



Flexibilidad conformacional de moléculas bioactivas: implicaciones en diseño de fármacos y función de globinas

Flavio Forti

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tdx.cat) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tdx.cat) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tdx.cat) service has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized neither its spreading and availability from a site foreign to the TDX service. Introducing its content in a window or frame foreign to the TDX service is not authorized (framing). This rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.



Departament de Fisicoquímica

Facultat de Farmàcia

Universitat de Barcelona

Flexibilidad conformacional de moléculas bioactivas: implicaciones en diseño de fármacos y función de globinas

Tesis para optar al título de Doctor por la Universitat de Barcelona

Flavio Forti, Junio 2011

Directores:

F. Javier Luque

Xavier Barril

Agradecimientos

Dicen que un doctorado es un largo camino, que hay mucho por equivocarse, por repetir, por insistir. En definitiva, por aprender. Al principio del camino definitivamente se ve así y así es. Ahora, mientras escribo esto y contemplo hacia atrás, el camino recorrido se siente largo pero corto a la vez. Parece que fue ayer que llegué a Barcelona y entré por primera vez a este laboratorio de Modelado Molecular el 6 de Noviembre de 2006. ¿Cuándo pasó todo este tiempo?

En el camino he compartido momentos con muchísima gente a la que me gustaría mencionar en esta página.

Para empezar quiero agradecer a mis directores de tesis, Javi y Xavi. No sólo por el rol que cumplieron al dirigirme a lo largo de estos años sino, y mucho más importante, por su calidad humana. Es, definitivamente, esto lo que ha hecho posible caminar sin grandes problemas por el sinuoso andar de la tesis.

A los compañeros que me recibieron en el laboratorio y con quienes compartí los primeros años. Oscar y Damián, por su compañerismo y ayuda en la primera etapa en que debía sintonizar con las costumbres del labo. A Ignacio, por las charlas cuánticas. A Axel, por sus consejos y opiniones como voz experimentada. A Jesús, por su predisposición constante a dar una mano con algún script de Perl. A Mousumi, por su honestidad y compañerismo. A Peter, por su buena onda y humor.

A los de la “segunda” etapa. A Ana, por su dulzura portuguesa. A Jordi, por su ironía “heavy metal” y crítica a todo. A Carles, por el doble sentido y constante apuesta al chiste. A Salomé por su cordialidad. A Haydee y Patricio con quienes no he compartido tanto pero siempre tuvieron gestos positivos hacia mi.

A los que pasaron por el labo. A Francesca por las correcciones en italiano y las charlas. A los coterráneos Lula y Adrián que siento como si hubieran estado desde el principio del doctorado y con quienes siempre es divertido compartir charlas. Al grupo de Darío que siempre me recibió como uno más del labo de Buenos Aires. A Leo, Marce, Damián y Pau con quienes trabajamos y reímos juntos. A Darío en particular por su siempre excelente calidad humana. A Claudio por su rigurosidad científica de la cual se aprende siempre y por recibirme en Houston.

A Ramón por la parsimonia al exponer sus principios y por proponer siempre otra visión de las cosas. A Assumpta por ser la madre del labo, regando todo de alegría y cariño. Al departamento de Fisicoquímica y las vecinas “de abajo” por darme la oportunidad de estar y hacerme sentir como en casa.

Finalmente, a mi familia y amigos y a Lur por estar ahí a lo largo y ancho del camino, mucho, muchísimo más allá de la tesis.

Barcelona, 14 de Junio de 2011

Índice

1. Introducción	9
1.1 Actividad biológica: receptor y ligando.....	11
1.2 Flexibilidad molecular, espacio conformacional y afinidad.....	12
1.3 Globinas y ligandos bimoleculares.....	13
1.4 Protoglobina de Methanosarcina Acetivorans.....	17
1.5 Ligandos flexibles: reconocimiento molecular, conformación bioactiva y coste de reorganización conformacional.....	19
1.6 Bibliografía.....	25
2. Objetivos	29
3. Métodos	33
3.1. Métodos basados en Mecánica Cuántica.....	35
3.1.1 Método Hartree-Fock (HF)	37
3.1.2 Método perturbativo Moller-Plesset (MP)	39
3.1.3 Métodos basados en el funcional de la densidad electrónica (DFT)	40
3.1.4 Métodos semiempíricos.....	42
3.1.5 Derivación de cargas parciales.....	44
3.2. Métodos basados en la mecánica clásica.....	45
3.2.1 Campos de fuerzas.....	46
3.2.1 Métodos basados en campos de fuerzas.....	49
3.3. Métodos híbridos cuántico-clásico (QM-MM)	58
3.4. Método de Metrópolis Monte Carlo.....	59
3.5. Perfiles de energía libre asociados a la migración de ligandos en proteínas.....	61
3.5.1 Método de Steered Molecular Dynamics acoplado a la ecuación de Jarzynski.....	62

3.5.2	Implicit Ligand Sampling (ILS)	63
3.5.3	FPocket	64
3.6.	Solvatación	65
3.6.1	Descripción termodinámica	66
3.6.2	Modelos de solvatación continuos	66
3.7.	Descomposición estructural de modos normales (NSD)	70
3.8.	Bibliografía	72
4.	Metodología <i>Multinivel</i>	75
4.1.	Extensión del modelo de solvatación MST al hamiltoniano semiempírico RM1	77
4.1.1	Compuestos neutros	78
4.1.2	Compuestos iónicos en agua	88
4.2.	Estrategia multinivel	95
4.2.1	Descripción de la metodología Multinivel	95
4.2.2	Exploración conformacional de moléculas flexibles: histamina	103
4.3.	Discusión	115
4.4.	Bibliografía	116
5.	Protoglobina de <i>Methanosarcina Acetivorans</i>	119
5.1	Impacto de la distorsión del hemo en la afinidad por O ₂	121
5.1.1	Estructuras hexa y pentacoordinada y limitaciones del modelo	121
5.1.2	Efecto de las distorsiones fuera del plano (OOP)	122
5.1.3	Efecto de las distorsiones en el plano (IP)	125
5.1.4	Análisis de la alta afinidad de MaPgb por O ₂	127
5.1.5	Efectos en la distribución de densidad electrónica	132
5.2	Dinámica Molecular de MaPgb: relación entre la estructura cuaternaria y la unión de ligando	133
5.2.1	Evolución estructural de MaPgb	134

5.2.2	Detección del ligando unido y comunicación al resto de la proteína.....	138
5.2.3	Migración de ligandos.....	140
5.3	Discusión acerca de las implicaciones funcionales.....	144
5.4	Bibliografía.....	148
6.	Conclusiones	151
7.	Publicaciones	155

Capítulo 1

Introducción

1.1 Actividad biológica: receptor y ligando

La expresión actividad biológica suele englobar toda la serie de procesos que tienen lugar en un organismo como resultado de la influencia de un compuesto, en general, xenobiótico¹. Un compuesto es biológicamente activo suele ejercer su actividad interaccionando con una macromolécula endógena denominada receptor. En adelante nos centraremos en los receptores más usuales, es decir, aquéllos de naturaleza proteica. Al compuesto biológicamente activo lo denominaremos ligando, dado que su unión al receptor desencadena los procesos asociados a su efecto biológico.

