCG str, Mycobacterium colombiense CECT 3035, Mycobacterium tuberculosis, Mycobacterium ulcerans Agy99
YHSLW	3,96E-03	3,04E-05	O: gamma proteobacterium HTCC5015, Pseudoalteromonas haloplanktis TAC125, Pseudoalteromonas haloplanktis TAC125, Shewanella pealeana ATCC 700345, Shewanella woodyi ATCC 51908, Shewanella loihica PV-4, Thiobacillus denitrificans ATCC 25259, Shewanella denitrificans OS217, Shewanella frigidimarina NCIMB 400, Gemmatimonas aurantiaca T-27, Shewanella sp MR-7, gamma proteobacterium IMCC2047, Pseudoalteromonas sp BSi20429, Nitrosospira multiformis ATCC 25196, Shewanella oneidensis MR-1, Shewanella amazonensis SB2B, Bermanella marisrubri, Shewanella sp HN-41, Acidovorax sp NO-1, Acidovorax sp NO-1, Saccharophagus degradans 2-40, Shewanella piezotolerans WP3, Psychromonas ingrahamii 37, Shewanella baltica OS625
YFAQW	5,79E-02	5,27E-04	O: Vitis vinifera, Paenibacillus mucilaginosus KNP414, Paenibacillus dendritiformis C454, Carica papaya, Arabidopsis thaliana, Zea mays, Medicago truncatula, Medicago truncatula, Wolffia australiana, Physcomitrella patens, Physcomitrella patens, Arabidopsis lyrata subsp lyrata, Desmospora sp 8437, Selaginella moellendorffii, Picea sitchensis, Triticum aestivum, Paenibacillus sp oral taxon, Datisca glomerata, Populus trichocarpa, Eutrema halophilum, Thalassiosira pseudonana, Oryza sativa Japonica Group, Glycine max, Glycine max, Paenibacillus larvae subsp larvae, Paenibacillus sp Aloe-11, Ricinus communis, Phaeodactylum tricornutum CCAP 10551, Paenibacillus lactis 154

Aminoácidos claves del portal E7 B10-CD1-E7-E11-G8	k_{E7G}	**Especies**
YFTQ	1,22E-03	O: Bacillus clausii KSM-K16, Bacillus megaterium QM B1551, Bacillus cytotoxicus NVH 391-98, Bacillus cereus Rock3-44, Geobacillus sp WCH70, Bacillus thuringiensis serovar huazhongensis
YFLK	1,18E-11	P: Novosphingobium aromaticivorans DSM 12444, Sphingomonas wittichii RW1, Sphingomonas wittichii RW1, Afipia sp 1NLS2

YFLV	1,18E-11	P: Alicycliphilus denitrificans K601, Acidovorax avenae subsp avenae, Polaromonas naphthalenivorans CJ2, Polaromonas naphthalenivorans CJ2, Burkholderia cenocepacia PC184, Lautropia mirabilis ATCC 51599, Roseibium sp TrichSKD4, Burkholderia gladioli BSR3, Burkholderia ambifaria MEX-5, Mucilaginibacter paludis DSM 18603, Burkholderia multivorans ATCC 17616, Sphingobacterium sp 21, Bordetella bronchiseptica RB50, Nitrosomonas eutropha C91, Nitrosomonas eutropha C91, Acidovorax citrulli AAC00-1, Hahella chejuensis KCTC 2396, Azoarcus sp BH72, Burkholderia glumae BGR1, Burkholderia vietnamiensis G4, Burkholderia sp, Acidovorax sp NO-1
YFLQ	4,00E-21	N: Mycobacterium intracellulare MOTT-64, Mycobacterium parascrofulaceum ATCC BAA-614, Mycobacterium bovis BCG str, Mycobacterium colombiense CECT 3035, Mycobacterium tuberculosis, Mycobacterium ulcerans Agy99
LFFF	2,69E-15	N: Halomonas sp TD01, Congregibacter litoralis KT71
YHSL	6,48E-03	O: gamma proteobacterium HTCC5015, Pseudoalteromonas haloplanktis TAC125, Pseudoalteromonas haloplanktis TAC125, Shewanella pealeana ATCC 700345, Shewanella woodyi ATCC 51908, Shewanella loihica PV-4, Thiobacillus denitrificans ATCC 25259, Shewanella denitrificans OS217, Shewanella frigidimarina NCIMB 400, Gemmatimonas aurantiaca T-27, Shewanella sp MR-7, gamma proteobacterium IMCC2047, Pseudoalteromonas sp BSi20429, Nitrosospira multiformis ATCC 25196, Shewanella oneidensis MR-1, Shewanella amazonensis SB2B, Bermanella marisrubri, Shewanella sp HN-41, Acidovorax sp NO-1, Acidovorax sp NO-1, Saccharophagus degradans 2-40, Shewanella piezotolerans WP3, Psychromonas ingrahamii 37, Shewanella baltica OS625
YFAQ	9,04E-03	O: Vitis vinifera, Paenibacillus mucilaginosus KNP414, Paenibacillus dendritiformis C454, Carica papaya, Arabidopsis thaliana, Zea mays, Medicago truncatula, Medicago truncatula, Wolffia australiana, Physcomitrella patens, Physcomitrella patens, Arabidopsis lyrata subsp lyrata, Desmospora sp 8437, Selaginella moellendorffii, Picea sitchensis, Triticum aestivum, Paenibacillus sp oral taxon, Datisca glomerata, Populus trichocarpa, Eutrema halophilum, Thalassiosira pseudonana, Oryza sativa Japonica Group, Glycine max, Glycine max, Paenibacillus larvae subsp larvae, Paenibacillus sp Aloe-11, Ricinus communis, Phaeodactylum tricornutum CCAP 10551, Paenibacillus lactis 154
LFFL	2,69E-15	N: Shewanella denitrificans OS217, Marinobacter manganoxydans MnI7-9, Marinobacter adhaerens HP15, Marinobacter sp ELB17, Marinobacter algicola DG893, Halomonas sp HAL1
YFMK	2,69E-15	P: Rhizobium leguminosarum bv trifolii, Agrobacterium sp ATCC 31749, Agrobacterium tumefaciens F2, Sinorhizobium medicae WSM419, Agrobacterium tumefaciens CCNWGS0286, Sinorhizobium fredii NGR234, Rhizobium sp PDO1-076, Rhizobium etli CFN 42, Rhizobium etli CIAT 652, Rhizobium leguminosarum bv viciae, Sinorhizobium meliloti CCNWSX0020
LFLL	2,78E-07	N: Pseudomonas psychrotolerans L19, Haliscomenobacter hydrossis DSM 1100, Pseudomonas syringae Cit 7, Pseudomonas syringae pv oryzae, Pseudomonas syringae pv maculicola, Idiomarina sp A28L, Parvularcula bermudensis HTCC2503, Pseudomonas fulva 12-X, Pseudomonas savastanoi pv savastanoi, Pseudomonas fluorescens SBW25, Xanthomonas oryzae pv oryzae, Xanthomonas albilineans GPE PC73, Pseudomonas mendocina ymp, Salinisphaera shabanensis E1L3A, Xanthomonas gardneri ATCC 19865, Xanthomonas campestris pv campestris, Pseudomonas syringae pv aptata, Pseudomonas syringae pv morsprunorum, Pseudomonas syringae pv tomato, Xanthomonas vesicatoria ATCC 35937, Ricinus communis, Xanthomonas citri pv mangiferaeindicae
YHML	2,28E-04	O: Sphingobium sp SYK-6, marine gamma proteobacterium HTCC2148
VFLL	1,18E-11	N: Acidithiobacillus caldus SM-1, Limnobacter sp MED105, Halothiobacillus neapolitanus c2, Oxalobacteraceae bacterium IMCC9480, Burkholderiales bacterium JOSHI_001, Cupriavidus necator N-1, Candidatus Nitrospira defluvii, Candidatus Nitrospira defluvii, Candidatus Nitrospira defluvii, Candidatus Nitrospira defluvii, Acidovorax sp NO-1
YHAL	6,48E-03	O: Methylotenera versatilis 301, Methylotenera versatilis 301, Bordetella petrii DSM 12804, Methylobacter tundripaludum SV96, Alicycliphilus denitrificans K601, Methylomonas methanica MC09, Novosphingobium aromaticivorans DSM 12444, Novosphingobium aromaticivorans DSM 12444, Marinomonas mediterranea MMB-1, Methylovorus sp MP688, Beijerinckia indica subsp indica, Acidovorax ebreus TPSY, Novosphingobium nitrogenifigens DSM 19370, Novosphingobium nitrogenifigens DSM 19370, Acidovorax delafieldii 2AN, Methylotenera mobilis JLW8, Curvibacter putative symbiont of, Bacteriovorax marinus SJ, Methylomonas sp 16a