Para llegar hasta el receptor macromolecular, la estructura química de un ligando exógeno debe facilitar su biodistribución en el organismo, permitiendo superar procesos tales como la absorción intestinal, el paso de la barrera hematoencefálica o bien, ya a nivel celular, el acceso a la estructura celular donde se halla el receptor, atravesando la membrana de la célula. Una vez próximo al receptor, el ligando debe alcanzar el sitio activo (o de unión), que es donde se concreta específicamente la interacción ligando-receptor. A nivel molecular, el sitio activo puede estar parcialmente expuesto al solvente o bien enterrado dentro de la estructura del receptor. En este caso, el ligando hace uso de cavidades y canales presentes en la estructura molecular del receptor para migrar a través de la matriz proteica y alcanzar el sitio activo.

En general el ligando es una molécula pequeña en comparación con su receptor macromolecular. Su tamaño oscila desde moléculas diatómicas sencillas como O₂, CO y NO hasta aquéllas de mayor tamaño que pueden alcanzar algunas decenas de átomos. Un típico ejemplo de actividad biológica relacionada con moléculas diatómicas es la respiración pulmonar, pues la captación de O₂ está mediada por la hemoglobina, siendo posteriormente transportada por la sangre hasta alcanzar tejidos como el músculo, donde se almacena en la mioglobina. La interacción de O₂ con dichas proteínas implica la migración a través de diversos sitios de unión en la matriz proteica hasta alcanzar la cavidad que contiene el grupo prostético hemo.

Por otro lado, los compuestos bioactivos de mayor tamaño incluyen el principio activo (fármaco) de medicamentos, o bien moléculas endógenas sintetizadas y liberadas por el mismo organismo. La mayor complejidad asociada a la estructura química de dichos ligandos hace que su actividad biológica refleje el balance de diversos factores,

que incluyen no sólo la migración hacia el centro de unión, sino la interacción entre los grupos químicos del compuesto bioactivo y los residuos presentes en el centro activo, la desolvatación del ligando (y del centro de unión del receptor), el coste energético asociado a la selección de a conformación bioactiva, así como contribuciones entrópicas debido a la reducción en la flexibilidad conformacional de ligando y receptor y la expulsión de moléculas de agua de hidratación.

En el presente estudio nos centraremos en procesos asociados a la difusión molecular del ligando a través de la matriz proteica hasta el sitio de unión, así como la selección de la conformación bioactiva implicada en el reconocimiento ligando-receptor.

1.2 Flexibilidad molecular, espacio conformacional y afinidad

Exceptuando moléculas sencillas como O₂, tanto ligandos como receptores son moléculas flexibles. Por tanto, la comprensión de los procesos descritos anteriormente es extremadamente compleja dado que no existe una única representación tridimensional de estas moléculas cuando están libres en solución o formando el complejo ligando-receptor a la temperatura del organismo. Por el contrario, receptores y ligandos poseen una gran cantidad de estados térmicamente accesibles, que definen el espacio conformacional accesible a una temperatura determinada, y que reflejan esencialmente la flexibilidad interna debida a rotaciones a través de enlaces simples.

La riqueza conformacional de una molécula aumenta exponencialmente con su tamaño dado que cuantos más átomos –y, por lo tanto, enlaces, ángulos y torsiones– posea, tantos más grados de libertad existirán. Estos grados de libertad no son de naturaleza aditiva, sino que se multiplican dando origen a un número muy elevado de posibles estados conformacionales, que deben ser considerados a priori en el proceso de formación del complejo ligando-receptor.

Tanto la difusión del ligando a través de la matriz proteica como la unión en el centro activo del receptor son procesos íntimamente relacionados con la flexibilidad conformacional de ligando, receptor y complejo ligando-receptor. Dichos procesos determinan la afinidad entre ligando y receptor¹, que a su vez está relacionada con la cinética de asociación y de disociación entre ligando y receptor, tal como se expresa en

la Figura 1.2.1.

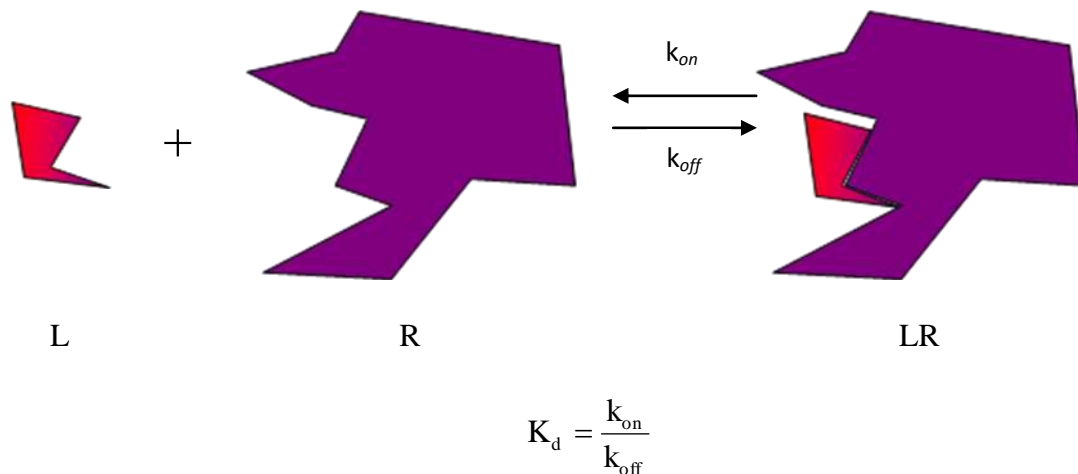


Figura 1.2.1. Esquema ilustrativo del proceso de unión entre ligando (L) y receptor (R) para dar el complejo (LR). Definición de la constante de equilibrio K_d asociada al proceso.

Las constantes cinéticas k_{on} y k_{off} dependen tanto de los procesos de entrada y salida del ligando, así como de la interacción entre ligando y receptor. La cinética de disociación, regulada por k_{off} , es unimolecular e inversamente proporcional al tiempo de vida medio del complejo ligando-receptor (LR). Por otro lado, la cinética de asociación, regulada por k_{on} , es bimolecular, reflejando la probabilidad de que ligando (L) y receptor (R) den lugar a la formación del complejo.

1.3 Globinas y ligandos bimoleculares

Las globinas son proteínas presentes en todos los organismos vivos que poseen el grupo prostético hemo en su estructura, lo que les permite exhibir una elevada sensibilidad a moléculas diatómicas, como O_2 , NO y CO . Pertenecen a la superfamilia de las hemoproteínas, aunque comparten algunas características distintivas. Recientemente Janeth Thornton² y colaboradores han estudiado más de 2300 hemoproteínas presentes en el Protein Data Bank (PDB), observando una amplia diversidad en las funciones de los 34 grupos en que se han clasificado. Ello refleja variaciones en la forma, volumen y entorno químico del sitio de unión y en el modo de unión del hemo y la naturaleza de las interacciones hemo-proteína. Este estudio

demuestra la diversidad estructural y funcional existente dentro de esta superfamilia.