YFKV	7,95E-06	P: Leadbetterella byssophila DSM 17132, Leadbetterella byssophila DSM 17132
YHTL	7,95E-06	O: Xanthobacter autotrophicus Py2, marine gamma proteobacterium HTCC2143
YYAF	6,48E-03	O: Arthrobacter aurescens TC1, Streptomyces sp AA4, Arthrobacter sp FB24, Streptomyces clavuligerus ATCC 27064, Arthrobacter chlorophenolicus A6, Kribbella flavida DSM 17836, Streptomyces sp Mg1, Dermacoccus sp Ellin185, Streptomyces albus J1074, Kitasatospora setae KM-6054, Streptomyces sp SirexAA-E, Streptomyces griseus XylebKG-1, Streptosporangium roseum DSM 43021, Amycolicicoccus subflavus DQS3-9A1, Streptomyces avermitilis MA-4680 =, Actinosynnema mirum DSM 43827, Arthrobacter phenanthrenivorans Sphe3, Streptomyces sp SPB74, Microlunatus phosphovorus NM-1, Streptomyces pratensis ATCC 33331, Arthrobacter globiformis NBRC 12137, Streptomyces coelicolor A3(2), Streptomyces sp C, Amycolatopsis mediterranei S699
YYAL	9,04E-03	O: Mycobacterium tuberculosis NCGM2209, Clavibacter michiganensis subsp sepedonicus, Micrococcus luteus SK58, Corynebacterium efficiens YS-314, Streptomyces sviceus ATCC 29083, Microbacterium laevaniformans OR221, Mycobacterium rhodesiae NBB3, Leifsonia xyli subsp xyli, Actinoplanes sp SE50110, Mycobacterium intracellulare MOTT-64, Cellulomonas flavigena DSM 20109, Corynebacterium glutamicum ATCC 14067, Streptomyces venezuelae ATCC 10712, Blastococcus saxobsidens DD2, Xylanimonas cellulosilytica DSM 15894, Gordonia effusa NBRC 100432, Janibacter sp HTCC2649, Streptomyces zinciresistens K42, Beutenbergia cavernae DSM 12333, Frankia alni ACN14a, Segniliparus rotundus DSM 44985, Mycobacterium tusciae JS617, Corynebacterium amycolatum SK46, Corynebacterium variabile DSM 44702, Corynebacterium accolens ATCC 49725, Arthrobacter arilaitensis Re117, Saccharomonospora marina XMU15, Mycobacterium parascrofulaceum ATCC BAA-614, Mycobacterium avium 104, Streptomyces ghanaensis ATCC 14672, Mycobacterium vanbaalenii PYR-1, Nocardia farcinica IFM 10152, Corynebacterium diphtheriae NCTC 13129, Saccharomonospora cyanea NA-134, Brachybacterium faecium DSM 4810, Mycobacterium colombiense CECT 3035, Corynebacterium lipophiloflavum DSM 44291, Corynebacterium resistens DSM 45100, Frankia sp CN3, Streptomyces griseoaurantiacus M045, Mycobacterium abscessus subsp bolletii, Nakamurella multipartita DSM 44233, marine actinobacterium PHSC20C1, Isoptericola variabilis 225, Jonesia denitrificans DSM 20603, Frankia sp EAN1pec, Thermobifida fusca, Corynebacterium ammoniagenes DSM 20306, Dietzia cinnamea P4, Algoriphagus machipongonensis, Corynebacterium casei UCMA 3821, Mycobacterium gilvum PYR-GCK, Streptomyces cattleya NRRL 8057, Gordonia araii NBRC 100433, Saccharomonospora paurometabolica YIM 90007, Streptomyces griseoflavus Tu4000, Streptomyces viridochromogenes DSM 40736, Corynebacterium pseudogenitalium ATCC 33035, Micromonospora sp ATCC 39149, Corynebacterium glutamicum ATCC 13032, Stackebrandtia nassauensis DSM 44728, Saccharomonospora glauca K62, Frankia sp CcI3, Mycobacterium rhodesiae JS60, Streptomyces hygroscopicus subsp jinggangensis, Mycobacterium thermoresistibile ATCC 19527, Gordonia sputi NBRC 100414, Mycobacterium leprae, Gordonia otitidis NBRC 100426, Nocardia brasiliensis ATCC 700358, Mycobacterium sp MCS, Kocuria rhizophila DC2201, Corynebacterium striatum ATCC 6940, Corynebacterium jeikeium K411, Corynebacterium glucuronolyticum ATCC 51866, Micromonospora aurantiaca ATCC 27029, Streptomyces violaceusniger Tu 4113, Microbacterium testaceum StLB037, Streptomyces sp SPB78, Mycobacterium marinum M, Frankia sp EuI1c, Saccharopolyspora erythraea NRRL 2338, Saccharomonospora viridis DSM 43017, Streptomyces himastatinicus ATCC 53653, Gordonia alkanivorans NBRC 16433, Corynebacterium aurimucosum ATCC 700975, Salinispora tropica CNB-440, Mycobacterium smegmatis str MC2, Streptomyces bingchenggensis BCW-1, Saccharomonospora azurea NA-128, Mycobacterium sp JDM601, Streptomyces sp e14, Frankia sp EUN1f, Gordonia neofelifaecis NRRL B-59395, Kytococcus sedentarius DSM 20547, Gordonia bronchialis DSM 43247, Corynebacterium ulcerans 809, Salinispora arenicola CNS-205, Cellulomonas fimi ATCC 484, Nocardiopsis dassonvillei subsp dassonvillei, Catenulispora acidiphila DSM 44928, Nocardia cyriacigeorgica GUH-2, Gordonia polyisoprenivorans VH2, Thermomonospora curvata DSM 43183, Geodermatophilus obscurus DSM 43160, Mobilicoccus pelagius NBRC 104925, Corynebacterium pseudotuberculosis 106-A, Planctomyces limnophilus DSM 3776, Frankia symbiont of Datisca, Corynebacterium genitalium ATCC 33030, Brevibacterium mcbrellneri ATCC 49030, Streptomyces scabiei 8722, Acidothermus cellulolyticus 11B
YFQQ	1,15E-19	N: Plesiocystis pacifica SIR-1, Plesiocystis pacifica SIR-1, Cyanothece sp PCC 8802

Aminoácidos claves del canal corto G8 H9-G8-G9	k_{STG8}	Especies
FWL	2,69E-15	O: Vitis vinifera, Janibacter sp HTCC2649, Arthrobacter arilaitensis Re117, Carica papaya, Arabidopsis thaliana, Nitrosospira multiformis ATCC 25196, Algoriphagus machipongonensis, Zea mays, Medicago truncatula, Medicago truncatula, Wolffia australiana, Beijerinckia indica subsp indica, Physcomitrella patens, Physcomitrella patens, Arabidopsis lyrata subsp lyrata, Selaginella moellendorffii, Picea sitchensis, Triticum aestivum, Kocuria rhizophila DC2201, Corynebacterium glucuronolyticum ATCC 51866, Saccharopolyspora erythraea NRRL 2338, Datisca glomerata, Populus trichocarpa, Eutrema halophilum, Oryza sativa Japonica Group, Glycine max, Glycine max, Kytococcus sedentarius DSM 20547, Curvibacter putative symbiont of, Bacteriovorax marinus SJ, Acidovorax sp NO-1, Acidovorax sp NO-1, Ricinus communis, Brevibacterium mcbrellneri ATCC 49030, Acidothermus cellulolyticus 11B
LVV	1,22E-03	N: Mycobacterium intracellulare MOTT-64, Mycobacterium parascrofulaceum ATCC BAA-614, Mycobacterium bovis BCG str, Mycobacterium colombiense CECT 3035, Halomonas sp TD01, Congregibacter litoralis KT71, Mycobacterium tuberculosis
AWV	2,69E-15	P: Acidovorax avenae subsp avenae, Acidovorax citrulli AAC00-1, Acidovorax sp NO-1
LLV	2,28E-04	N: Acidithiobacillus caldus SM-1, Limnobacter sp MED105, Halothiobacillus neapolitanus c2, Pseudomonas psychrotolerans L19, Haliscomenobacter hydrossis DSM 1100, Pseudomonas syringae Cit 7, Shewanella denitrificans OS217, Pseudomonas syringae pv oryzae, Oxalobacteraceae bacterium IMCC9480, Marinobacter manganoxydans MnI7-9, Pseudomonas syringae pv maculicola, Idiomarina sp A28L, Parvularcula bermudensis HTCC2503, Pseudomonas fulva 12-X, Marinobacter adhaerens HP15, Pseudomonas savastanoi pv savastanoi, Burkholderiales bacterium JOSHI_001, Pseudomonas fluorescens SBW25, Xanthomonas oryzae pv oryzae, Xanthomonas albilineans GPE PC73, Pseudomonas mendocina ymp, Salinisphaera shabanensis E1L3A, Cupriavidus necator N-1, Marinobacter sp ELB17, Marinobacter algicola DG893, Xanthomonas gardneri ATCC 19865, Xanthomonas campestris pv campestris, Candidatus Nitrospira defluvii, Candidatus Nitrospira defluvii, Candidatus Nitrospira defluvii, Candidatus Nitrospira defluvii, Pseudomonas syringae pv aptata, Pseudomonas syringae pv morsprunorum, Pseudomonas syringae pv tomato, Acidovorax sp NO-1, Xanthomonas vesicatoria ATCC 35937, Ricinus communis, Halomonas sp HAL1, Xanthomonas citri pv mangiferaeindicae
AWL	2,69E-15	O: Novosphingobium aromaticivorans DSM 12444, Novosphingobium aromaticivorans DSM 12444, Sphingobium sp SYK-6, Novosphingobium nitrogenifigens DSM 19370, Novosphingobium nitrogenifigens DSM 19370
AWL	2,69E-15	P: Rhizobium leguminosarum bv trifolii, Agrobacterium sp ATCC 31749, Agrobacterium tumefaciens F2, Sinorhizobium medicae WSM419, Agrobacterium tumefaciens CCNWGS0286, Alicycliphilus denitrificans K601, Novosphingobium aromaticivorans DSM 12444, Polaromonas naphthalenivorans CJ2, Polaromonas naphthalenivorans CJ2, Sinorhizobium fredii NGR234, Burkholderia cenocepacia PC184, Lautropia mirabilis ATCC 51599, Roseibium sp TrichSKD4, Rhizobium sp PDO1-076, Burkholderia gladioli BSR3, Burkholderia ambifaria MEX-5, Mucilaginibacter paludis DSM 18603, Burkholderia multivorans ATCC 17616, Sphingobacterium sp 21, Leadbetterella byssophila DSM 17132, Leadbetterella byssophila DSM 17132, Bordetella bronchiseptica RB50, Nitrosomonas eutropha C91, Nitrosomonas eutropha C91, Rhizobium etli CFN 42, Hahella chejuensis KCTC 2396, Azoarcus sp BH72, Burkholderia glumae BGR1, Rhizobium etli CIAT 652, Burkholderia vietnamiensis G4, Sphingomonas wittichii RW1, Sphingomonas wittichii RW1, Afipia sp 1NLS2, Rhizobium leguminosarum bv viciae, Burkholderia sp, Sinorhizobium meliloti CCNWSX0020