La secuencia de las globinas, como subgrupo de las hemoproteínas, puede variar considerablemente de una a otra. Sin embargo, poseen una estructura terciaria común conocida como plegamiento globina, formado esencialmente por la disposición de tres hélices alfa apiladas sobre otras tres hélices alfa. Además, algunos residuos se hallan extremadamente conservados en posiciones específicas. Sin embargo, la evolución ha colocado determinados aminoácidos en zonas sensibles de la estructura tridimensional (3D) de cada globina, que modulan en gran medida tanto la difusión como la afinidad de ligandos, condicionando por tanto la función biológica de dichas hemoglobinas.

El caso prototípico de globina es el de la mioglobina (Mb), no sólo por ser miembro del primer linaje sino también por tratarse de una de las proteínas más estudiadas. Esto le ha valido la etiqueta del “átomo de hidrógeno de la biología”³. Kendrew^{4,5} resolvió hace más de 50 años su estructura 3D lo que le valió el Premio Nobel de Química en 1962. Pese a la multitud de estudios acerca de esta globina, aún hoy es motivo de debate y estudio por su complejidad, relevancia y función biológica.⁶

La Mb es una globina citoplasmática de 154 residuos que, como la hemoglobina (Hb), une reversiblemente O₂. En tanto que Mb exhibe un comportamiento de Michaelis-Menten, el tetrámero de Hb posee una curva sigmoidea de saturación de O₂. Hb posee, de este modo, el conocido efecto alostérico. La función biológica que se atribuye a Mb es la reserva de O₂ en el músculo, reflejado por ejemplo en mamíferos marinos que deben pasar largos períodos de apnea. Entre otras funciones, se ha propuesto que sirve de proveedor de O₂ intracelular, manteniendo su concentración constante⁶. Se han propuesto otras funciones, que son aún motivo de controversia, y ha sido relacionada con la inactivación de NO así como de especies reactivas de oxígeno⁶.

La naturaleza y disposición 3D de los residuos en el plegamiento globina condiciona la existencia de una red de canales y cavidades en la matriz proteica, que regulan el acceso selectivo de ligandos. Así, pues, la accesibilidad de los ligandos puede verse afectada por la presencia de barreras asociadas a la transición entre diversas cavidades, o incluso por la presencia de ciertos aminoácidos que pueden actuar como puertas de acceso a través de túneles⁷. Por otro lado, la presencia de aminoácidos dadores de puente de hidrógeno en el sitio de unión, generalmente ubicados en la cavidad distal, modulan la energía del enlace que se establece entre el ligando y el grupo hemo⁸. A

mayor número de puentes de hidrógeno entre la proteína y el ligando unido al hemo mayor es la energía asociada a la unión ligando-receptor. Este efecto es conocido como efecto distal.

Por otro lado, la reactividad del grupo hemo se ve afectada por la estructura local de la histidina proximal, único residuo conservado en todas las globinas conocidas. Esto redundante en una modulación de la afinidad por ligandos como oxígeno que se conoce como efecto proximal⁸. En algunos casos el plano del anillo imidazólico de la histidina proximal se halla orientado en una conformación alternada respecto a los nitrógenos pirrólicos del hemo mientras que en otros está eclipsado. La leghemoglobina (Lba) es un ejemplo del primer caso y la Mb del segundo. La conformación alternada permite una mejor donación de carga de la histidina al hierro del hemo, lo que refuerza la retrodonación π de éste al ligando, aumentando la fuerza del enlace Fe-ligando⁸. Es por ello que, pese a tener entornos distales similares, la k_{off} de Lba es menor que la de Mb.

La combinación de estas diferencias estructurales en la cavidad del grupo hemo abre un abanico de propiedades muy diferenciadas entre las globinas. Es por ello que éstas cumplen funciones tan diversas como el transporte y almacenamiento de gases, propiedades sensoras relacionadas con el control de determinados procesos, la oxidación de compuestos orgánicos, el transporte de electrones y la catálisis de reacciones entre especies reactivas de oxígeno y nitrógeno^{9,10,11}.

Estas funciones están determinadas por la afinidad selectiva de las distintas globinas por los diferentes ligandos bimoleculares. A su vez, ello depende de la relación entre las constantes cinéticas k_{on} y k_{off} para diversos ligandos di y triatómicos (Figura 1.3.1).

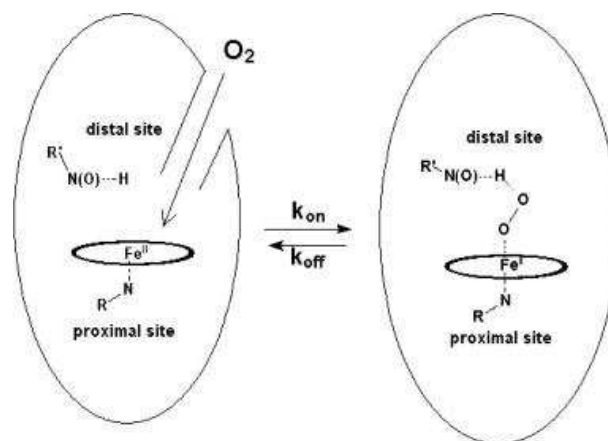


Figura 1.3.1. Proceso de asociación (k_{on}) y disociación (k_{off}) en globinas

El proceso de asociación, determinado por k_{on} , depende esencialmente de la accesibilidad del ligando a la cavidad distal, donde tendrá lugar la coordinación con la posición axial vacante del grupo prostético hemo. Dado que la formación del enlace con el Fe es generalmente rápida,⁸ el proceso de asociación está determinado en general por la migración del ligando a través de la proteína. La migración sucede a través de cavidades contiguas por las que el ligando debe difundir, superando barreras energéticas asociadas al paso entre las mismas gracias a las fluctuaciones térmicas. El conjunto de cavidades conforma un sistema de túneles que pueden estar regulados por diversos mecanismos, como la presencia de residuos que limitan o facilitan el acceso de ligandos, tal como se ha descrito en las hemoglobinas truncadas (TrHbs)¹².

El proceso de disociación, dado por k_{off} , está determinado por varios factores relacionados con la estabilidad del enlace ligando-Fe. Entre ellos podemos mencionar la fuerza del propio enlace ligando-Fe, la interacción del ligando con residuos dadores de puente de hidrógeno en la cavidad distal, así como otras formas de interacción no covalente con residuos distales. A su vez, una vez completada la ruptura del enlace ligando-Fe, la probabilidad de escape del ligando del sitio distal y la difusión a través de la matriz proteica también pueden afectar el proceso disociativo. La hemoglobina truncada N de *Mycobacterium tuberculosis* (MtTrHbN) es un ejemplo de baja k_{off} por estabilización del ligando por varios puentes de hidrógeno con residuos distales (TyrB10, GlnE11)^{13,14}.

La función biológica de hemoglobinas también puede verse afectada por otros factores, como la adopción de una estructura cuaternaria en condiciones fisiológicas. Mientras algunas de ellas se hallan como monómeros, como Mb, en otras la estructura fisiológica involucra diversas subunidades, formando agregados poliméricos, como la Hb de mamíferos,^{11,15-17} cuya estructura tetramérica está unida al efecto alostérico presente en la unión de O₂.