LWL	2,69E-15	O: Mycobacterium tuberculosis NCGM2209, Clavibacter michiganensis subsp sepedonicus, Arthrobacter aurescens TC1, Micrococcus luteus SK58, Corynebacterium efficiens YS-314, Streptomyces sviceus ATCC 29083, Microbacterium laevaniformans OR221, Streptomyces sp AA4, Arthrobacter sp FB24, Methylotenera versatilis 301, Methylotenera versatilis 301, gamma proteobacterium HTCC5015, Mycobacterium rhodesiae NBB3, Leifsonia xyli subsp xyli, Methylobacter tundripaludum SV96, Paenibacillus mucilaginosus KNP414, Actinoplanes sp SE50110, Paenibacillus dendritiformis C454, Pseudoalteromonas haloplanktis TAC125, Pseudoalteromonas haloplanktis TAC125, Shewanella pealeana ATCC 700345, Streptomyces clavuligerus ATCC 27064, Mycobacterium intracellulare MOTT-64, Cellulomonas flavigena DSM 20109, Corynebacterium glutamicum ATCC 14067, Streptomyces venezuelae ATCC 10712, Methylomonas methanica MC09, Blastococcus saxobsidens DD2, Xylanimonas cellulosilytica DSM 15894, Arthrobacter chlorophenolicus A6, Gordonia effusa NBRC 100432, Kribbella flavida DSM 17836, Marinomonas mediterranea MMB-1, Streptomyces zinciresistens K42, Beutenbergia cavernae DSM 12333, Shewanella woodyi ATCC 51908, Frankia alni ACN14a, Segniliparus rotundus DSM 44985, Mycobacterium tusciae JS617, Corynebacterium amycolatum SK46, Corynebacterium variabile DSM 44702, Corynebacterium accolens ATCC 49725, Streptomyces sp Mg1, Shewanella loihica PV-4, Thiobacillus denitrificans ATCC 25259, Methylovorus sp MP688, Saccharomonospora marina XMU15, Mycobacterium parascrofulaceum ATCC BAA-614, Mycobacterium avium 104, Streptomyces ghanaensis ATCC 14672, Shewanella denitrificans OS217, Mycobacterium vanbaalenii PYR-1, Dermacoccus sp Ellin185, Nocardia farcinica IFM 10152, Corynebacterium diphtheriae NCTC 13129, Saccharomonospora cyanea NA-134, Shewanella frigidimarina NCIMB 400, Gemmatimonas aurantiaca T-27, Streptomyces albus J1074, Kitasatospora setae KM-6054, Brachybacterium faecium DSM 4810, Mycobacterium colombiense CECT 3035, Corynebacterium lipophiloflavum DSM 44291, Streptomyces sp SirexAA-E, Corynebacterium resistens DSM 45100, Streptomyces griseus XylebKG-1, Frankia sp CN3, Streptomyces griseoaurantiacus M045, Xanthobacter autotrophicus Py2, Mycobacterium abscessus subsp bolletii, Nakamurella multipartita DSM 44233, Shewanella sp MR-7, Streptosporangium roseum DSM 43021, marine actinobacterium PHSC20C1, Isoptericola variabilis 225, gamma proteobacterium IMCC2047, Pseudoalteromonas sp BSi20429, Jonesia denitrificans DSM 20603, Frankia sp EAN1pec, Thermobifida fusca, Corynebacterium ammoniagenes DSM 20306, Dietzia cinnamea P4, Bacillus clausii KSM-K16, Corynebacterium casei UCMA 3821, Amycolicicoccus subflavus DQS3-9A1, Mycobacterium gilvum PYR-GCK, Shewanella oneidensis MR-1, Streptomyces cattleya NRRL 8057, Streptomyces avermitilis MA-4680 =, Gordonia araii NBRC 100433, Saccharomonospora paurometabolica YIM 90007, Streptomyces griseoflavus Tu4000, Actinosynnema mirum DSM 43827, Streptomyces viridochromogenes DSM 40736, Corynebacterium pseudogenitalium ATCC 33035, Micromonospora sp ATCC 39149, Corynebacterium glutamicum ATCC 13032, Stackebrandtia nassauensis DSM 44728, Saccharomonospora glauca K62, Desmospora sp 8437, Shewanella amazonensis SB2B, Frankia sp CcI3, Arthrobacter phenanthrenivorans Sphe3, Mycobacterium rhodesiae JS60, Streptomyces hygroscopicus subsp jinggangensis, Mycobacterium thermoresistibile ATCC 19527, Gordonia sputi NBRC 100414, Bermanella marisrubri, Mycobacterium leprae, Gordonia otitidis NBRC 100426, Nocardia brasiliensis ATCC 700358, Mycobacterium sp MCS, Corynebacterium striatum ATCC 6940, Corynebacterium jeikeium K411, Micromonospora aurantiaca ATCC 27029, Streptomyces sp SPB74, Streptomyces violaceusniger Tu 4113, Microbacterium testaceum StLB037, Streptomyces sp SPB78, Mycobacterium marinum M, Frankia sp EuI1c, Microlunatus phosphovorus NM-1, Paenibacillus sp oral taxon, Shewanella sp HN-41, Saccharomonospora viridis DSM 43017, Streptomyces himastatinicus ATCC 53653, Bacillus megaterium QM B1551, Gordonia alkanivorans NBRC 16433, Corynebacterium aurimucosum ATCC 700975, Salinispora tropica CNB-440, Mycobacterium smegmatis str MC2, Bacillus cytotoxicus NVH 391-98, Streptomyces bingchenggensis BCW-1, Saccharomonospora azurea NA-128, Streptomyces pratensis ATCC 33331, Mycobacterium sp JDM601, Bacillus cereus Rock3-44, Streptomyces sp e14, Frankia sp EUN1f, Gordonia neofelifaecis NRRL B-59395, Paenibacillus larvae subsp larvae, Gordonia bronchialis DSM 43247, Corynebacterium ulcerans 809, Salinispora arenicola CNS-205, Methylotenera mobilis JLW8, Cellulomonas fimi ATCC 484, Nocardiopsis dassonvillei subsp dassonvillei, Arthrobacter globiformis NBRC 12137, Streptomyces coelicolor A3(2), Streptomyces sp C, marine gamma proteobacterium HTCC2148, marine gamma proteobacterium HTCC2143, Catenulispora acidiphila DSM 44928, Paenibacillus sp Aloe-11, Nocardia cyriacigeorgica GUH-2, Saccharophagus degradans 2-40, Gordonia polyisoprenivorans VH2, Thermomonospora curvata DSM 43183, Geodermatophilus obscurus DSM 43160, Mobilicoccus pelagius NBRC 104925, Geobacillus sp WCH70, Shewanella piezotolerans WP3, Bacillus thuringiensis serovar huazhongensis, Methylomonas sp 16a, Corynebacterium pseudotuberculosis 106-A, Amycolatopsis mediterranei S699, Planctomyces limnophilus DSM 3776, Psychromonas ingrahamii 37, Frankia symbiont of Datisca, Corynebacterium genitalium ATCC 33030, Shewanella baltica OS625, Streptomyces scabiei 8722, Paenibacillus lactis 154
FWV	2,69E-15	O: Bordetella petrii DSM 12804, Alicycliphilus denitrificans K601, Acidovorax ebreus TPSY, Thalassiosira pseudonana, Acidovorax delafieldii 2AN, Phaeodactylum tricornutum CCAP 10551
AVV	1,57E-01	N: Plesiocystis pacifica SIR-1, Plesiocystis pacifica SIR-1, Cyanothece sp PCC 8802, Mycobacterium ulcerans Agy99

Aminoácidos claves del canal largo H5-B2-H9-E15-E11-G8	k_{LT}	Especies
LFFLLW	2,21E-12	O: Janibacter sp HTCC2649, Arthrobacter arilaitensis Re117, Kocuria rhizophila DC2201, Corynebacterium glucuronolyticum ATCC 51866, Saccharopolyspora erythraea NRRL 2338, Kytococcus sedentarius DSM 20547, Brevibacterium mcbrellneri ATCC 49030, Acidothermus cellulolyticus 11B
LILLQW	6,48E-03	O: Paenibacillus mucilaginosus KNP414, Paenibacillus dendritiformis C454, Bacillus clausii KSM-K16, Desmospora sp 8437, Paenibacillus sp oral taxon, Bacillus megaterium QM B1551, Bacillus cytotoxicus NVH 391-98, Bacillus cereus Rock3-44, Paenibacillus larvae subsp larvae, Paenibacillus sp Aloe-11, Geobacillus sp WCH70, Bacillus thuringiensis serovar huazhongensis, Paenibacillus lactis 154
LFFLQW	4,26E-05	O: Zea mays, Medicago truncatula, Medicago truncatula, Triticum aestivum, Thalassiosira pseudonana, Oryza sativa Japonica Group
LIAFQV	3,37E-02	N: Mycobacterium ulcerans Agy99
LFLLFW	6,33E-11	O: Arthrobacter aurescens TC1, Streptomyces sp AA4, Arthrobacter sp FB24, Streptomyces clavuligerus ATCC 27064, Arthrobacter chlorophenolicus A6, Kribbella flavida DSM 17836, Streptomyces sp Mg1, Dermacoccus sp Ellin185, Streptomyces albus J1074, Kitasatospora setae KM-6054, Streptomyces sp SirexAA-E, Streptomyces griseus XylebKG-1, Streptosporangium roseum DSM 43021, Amycolicicoccus subflavus DQS3-9A1, Streptomyces avermitilis MA-4680 =, Actinosynnema mirum DSM 43827, Arthrobacter phenanthrenivorans Sphe3, Streptomyces sp SPB74, Microlunatus phosphovorus NM-1, Streptomyces pratensis ATCC 33331, Arthrobacter globiformis NBRC 12137, Streptomyces coelicolor A3(2), Streptomyces sp C, Amycolatopsis mediterranei S699
LIAEVW	1,22E-03	P: Roseibium sp TrichSKD4, Mucilaginibacter paludis DSM 18603, Sphingobacterium sp 21, Leadbetterella byssophila DSM 17132, Leadbetterella byssophila DSM 17132, Hahella chejuensis KCTC 2396
LILLLL	1,22E-03	N: Haliscomenobacter hydrossis DSM 1100, Pseudomonas syringae Cit 7, Shewanella denitrificans OS217, Pseudomonas syringae pv oryzae, Marinobacter manganoxydans MnI7-9, Pseudomonas syringae pv maculicola, Idiomarina sp A28L, Parvularcula bermudensis HTCC2503, Marinobacter adhaerens HP15, Pseudomonas savastanoi pv savastanoi, Pseudomonas mendocina ymp, Salinisphaera shabanensis E1L3A, Cupriavidus necator N-1, Marinobacter sp ELB17, Marinobacter algicola DG893, Candidatus Nitrospira defluvii, Candidatus Nitrospira defluvii, Pseudomonas syringae pv aptata, Pseudomonas syringae pv morsprunorum, Pseudomonas syringae pv tomato, Acidovorax sp NO-1, Halomonas sp HAL1
LILLLW	1,49E-06	O: Methylotenera versatilis 301, Methylotenera versatilis 301, gamma proteobacterium HTCC5015, Methylobacter tundripaludum SV96, Pseudoalteromonas haloplanktis TAC125, Pseudoalteromonas haloplanktis TAC125, Shewanella pealeana ATCC 700345, Methylomonas methanica MC09, Marinomonas mediterranea MMB-1, Shewanella woodyi ATCC 51908, Shewanella loihica PV-4, Thiobacillus denitrificans ATCC 25259, Methylovorus sp MP688, Shewanella denitrificans OS217, Shewanella frigidimarina NCIMB 400, Gemmatimonas aurantiaca T-27, Xanthobacter autotrophicus Py2, Shewanella sp MR-7, gamma proteobacterium IMCC2047, Pseudoalteromonas sp BSi20429, Shewanella oneidensis MR-1, Shewanella amazonensis SB2B, Bermanella marisrubri, Shewanella sp HN-41, Methylotenera mobilis JLW8, marine gamma proteobacterium HTCC2148, marine gamma proteobacterium HTCC2143, Saccharophagus degradans 2-40, Shewanella piezotolerans WP3, Methylomonas sp 16a, Planctomyces limnophilus DSM 3776, Psychromonas ingrahamii 37, Shewanella baltica OS625
LILSLL	1,49E-06	N: Acidithiobacillus caldus SM-1, Limnobacter sp MED105, Halothiobacillus neapolitanus c2, Burkholderiales bacterium JOSHI_001, Candidatus Nitrospira defluvii, Candidatus Nitrospira defluvii