Estructuralmente, el motivo de plegamiento 3D más extendido es el conocido como apilamiento de hélices 3-sobre-3, aunque en las últimas décadas se han hallado nuevas estructuras como el apilamiento 2-sobre-2, que define al subgrupo de hemoglobinas 2/2Hbs, cuya secuencia está acortada en unos 20 o 30 aminoácidos con respecto a la mioglobina. Este motivo se halla presente en MtTrHbN, que participa en la regulación de la detoxificación de NO¹⁸⁻²¹. Otro subgrupo con peculiaridades estructurales lo

forman la protoglobina y las proteínas sensoras acopladas a globinas (globin-coupled sensors; GCS), que contienen alrededor de 190 aminoácidos^{18,19,22,23}. Las protoglobinas son las primeras globinas de dominio único relacionadas con GCS encontradas en Archaea. Aunque se desconoce aún su función fisiológica, se sabe que pueden unir O₂, CO y NO reversiblemente in vitro^{18,19,22,23}. En todos los casos mencionados el único aminoácido conservado es la histidina proximal HisF8²⁰, lo que deja el sexto sitio de coordinación del Fe en el lado distal para la unión de ligandos exógenos.

En estos casos la plasticidad conformacional está únicamente asociada al receptor, que cumple un papel fundamental en el proceso de migración de ligandos. De hecho, gran parte de los esfuerzos realizados en estudios computacionales de hemoproteínas²⁴⁻²⁷ están dirigidos a la exploración del espacio conformacional de la macromolécula y/o de aminoácidos específicos con objeto de entender las propiedades bioquímicas fruto de su interacción con ligandos.

1.4 Protoglobina de *Methanosarcina Acetivorans*

El grupo de Martino Bolognesi publicó recientemente la estructura de rayos X de la protoglobina de *Methanosarcina acetivorans* (MaPgb)²⁸. En el PDB hay 2 estructuras para MaPgb: 2VEB (forma oxigenada, 1.3 Å de resolución) y 2VEE (forma libre, 2.6 Å de resolución). La inspección de dichas estructuras revela una serie de características inusuales respecto a las globinas conocidas. Así, MaPgb contiene un total de 9 hélices entre las cuales existe una hélice pre-A denominada Z. El grupo hemo está altamente distorsionado (Figura 1.4.1) y completamente enterrado en la matriz proteica, dejando a los grupos propionato inaccesibles al solvente por la presencia del loop EF, así como por 20 residuos sin estructura secundaria del extremo N-terminal. Habitualmente las globinas poseen los propionatos expuestos a solvente y el hemo prácticamente plano, estructura termodinámicamente más estable para el hemo libre. Esto sugiere que la distorsión es debida a cierta tensión generada por los residuos que rodean al hemo, aunque no existen evidencias directas sobre la naturaleza concreta de dichos residuos.



Figura 1.4.1. Hemo de mioglobina (izquierda) y MaPgb (derecha)

La estructura oxigenada (2VEB) muestra al O₂ unido al grupo hemo sin la estabilización por puentes de hidrógeno de residuos distales, típica de otras globinas mencionadas en esta memoria como Mt-trHbN o Mb. El único residuo distal capaz de formar un puente es la Tyr(B10)61, aunque en el cristal se halla a 5.3 Å del ligando. El acceso de ligandos diatómicos al sitio distal sería posible debido a la presencia de dos túneles apolares que conectan el solvente con el hemo y están delimitados por las hélices B/G (túnel 1) y B/E (túnel 2), tal como se muestra en la Figura 1.4.2; **Error! No se encuentra el origen de la referencia.** Es importante notar que estos túneles son topológicamente distintos de aquéllos encontrados en otras globinas.

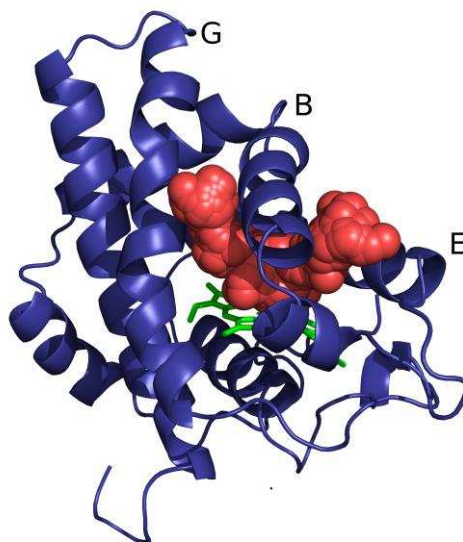


Figura 1.4.2. Túneles de MaPgb

MaPgb fue cristalizada como homodímero y eluye como tal en solución en los experimentos de filtración en gel. Como en otros GCS, se supone que el dímero es el estado biológicamente activo en solución. El área de contacto entre las subunidades que conforman este homodímero simétrico está básicamente definido por los pares de hélices G/H y H/G, que conforman un agregado intermolecular de 4 hélices con parte de la hélice Z y los loops BC y FG, totalizando una interfase de contacto de 2086 Å².

Estas características estructurales distintivas y la inusualmente baja k_{off} para O₂

(0.092-0.0094 s⁻¹) reportada por el grupo de Martino Bolognesi²⁸ en ausencia de estabilización por puentes de hidrógeno en la cavidad distal, convierten a la MaPgb en un caso relevante para el estudio de estructura-función en hemoproteínas. En particular, son de especial interés la relación existente entre la alta distorsión del grupo hemo, la baja k_{off} y la dinámica de la proteína y migración de ligandos a través de los túneles.

1.5 Ligandos flexibles: reconocimiento molecular, conformación bioactiva y coste de reorganización conformacional

En el caso de ligandos de mayor tamaño, la afinidad ligando-receptor depende del reconocimiento molecular selectivo y específico entre ambos, que está regulado por un delicado balance donde el tamaño, la forma y las propiedades fisicoquímicas del ligando juegan un papel central. Dicho balance se traduce en una elevada complementariedad entre las propiedades electrostáticas, estéricas e hidrofóbicas del ligando con aquéllas del centro de unión en el receptor. Cuanto mayor sea dicha complementariedad, mayor será la afinidad entre ligando y receptor, y mayor será entonces la actividad biológica.

El reconocimiento molecular, punto de partida de casi todo proceso biológico²⁹, fue establecido hace más de 100 años por Emil Fischer,³⁰ quien escribió: “una enzima y un glucósido deben encajar como una llave en una cerradura para poder iniciar una reacción química entre ellos”. El reconocimiento molecular es aún motivo de intenso estudio y se sabe que su complejidad va más allá de lo que enunciara Fischer. Hoy se acepta que se debe tener en cuenta la flexibilidad de las biomoléculas a la hora de estudiar cómo se logra el reconocimiento molecular. Esto da origen los modelos de ajuste inducido y la existencia de colectivos de conformaciones preequilibradas³¹.