LIALLW	4,26E-05	O: Novosphingobium aromaticivorans DSM 12444, Novosphingobium aromaticivorans DSM 12444, Sphingobium sp SYK-6, Novosphingobium nitrogenifigens DSM 19370, Novosphingobium nitrogenifigens DSM 19370
LFLLLW	1,49E-06	O: Mycobacterium tuberculosis NCGM2209, Clavibacter michiganensis subsp sepedonicus, Micrococcus luteus SK58, Corynebacterium efficiens YS-314, Streptomyces sviceus ATCC 29083, Microbacterium laevaniformans OR221, Mycobacterium rhodesiae NBB3, Leifsonia xyli subsp xyli, Actinoplanes sp SE50110, Mycobacterium intracellulare MOTT-64, Cellulomonas flavigena DSM 20109, Corynebacterium glutamicum ATCC 14067, Streptomyces venezuelae ATCC 10712, Blastococcus saxobsidens DD2, Xylanimonas cellulosilytica DSM 15894, Gordonia effusa NBRC 100432, Streptomyces zinciresistens K42, Beutenbergia cavernae DSM 12333, Frankia alni ACN14a, Segniliparus rotundus DSM 44985, Mycobacterium tusciae JS617, Corynebacterium amycolatum SK46, Corynebacterium variabile DSM 44702, Corynebacterium accolens ATCC 49725, Saccharomonospora marina XMU15, Mycobacterium parascrofulaceum ATCC BAA-614, Mycobacterium avium 104, Streptomyces ghanaensis ATCC 14672, Mycobacterium vanbaalenii PYR-1, Nocardia farcinica IFM 10152, Corynebacterium diphtheriae NCTC 13129, Saccharomonospora cyanea NA-134, Brachybacterium faecium DSM 4810, Mycobacterium colombiense CECT 3035, Corynebacterium lipophiloflavum DSM 44291, Corynebacterium resistens DSM 45100, Frankia sp CN3, Streptomyces griseoaurantiacus M045, Mycobacterium abscessus subsp bolletii, Nakamurella multipartita DSM 44233, marine actinobacterium PHSC20C1, Isoptericola variabilis 225, Jonesia denitrificans DSM 20603, Frankia sp EAN1pec, Thermobifida fusca, Corynebacterium ammoniagenes DSM 20306, Dietzia cinnamea P4, Corynebacterium casei UCMA 3821, Mycobacterium gilvum PYR-GCK, Streptomyces cattleya NRRL 8057, Gordonia araii NBRC 100433, Saccharomonospora paurometabolica YIM 90007, Streptomyces griseoflavus Tu4000, Streptomyces viridochromogenes DSM 40736, Corynebacterium pseudogenitalium ATCC 33035, Micromonospora sp ATCC 39149, Corynebacterium glutamicum ATCC 13032, Stackebrandtia nassauensis DSM 44728, Saccharomonospora glauca K62, Frankia sp CcI3, Mycobacterium rhodesiae JS60, Streptomyces hygroscopicus subsp jinggangensis, Mycobacterium thermoresistibile ATCC 19527, Gordonia sputi NBRC 100414, Mycobacterium leprae, Gordonia otitidis NBRC 100426, Nocardia brasiliensis ATCC 700358, Mycobacterium sp MCS, Corynebacterium striatum ATCC 6940, Corynebacterium jeikeium K411, Micromonospora aurantiaca ATCC 27029, Streptomyces violaceusniger Tu 4113, Microbacterium testaceum StLB037, Streptomyces sp SPB78, Mycobacterium marinum M, Frankia sp EuI1c, Saccharomonospora viridis DSM 43017, Streptomyces himastatinicus ATCC 53653, Gordonia alkanivorans NBRC 16433, Corynebacterium aurimucosum ATCC 700975, Salinispora tropica CNB-440, Mycobacterium smegmatis str MC2, Streptomyces bingchenggensis BCW-1, Saccharomonospora azurea NA-128, Mycobacterium sp JDM601, Streptomyces sp e14, Frankia sp EUN1f, Gordonia neofelifaecis NRRL B-59395, Gordonia bronchialis DSM 43247, Corynebacterium ulcerans 809, Salinispora arenicola CNS-205, Cellulomonas fimi ATCC 484, Nocardiopsis dassonvillei subsp dassonvillei, Catenulispora acidiphila DSM 44928, Nocardia cyriacigeorgica GUH-2, Gordonia polyisoprenivorans VH2, Thermomonospora curvata DSM 43183, Geodermatophilus obscurus DSM 43160, Mobilicoccus pelagius NBRC 104925, Corynebacterium pseudotuberculosis 106-A, Frankia symbiont of Datisca, Corynebacterium genitalium ATCC 33030, Streptomyces scabiei 8722
LILFLL	3,37E-02	N: Pseudomonas psychrotolerans L19, Oxalobacteraceae bacterium IMCC9480, Pseudomonas fulva 12-X, Pseudomonas fluorescens SBW25, Xanthomonas oryzae pv oryzae, Xanthomonas albilineans GPE PC73, Xanthomonas gardneri ATCC 19865, Xanthomonas campestris pv campestris, Xanthomonas vesicatoria ATCC 35937, Ricinus communis, Xanthomonas citri pv mangiferaeindicae
FIASVW	1,22E-03	P: Burkholderia cenocepacia PC184, Burkholderia gladioli BSR3, Burkholderia ambifaria MEX-5, Burkholderia multivorans ATCC 17616, Burkholderia glumae BGR1, Burkholderia vietnamiensis G4, Burkholderia sp
LIASVW	6,48E-03	P: Alicycliphilus denitrificans K601, Acidovorax avenae subsp avenae, Polaromonas naphthalenivorans CJ2, Polaromonas naphthalenivorans CJ2, Lautropia mirabilis ATCC 51599, Bordetella bronchiseptica RB50, Nitrosomonas eutropha C91, Nitrosomonas eutropha C91, Acidovorax citrulli AAC00-1, Azoarcus sp BH72, Acidovorax sp NO-1
LIALQV	6,48E-03	N: Plesiocystis pacifica SIR-1, Plesiocystis pacifica SIR-1, Cyanothece sp PCC 8802
LIFLLW	7,95E-06	O: Bordetella petrii DSM 12804, Alicycliphilus denitrificans K601, Nitrosospira multiformis ATCC 25196, Algoriphagus machipongonensis, Beijerinckia indica subsp indica, Acidovorax ebreus TPSY, Acidovorax delafieldii 2AN, Curvibacter putative symbiont of, Bacteriovorax marinus SJ, Acidovorax sp NO-1, Acidovorax sp NO-1

LILLFV	1,81E-09	N: Halomonas sp TD01, Congregibacter litoralis KT71
FIASKW	6,48E-03	P: Rhizobium leguminosarum bv trifolii, Agrobacterium sp ATCC 31749, Agrobacterium tumefaciens F2, Sinorhizobium medicae WSM419, Agrobacterium tumefaciens CCNWGS0286, Novosphingobium aromaticivorans DSM 12444, Sinorhizobium fredii NGR234, Rhizobium sp PDO1-076, Rhizobium etli CFN 42, Rhizobium etli CIAT 652, Sphingomonas wittichii RW1, Sphingomonas wittichii RW1, Afipia sp 1NLS2, Rhizobium leguminosarum bv viciae, Sinorhizobium meliloti CCNWSX0020
LFFFQW	6,48E-03	O: Vitis vinifera, Carica papaya, Arabidopsis thaliana, Wolffia australiana, Physcomitrella patens, Physcomitrella patens, Arabidopsis lyrata subsp lyrata, Selaginella moellendorffii, Picea sitchensis, Datisca glomerata, Populus trichocarpa, Eutrema halophilum, Glycine max, Glycine max, Ricinus communis, Phaeodactylum tricornutum CCAP 10551
LILFQV	3,37E-02	N: Mycobacterium intracellulare MOTT-64, Mycobacterium parascrofulaceum ATCC BAA-614, Mycobacterium bovis BCG str, Mycobacterium colombiense CECT 3035, Mycobacterium tuberculosis

Tabla A.V.2. Combinaciones de aminoácidos de sitios activos y de canales con sus respectivos parámetros calculados para calcular las constantes $k_{on}chem$ y $k_{off}chem$ para todas las trHbs de cada organismo.

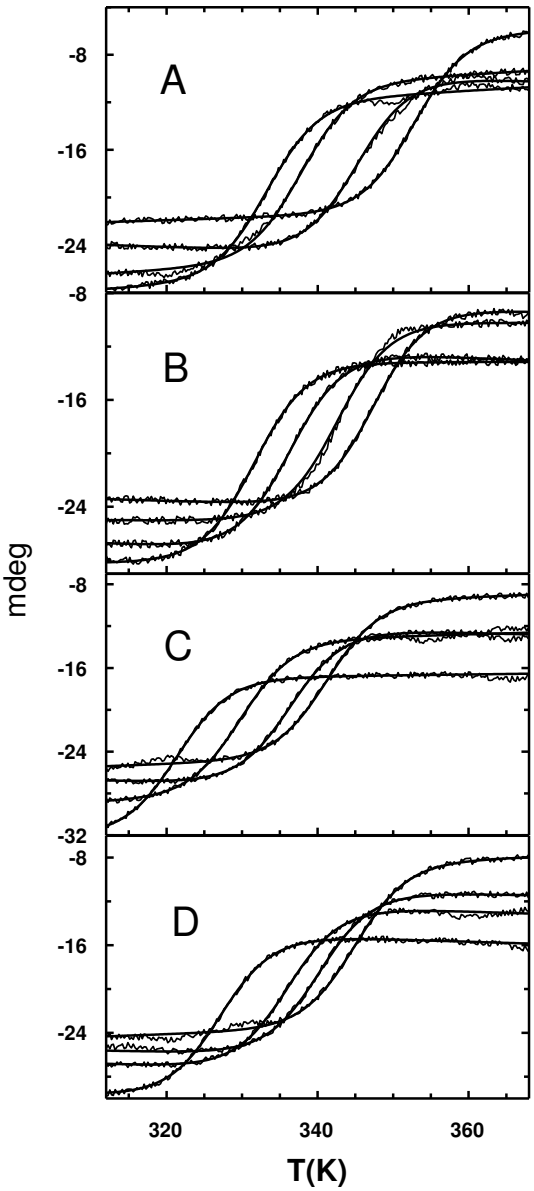


Figura A.VI.1. Curvas de desplegado térmico seguidas por dicroísmo circular a 225nm, para la Tf-trHbO wt (A) y su mutante Pro74Gly (B), en comparación a la Mt-trHbO wt (C) y su mutante Gly43Pro (D).

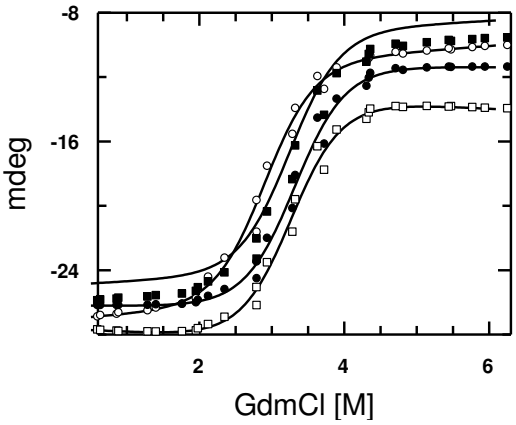


Figura A.VI.2. Perfiles de desnaturalización utilizando GdmCl seguidos por dicroísmo circular a 225nm, para la Tf-trHbO wt (□) y su mutante Pro74Gly (■) y para la Mt-trHbO wt (○) y su mutante Gly43Pro (●).

X. Agradecimientos

Llevar a cabo la presente tesis doctoral ha implicado un esfuerzo enorme, y este documento refleja el resultado de años de intenso trabajo y dedicación, haciendo lo que me gusta. Llegar a este momento esperado sería impensado si no estuviese rodeado de excelentes personas, muchas de ellas guiándome, otras aconsejándome, otras integrando un hermoso clima de trabajo, y muchas, muchas otras, haciéndome el aguante. Este agradecimiento, para todos ustedes!!!