Una de las contribuciones electrostáticas más relevantes en el reconocimiento molecular es el puente de hidrógeno debido básicamente a su direccionalidad³²⁻³⁵. Esta interacción resulta de la atracción electrostática entre un átomo electronegativo Y (O, N, S) y un átomo de hidrógeno unido a otro átomo electronegativo X. La distancia típica entre los átomos pesados X(H) e Y varía de 2.5 a 3.2 Å, y el ángulo X-H...Y de 130 a 180 grados¹. A distancias más cortas (hasta 2.3 Å) la interacción adopta un carácter más

covalente que aumenta la energía de interacción³⁶⁻³⁸. También son importantes los puentes salinos y las interacciones π -catión, donde entran en juego átomos con carga neta^{1,39-41}.

Otra contribución al reconocimiento molecular es la interacción de van der Waals, que involucra un componente de carácter repulsivo y otro atractivo. La contribución atractiva está determinada por la formación de dipolos instantáneos (no permanentes) originados por la fluctuación en la distribución electrónica de las partes interaccionantes y es de corto alcance⁴². La contribución repulsiva es aún de más corto alcance y está determinada por el recubrimiento de las nubes electrónicas de las partes interaccionantes, siendo el reflejo del principio de exclusión de Pauli.

Para el caso de ligandos flexibles, la riqueza conformacional no reside únicamente en el receptor, sino que el ligando también visita múltiples conformaciones en el medio celular. Por tanto, el reconocimiento molecular, que requiere de la complementariedad ligando-receptor, exige que el ligando sea capaz de adoptar una conformación capaz de amoldarse a las características del centro de unión del receptor. En general, solamente un número muy reducido de todas las conformaciones exploradas por el ligando a una temperatura determinada en el medio fisiológico satisfacen los requerimientos necesarios para interactuar con el receptor. Por ello, para compuestos flexibles, la afinidad entre ligando y receptor está modulada no sólo por el grado de complementariedad, sino también por el coste de reorganización conformacional asociado a la adopción de la conformación bioactiva.

El coste de reorganización conformacional se ve compensado por la interacción favorable entre ligando y receptor, tal como se ilustra de forma simplificada en la Figura 1.5.1, donde $\Delta G_{\text{conf}}^{\text{L}}$ es el coste requerido para adoptar la conformación bioactiva L' a partir de la conformación estable en solución L. Asimismo, $\Delta G_{\text{conf}}^{\text{R}}$ representa el coste asociado a la reorganización del (centro de unión del) receptor. Claramente, cuanto mayor sea el coste asociado a la selección de la conformación bioactiva del ligando y a la reorganización del receptor, menor será la afinidad ligando-receptor.

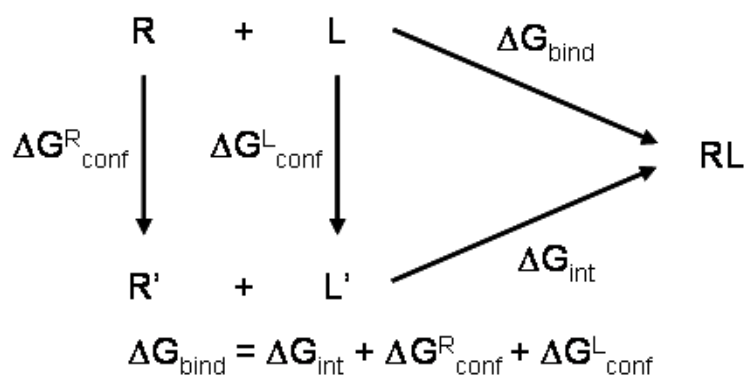


Figura 1.5.1. Ciclo termodinámico para la formación del complejo ligando-receptor.

Los estudios de afinidad ligando-receptor exigen, por lo tanto, no sólo una adecuada identificación de los determinantes moleculares asociados a la interacción entre ambos¹, sino también una correcta estimación del coste de reorganización conformacional⁴³. Dado que el coste de reorganización conformacional va en detrimento de la afinidad, es lógico suponer que un buen ligando se unirá al receptor adoptando una conformación de baja energía. No obstante, diversos estudios han llegado a conclusiones muy distintas a este respecto. Así, Perola et al.⁴⁴ analizaron el coste energético y la reorganización conformacional de 150 ligandos, cuya estructura cristalográfica en el complejo con el receptor era conocida. El estudio confirmó la noción general de que los ligandos raramente se unen en su conformación de más baja energía, estimándose que “el 60% de los ligandos no se unen en una conformación de mínimo local”. Asimismo, sus resultados muestran que “mientras que el coste de reorganización conformacional calculado fue de menos de 5 kcal/mol para aproximadamente el 60% de los ligandos, se obtuvieron costes de más de 9 kcal/mol para el 10% de los casos independientemente del método utilizado”.

Tirado-Rives y colaboradores⁴⁵ realizaron una comparación de diversos métodos computacionales para estimar el coste conformacional de cuatro moléculas con distinto grado de flexibilidad conformacional. En su trabajo concluyen que “cuando un ligando se une a una proteína, lo hace generalmente en una conformación que no es un mínimo de energía para el ligando libre y existe una pérdida de grados de libertad conformacional”. Cuanto más flexible es el ligando, más puede alejarse de los mínimos de energía potencial para adoptar la conformación bioactiva, lo que puede significar que

el ligando libre no tenga una población significativa de la conformación bioactiva. Por otra parte, los autores concluyen que la incertidumbre de la estimación del coste de reorganización conformacional es aproximadamente de 5 kcal/mol si se evalúa incluyendo una búsqueda conformacional con un método de solvatación GB/SA y cálculo de energías para los conformeros clave con métodos ab initio o DFT. Si se utilizan campos de fuerzas, la incertidumbre puede estimarse de manera conservadora en el rango 5-10 kcal/mol.

Otros estudios realizados sobre colectivos conformacionales sugieren que el coste entrópico es menor a 3 kcal/mol¹. Asimismo, Vieth y colaboradores encontraron que la distribución espacial de los puntos de anclaje responsables de la unión del ligando al receptor coinciden en las conformaciones bioactivas con la distribución en las conformaciones de mínima energía¹.

Estos resultados demuestran la dificultad que conlleva la estimación cuantitativa del coste conformacional asociado a la unión de ligando al receptor. La situación es aún más delicada si se tiene en cuenta que existen otras contribuciones a la energía de unión, como puede ser el coste energético que conlleva la distorsión estructural inducido en el receptor debido a la unión con el ligando. Por otra parte, dado que por cada 1.4 kcal/mol de incremento en el coste implicado en la selección de la conformación bioactiva se pierde un orden en la afinidad de unión, la incertidumbre que ello conlleva en la estimación de la afinidad entre ligando y receptor puede ser muy elevada.

Ciertamente, es razonable pensar que el impacto de la incertidumbre “conformacional” será menor en el caso de ligandos que tengan reducida flexibilidad interna, o bien si se analizan diferencias de afinidad entre series homólogas de compuestos, dado que en gran medida debe existir una cancelación de la contribución asociada a la reorganización conformacional de ligando y receptor cuando dichos ligandos se unen al mismo centro de unión. No obstante, las consideraciones precedentes obligan a ser cautos a la hora de comparar ligandos estructuralmente diversos y/o que interaccionan en centros diferentes del receptor, puesto que en estos casos asumir dicha cancelación podría dar lugar a errores considerables en la predicción de la afinidad. De hecho, cabe pensar que esta limitación es una de las causas que subyacen en la limitada validez de las funciones de puntuación empleadas en métodos de docking⁴⁵.

Desde el punto de vista fisicoquímico, el coste de reorganización conformacional puede dividirse en una contribución de naturaleza entálpica y una entrópica^{1,46}. La primera está asociada esencialmente a la diferencia de energía que existe entre la conformación bioactiva y los mínimos en la superficie de energía potencial (SEP) del ligando libre en solución. La contribución entrópica, por otro lado, está relacionada con la diferencia en el espacio conformacional accesible para el ligando libre en solución y unido al receptor.

En el trabajo de Chang y colaboradores⁴⁷ se aborda la cuestión de la pérdida de entropía configuracional del ligando en su unión al receptor, lo cual va en detrimento de la afinidad. Ésta se separa en 2 términos aditivos: la entropía conformacional, asociada al número de mínimos de energía, y la vibracional, relacionada con el volumen de estos mínimos. En muchos modelos la pérdida de entropía configuracional es ignorada o se recurre a aproximaciones basadas en supuestos físicos discutibles. Uno de estos supuestos es la disminución de rotámeros accesibles al ligando por la restricción al movimiento que presenta al unirse al receptor. Utilizando el método Mining Mínima se estudia en este trabajo la pérdida de entropía configuracional de amprenavir en su unión a la proteasa del VIH. El estudio muestra que la mayor contribución a la disminución de la entropía configuracional no es debida a una reducción en el número de rotámeros accesibles sino al angostamiento de los mínimos de energía del ligando unido respecto al libre, es decir, a una disminución esencialmente de la entropía vibracional. Es importante recalcar que aún la disminución más drástica de rotámeros accesibles del ligando libre versus ligando unido conlleva una reducción en la entropía configuracional menor. Esto es así dado que la entropía conformacional escala con el logaritmo del cociente entre el número de rotámeros ligado y libre.

La aplicación estricta de los principios de la termodinámica estadística demuestra que la energía libre es una propiedad global del sistema en consideración¹ por lo que depende del espacio conformacional total. El desarrollo de nuevas metodologías computacionales para la estimación del coste de reorganización conformacional requiere, por lo tanto, de la completa exploración del espacio conformacional del ligando libre en solución, así como la identificación de las estructuras estables en solución. Finalmente, para cada familia conformacional presente solución, es necesaria una estimación precisa tanto de su energía como de su entropía.

En resumen, existe aún una importante incertidumbre en la determinación de las

energías libres relativas de los distintos conformeros de un ligando en solución y de la correcta geometría y energía relativa de la conformación bioactiva. Esta incertidumbre, por sí misma, ya es suficiente para considerar inadecuada la viabilidad de los métodos de docking actuales,⁴⁵ y representa todo un desafío al estudio de la interacción ligando-receptor. Si se pudieran obtener mejores estimaciones de las energías libres relativas en solución de los distintos conformeros de un ligando, las energías de corte recomendadas en métodos de generación conformacional de compuestos flexibles podrían ser sensiblemente inferiores al umbral de 20 kcal/mol, que se emplea habitualmente como umbral para seleccionar conformaciones en el cribado virtual de compuestos⁴⁸. Esto redundaría en un menor número de conformaciones a considerar y, consecuentemente en métodos más eficientes.

1.6 Bibliografía

- (1) Gohlke, H.; Klebe, G. *Angew. Chem. Intl. Ed.* **2002**, 41, 2644.
- (2) Smith, L. J.; Kahraman, A.; Thornton, J. M. *Proteins* **2010**, 78, 2349.
- (3) Tomita, A.; Sato, T.; Ichiyanagi, K.; Nozawa, S.; Ichikawa, H.; Chollet, M.; Kawai, F.; Park, S.-Y.; Tsuduki, T.; Yamato, T.; Koshihara, S.-y.; Adachi, S.-i. *P Natl Acad Sci USA* **2009**.
- (4) Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C. *Nature* **1958**, 181, 662.
- (5) Burmester, T.; Weich, B.; Reinhardt, S.; Hankeln, T. *Nature* **2000**, 407, 520.
- (6) Ordway, G. A.; Garry, D. J. *J Exp Biol* **2004**, 207, 3441.
- (7) Hornak, V.; Okur, A.; Rizzo, R. C.; Simmerling, C. *Proc Natl Acad Sci USA* **2006**, 103, 915.
- (8) Marti, M. A.; Crespo, A.; Capece, L.; Boechi, L.; Bikiel, D. E.; Scherlis, D. A.; Estrin, D. A. *J Inorg Biochem* **2006**, 100, 761.
- (9) Ghosh, A. *THE SMALLEST BIOMOLECULES: DIATOMICS AND THEIR INTERACTIONS WITH HEME PROTEINS*; Elsevier, 2007; Vol. 1.
- (10) Paoli, M.; Marles-Wright, J.; Smith, A. *DNA Cell Biol* **2002**, 21, 271.
- (11) Vinogradov, S. N.; Moens, L. *J Biol Chem* **2008**, 283, 8773.
- (12) Ruscio, J. Z.; Kumar, D.; Shukla, M.; Prisant, M. G.; Murali, T. M.; Onufriev, A. V. *Proc Natl Acad Sci USA* **2008**, 105, 9204.
- (13) 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, 97.
- (14) 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, 4433.
- (15) Frauenfelder, H.; McMahon, B. H.; Fenimore, P. W. *P Natl Acad Sci USA* **2003**, 100, 8615.
- (16) Scott, E. E.; Gibson, Q. H.; Olson, J. S. *J Biol Chem* **2001**, 276, 5177.
- (17) Wittenberg, J. B.; Wittenberg, B. A. *J Exp Biol* **2003**, 206, 2011.
- (18) Freitas TAK, H. S., Dioum EM, Saito JA, Newhouse J, Gonzalez G, Gilles-Gonzalez M-A, Alam M. *Proc Nat Acad Sci USA* **2004**, 101, 6675.