Quisiera agradecer en primer lugar a las entidades que han financiado de una u otra forma el trabajo y mi formación doctoral. Agradezco al CONICET y a la Universidad de Buenos Aires por el apoyo en becas, equipamiento y congresos a los que asistí. Al DQIAyQF y al INQUIMAE por el espacio concedido. También a la Unión Europea por financiar en gran medida los viajes internacionales realizados.

A los jurados, Gustavo Parisi, Rodolfo González Lebrero y Sebastián Fernández Alberti, por su buena predisposición y tiempo invertido para evaluar esta tesis doctoral.

A Darío, mi director de tesis. Fuiste y siempre serás mi maestro científico. Aprendí muchísimo con vos, y destaco que hacer el doctorado en tu grupo de modelado molecular me voló la cabeza, no solo por lo que aprendí, de muchas cosas, sino que además me permitiste viajar mucho. Esto permitió continuar diversas líneas de colaboración con distintos grupos de investigación, permitiéndome conocer y entablar relaciones con gente muy buena, tanto a nivel profesional como personal. Me guiaste para comenzar a formarme como científico, me transmitiste entusiasmo, ganas, pasión por esto que hacemos en tu laboratorio. Formaste un super grupo, y super con todas las letras, formaste un grupo de excelencia en lo que hace y de una calidad humana muy grande.

A Leo, mi co-director, con quien tuvimos decenas de miles de charlas sobre proyectos, técnicas, enfoques, e incontables debates de ciencia, caracterizados por extensas reuniones e interminables conversaciones por mails. Tu contribución hacia mi formación doctoral ha sido muy grande, me enseñaste a pensar de manera crítica y racional, de manera científica, debatiendo cada detalle importante, de principio a final. Tu didáctica al momento de transmitirme conocimientos me transmitió mucha pasión, e hizo que el proceso de aprendizaje sea muy intenso y aprovechador. Te agradezco también por tu participación como socio activo en la creación de nuestro querido “Jardín de Encalada”.

A Marce, designado como mi consejero de estudios, pero que para mí es mucho más que eso. Desde el primer momento me brindaste tu ayuda y activa participación en los proyectos en los que estuve involucrado. Todos los que te conocemos sabemos que sos una persona muy singular, para los entendidos, un *outlier* de la ciencia. Sos una persona que yo aproveché y sigo aprovechando muchísimo. Hemos tenido muchas discusiones científicas, y a tu estilo, discusiones a fondo, con gritos de por medio que parecía que se terminaba el universo y se iba el paper al joraca, pero tan fructíferas, tan jugosas, que siempre dará gusto seguir teniéndolas. También me aconsejaste mucho en distintas etapas sobre pasos a seguir en el ámbito profesional; fuiste, sos y seguirás siendo un guía científico para mí.

A nuestros colaboradores, principalmente a Giulietta Smulevich, Cristiano Viappiani, Stefania Abbruzzetti, Javier Luque, Cinzia Verde y Daniela Giordano. Aprendí muchísimo al momento de

trabajar con ellos de manera bien interdisciplinaria. Me siento muy afortunado de haber trabajado y aprendido de grandes maestros como ellos.

A todo el GMM, el numeroso grupo de trabajo con el que conviví estos años, rodeado de excelentes personas, con quienes se generó una muy linda relación. Compartimos muchas charlas, muchas de ellas científicas, y otras no tanto… Recuerdo esos divertidos almuerzos donde hablábamos sobre diversos temas, siempre dándoles el *touch* nerd, y, algunas otras veces, donde nos divertíamos divagando... y por qué no, citar algunos casos célebres, como por ejemplo el poliamor, la consultora de cadáveres, estrategias para matar al nóbel, etc etc… A ustedes chicos: Lu, Jenny, Lula, Ithu, Clau, Noosh, Facto, Ferchu ♥, Uri, Dieguito H., Ari, Pablito, Juan R., Eli, Rodri, Fran, Nico, Yami, Laia y Fede. También a quienes de una manera u otra son parte del mismo grupo: Adrián T., Ale Nadra, Nano, Damián S., Nacho, Lucas, Radu, Mode, Lanza, Juan B. Y a quienes ya no están en el GMM, pero que también hemos compartido muchas cosas: Charlito, Dieguito, Ariel, Santi y Diego M.

Mención aparte para Noosh, traductora profesional y *babysitter* (agregalo a tu CV, esta mención es tu certificado). Me has ayudado tanto a mejorar mis publicaciones con un inglés de primera, como a cuidar a Leoncito con un cariño inmenso.

A mis amigos de la infancia, con quienes seguimos manteniendo una hermosa relación: Lea, Juli, Jesi, Sofi, Anto, Eve, Vicky y Naty.

A Nora, Jorge, Pau, Flo y Joako, por acompañarme en esta importante etapa de mi vida.

A mamá, papá, Belu y Tita, por todo el amor que me dan, por todo lo compartido, por preocuparse por mí, por enseñarme a ser como soy, por ser mis ejemplos de vida.

A Vicky, por hacerme el aguante durante todo el doctorado, por aconsejarme, por acompañarme, por caminar juntos… y por darme el regalo más grande de la vida, León.

Muchísimas gracias a todos!!!!!!!!!

XI. Abreviaturas

Hemoglobina, Hb; hemoglobinas truncadas, trHb o 2/2 Hb; dinámica molecular, DM; método híbrido de mecánica cuántica / mecánica clásica, QM/MM (por su sigla en inglés, *Quantum Mechanics / Molecular Mechanics*); muestreo del ligando implícito, ILS (por su sigla en inglés, *Implicit Ligand Sampling*); perfil de energía libre, FEP (por su sigla en inglés, *Free Energy Profile*); canal largo, CL; canal corto G8, CCG8; portal E7, PE7; sensores acoplados a globinas, GCS (por su sigla en inglés, *globin coupled sensors*); mioglobina, Mb; neuroglobina, Ngb; Citoglobina, Cgb; alineamiento múltiple de secuencias, MSA (por su sigla en inglés, *Multiple Sequence Alignment*); verosililitud máxima, ML (por su sigla en inglés, *Maximum Likelihood*); *Thermobifida fusca*, Tf; *Mycobacterium tuberculosis*, Mt; *Basillus subtilis*, Bs.