- (19) Hou S, F. T., Larsen RW, Piatibratov M, Sivozhelezov V, Yamamoto A, Meleshkevitch EA, Zimmer M, Ordal GW, Alam M. *P Natl Acad Sci USA* **2001**, 98, 9353.
- (20) O. H. Kapp LM, J. V., C. N. Trotman, T. Suzuki, and S. N. Vinogradov. *Protein Sci* **1995**, 4, 2179–2190.
- (21) Ouellet H, O., Y., Richard, C., Labarre, M., Wittenberg, B., Wittenberg, J., Guertin. *P Natl Acad Sci USA* **2002**, 99, 5902.
- (22) Vinogradov, S. N.; Hoogewijs, D.; Bailly, X.; Mizuguchi, K.; Dewilde, S.; Moens, L.; Vanfleteren, J. R. *Gene* **2007**, 398, 132.
- (23) Vinogradov, S.; Hoogewijs, D.; Bailly, X.; Arredondo-Peter, R.; Gough, J.; Dewilde, S.; Moens, L.; Vanfleteren, J. *Evol Biol* **2006**, 6, 31.
- (24) Bidon-Chanal, A.; Marti, M. A.; Crespo, A.; Milani, M.; Orozco, M.; Bolognesi, M.; Luque, F. J.; Estrin, D. A. *Proteins* **2006**, 64, 457.
- (25) Bidon-Chanal, A.; Marti, M. A.; Estrin, D. A.; Luque, F. J. *J Am Chem Soc* **2007**, 129, 6782.
- (26) Boechi, L.; Manez, P. A.; Luque, F. J.; Marti, M. A.; Estrin, D. A. *Proteins* **2009**, 78, 962.
- (27) Boechi, L.; Marti, M. A.; Milani, M.; Bolognesi, M.; Luque, F. J.; Estrin, D. A. *Proteins* **2008**, 73, 372.
- (28) Nardini M, P. A., Thijs L, Saito JA, Dewilde S, Alam M, Ascenzi P, Coletta M, Ciaccio C, Moens L, Bolognesi M. *EMBO Rep.* **2008**, 9, 157.
- (29) Stryer, L., Berg, J. M., Tymoczko, J. L., *Biochemistry*, 2002, W.H.Freeman & Co Ltd, Voet, D., Voet, J. G. *Biochemistry*; John Wiley & Sons, 2004.
- (30) Fischer, E. *Ber. Dtsch. Chem. Ges.* **1894**, 27, 2985.
- (31) Spyrakis, F., Bidon Chanal, A., Barril, X., Luque, F., *J. Curr Top Med Chem* **2011**, 11.
- (32) Fersht, A. R. *Trends Biochem. Sci.* **1987**, 12, 301.
- (33) Murphy, K. P., Gill, S. J. *J. Mol. Biol.* **1991**, 222, 699.
- (34) Yang, A., Sharp, K. A., Honig, B. *J. Mol. Biol.* **1992**, 227, 889.
- (35) Zou, Q., Habermann-Rottinghaus, S. M., Murphy, K. P. *Proteins* **1998**, 31, 107.
- (36) Steiner, T., Saenger, W. *Acta Crystallogr. Sect. B* **1994**, 50, 348.

-
- (37) Wang, Z., Luecke, H., Nanhua, Y., Quioco, F. *Nat. Struct. Biol.* **1997**, 4, 519.
- (38) Warshel, A., Papazyan, A., Kollman, P. A. *Science* **1995**, 269, 102.
- (39) Meyer, E. A., Castellano, R. K., Diederich, F. *Angew. Chem. Int. Ed.* **2003**, 42, 1210.
- (40) Barril, X., Alemán, C., Orozco, M., Luque, F. J. *Proteins* **1998**, 32, 67.
- (41) Ma, J., Dougherty, D. A. *Chem. Rev.* **1997**, 97, 1303.
- (42) Burley, S. K., Petsko G. A. *Adv. Protein Chem.* **1988**, 39, 125.
- (43) Leach, A. R. *Molecular Modelling Principles and Applications*; Addison Wesley Longman Limited, 1996.
- (44) Perola, E. C., P. S. *J Med Chem* **2004**, 47, 2499.
- (45) Tirado-Rives, J. J. W. L. *J Med Chem.* **2006**, 49, 5880.
- (46) Hunter, C. A. *Angew. Chem. Int. Ed.* **2004**, 43, 5310.
- (47) Chang, C.-e. A.; Chen, W.; Gilson, M. K. *Proc Natl Acad Sci USA* **2007**, 104, 1534.
- (48) Butler, K. T.; Luque, F. J.; Barril, X. J. *Comput. Chem.* **2009**, 30, 601.

Capítulo 2

Objetivos

El presente trabajo persigue analizar el impacto de la flexibilidad conformacional de biomoléculas en la interacción ligando-receptor. En una primera parte el objeto central de estudio es el ligando, donde se propone el desarrollo de una metodología para la exploración conformacional de ligandos. En la segunda parte la atención se centra en el receptor, explorando las implicaciones funcionales de sus características estructurales y dinámicas.

El objetivo específico de la primera parte del trabajo es el desarrollo de una metodología robusta para la exploración conformacional completa de un ligando de relevancia biológica. Para ello, se persigue la identificación de las familias conformacionales más relevantes, así como de sus estabilidades relativas en solución. La metodología desarrollada se denomina Multinivel, dado que combina una exploración mediante la técnica de Monte Carlo mediante el modelo de solvatación continuo MST desarrollado por nuestro grupo para el hamiltoniano RM1 con cálculos ab initio de alto nivel a posteriori para los conformeros de interés. En este contexto, la memoria está estructurada en tres partes:

- 1) Extensión del modelo de solvatación continuo MST al hamiltoniano semiempírico RM1
- 2) Desarrollo y aplicación de la metodología Multinivel Monte Carlo-RM1/MST acoplada a cálculos ab initio.
- 3) Caracterización conformacional de cuatro compuestos antirretrovirales del VIH, dada su relevancia fisicoquímica y biológica.

En la segunda parte los objetivos específicos del estudio de la hemoproteína Protoglobina de *Methanosarcina acetivorans* se centran en dos aspectos principales. Por un lado, establecer la relación existente entre la distorsión excepcional del grupo hemo y la afinidad por ligando. Por otra parte, se pretende determinar la influencia ejercida por la dinámica estructural de la proteína en la difusión de ligandos. Los objetivos concretos son:

- 1) Impacto de la distorsión del grupo hemo en la afinidad por O₂.

- 2) Caracterizar la plasticidad dinámica de Protoglobina en solución mediante estudios de dinámica molecular, con el fin de profundizar en i) la migración de ligandos bimoleculares a través de canales presentes en la matriz proteica, y ii) la dinámica esencial de la proteína y su relación con la difusión de ligando.

Capítulo 3

Métodos

La Química Teórica precisa del desarrollo de metodologías capaces de reproducir, predecir o explicar resultados experimentales valiéndose del conocimiento de las leyes que rigen el comportamiento de la materia y sus transformaciones. La aplicación más rigurosa hace uso de los principios de la mecánica cuántica. Así, el reconocido físico y premio Nobel Richard Feynman estableció que “todas las cosas están hechas de átomos y todo lo que los organismos vivos hacen puede ser entendido en términos del menear y sacudir de los átomos”¹, y que “la parte más profunda de la Química Teórica debe terminar en la Mecánica Cuántica”¹.

A medida que el número de átomos y electrones de un sistema crece, el uso de métodos cuánticos se hace inviable y es necesario recurrir a modelos de menor complejidad, si bien ello supone buscar un compromiso entre la precisión y exactitud y la aplicabilidad de dichos modelos. Como dijo Dirac (también premio Nobel de Física) en 1929: “Las leyes subyacentes necesarias para la teoría matemática de gran parte de la física y toda la química son por tanto completamente conocidas, y la dificultad es sólo que la aplicación exacta de estas leyes deriva en ecuaciones demasiado complicadas como para ser resolubles. Es por lo tanto deseable que métodos prácticos aproximados en la aplicación de la mecánica cuántica deberían ser desarrollados, que deriven en una explicación de las características principales de los sistemas atómicos complejos sin demasiado cálculo.” Ésta ha sido una ingente tarea en estas décadas que, con el advenimiento de la informática y el creciente poder de cálculo, ha permitido alcanzar resultados inimaginables.

En el modelado de sistemas químicos existen básicamente dos grandes modelos físicos bien diferenciados: aquéllos que tienen en cuenta el carácter cuántico de los electrones, y aquéllos que lo ignoran recurriendo a la mecánica clásica. Estos últimos utilizan campos de fuerzas que, mediante el ajuste de parámetros, proporcionan una descripción simplificada del sistema molecular. En esta sección se presentarán los principios de los métodos utilizados en esta tesis.

3.1. Métodos basados en Mecánica Cuántica

El estudio de las propiedades químicas de una molécula, y en particular de su reactividad, requiere el conocimiento de su estructura electrónica. Un sistema compuesto por un conjunto de núcleos y electrones viene descrito por una función de

onda que depende de las coordenadas de núcleos (\mathbf{R}) y electrones del sistema (\mathbf{r}) y del tiempo. Su obtención pasa por resolver la ecuación de Schrödinger dependiente del tiempo:

3.1-1

$$\hat{H}\Psi = i\hbar\partial_t\Psi$$

donde \hat{H} es el operador Hamiltoniano, ∂_t representa la derivada parcial respecto al tiempo y Ψ es la función de onda que describe el estado del sistema.

La ecuación 3.1-1 describe cómo un sistema cuántico evoluciona con el tiempo. La función de onda contiene toda la información acerca del sistema, proporcionando la descripción más completa que puede darse a un sistema físico, sin tener en cuenta efecto relativistas.

La expresión de H(en unidades atómicas) es:

3.1-2

$$H = -\sum_i \frac{\nabla_i^2}{2M_i} - \sum_i \frac{\nabla_i^2}{2} + \sum_{ij} \frac{Z_i Z_j}{|\mathbf{R}_i - \mathbf{R}_j|} - \sum_{ij} \frac{Z_i}{|\mathbf{R}_i - \mathbf{r}_j|} + \sum_{i<j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}$$

donde los distintos términos corresponden por orden a la energía cinética de los núcleos, la energía cinética de los electrones, la repulsión núcleo-núcleo, la repulsión núcleo-electrón, y finalmente la repulsión electrón-electrón.

En el caso de "soluciones estacionarias", es decir, aquellas donde H no depende explícitamente del tiempo, la ecuación 3.1-1 se reduce a una expresión de autovalores y autovectores de la forma:

3.1-3

$$H|\Psi\rangle = E\Psi$$

Esta ecuación sólo tiene solución analítica para sistemas con un electrón como el átomo de hidrógeno o átomos hidrogenoides. Por ello, los métodos de la Química Cuántica recurren a diversas aproximaciones. Por un lado, emplean la aproximación de Born-Oppenheimer, donde se desacopla el movimiento nuclear del movimiento de los

electrones. Ello equivale a suponer que el movimiento de los electrones es suficientemente rápido para adaptarse instantáneamente al cambio en las posiciones de los núcleos. Por otro lado, dado que la posición de los núcleos es fija, el término de repulsión nuclear es constante, permitiendo definir el Hamiltoniano electrónico, que incluye una parte monoeléctronica y otra bielectrónica. Así, la función de onda pasa a depender solamente de la posición de los electrones, aunque paramétricamente de la posición de los núcleos.

3.1-4

$$H = \sum_i \left(\frac{\nabla_i^2}{2} + \sum_j \frac{Z_j}{|R_j - r_i|} \right) + \sum_{i < j} \frac{1}{|r_i - r_j|}$$

La resolución de la ecuación 3.1-3 usa el concepto de orbital, que es una función monoeléctronica empleada para expandir la función de onda en contribuciones monoatómicas. Por otro lado, la descripción de una partícula cuántica precisa de una característica adicional, el spin, de forma que la descripción de un orbital conlleva producto de una parte espacial $|\phi\rangle$ y una de espín $|s, m_s\rangle$, denominándose espín-orbital.

En base al principio de antisimetría, la función de onda de un sistema de dos o más electrones debe ser antisimétrica frente al intercambio simultáneo de las coordenadas espaciales y de espín. Por ello, la descripción un conjunto de N electrones independientes adopta la forma de un determinante de Slater².

3.1-5

$$|\Psi\rangle = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(x_1) & \chi_1(x_2) & \cdots & \chi_1(x_N) \\ \chi_2(x_1) & \chi_2(x_2) & \cdots & \chi_2(x_N) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_N(x_1) & \chi_N(x_2) & \cdots & \chi_N(x_N) \end{vmatrix}$$

3.1.1 Método Hartree-Fock (HF)

La teoría Hartree-Fock (HF) describe el estado fundamental de un sistema cuántico con un único determinante de Slater formado con el mejor conjunto posible de orbitales

moleculares. Para ello, seminimiza el valor de la energía del sistema alterando los espín-orbitales moleculares en base al principio variacional, según el cual para cualquier función de onda normalizada (Φ) que cumpla las condiciones de contorno el valor esperado para el hamiltoniano H es mayor o igual que el valor esperado para el estado fundamental.

La aplicación del principio variacional, sujeto a la condición de ortonormalidad de los espín-orbitales, conduce a la ecuación 3.1-6, donde se introducen los operadores de Coulomb (J; 3.1-7) y de intercambio (K; 3.1-8). El primero de ellos representa el potencial local promedio que experimenta el electrón 1 debido al electrón 2 y es fácilmente identificable en términos clásicos. Por otra parte, el segundo no tiene un análogo clásico y se origina debido a la naturaleza antisimétrica del determinante de Slater frente al intercambio de electrones.

3.1-6

$$\left[h(1) + \sum_{b \neq a} \widehat{\mathcal{J}}_b(1) - \sum_{b \neq a} \widehat{\mathcal{K}}_b(1) \right] \chi_a(1) = \epsilon_a \chi_a(1)$$

3.1-7

$$\widehat{\mathcal{J}}_b(1) = \int \chi_b^*(2) r_{12}^{-1} \chi_b(2) d\mathbf{x}_2$$

3.1-8

$$\widehat{\mathcal{K}}_b(1) \chi_a(1) = \left[\int \chi_b^*(2) r_{12}^{-1} \chi_a(2) d\mathbf{x}_2 \right] \chi_b(1)$$

En base a la ecuación 3.1-9, se puede definir el operador de Fock (f; ecuación 3.1-10), de modo que las ecuaciones de HF se convierten en 3.1-11. La función de onda resultante es el determinante de Slater formado por los espín-orbitales así obtenidos.

3.1-9

$$[\widehat{\mathcal{J}}_a(1) - \widehat{\mathcal{K}}_a(1)] \chi_a(1) = 0$$

3.1-10

$$f(1) = h(1) + v^{HF}(1) = h(1) + \sum_b (\widehat{\mathcal{J}}_b(1) - \widehat{\mathcal{K}}_b(1))$$

donde v^{HF} describe el campo efectivo que nota un electrón debido a la presencia de los N-1 electrones restantes.

3.1-11

$$f|\chi_a\rangle = \epsilon_a|\chi_a\rangle$$

Como el operador de Fock depende de los orbitales y sus autovectores son a su vez un nuevo conjunto de orbitales la solución a las ecuaciones de HF requiere de un proceso iterativo, conocido como método autoconsistente. Dado que el potencial v^{HF} no trata la interacción electrónica instantánea sino en forma promediada, el método HF describe de forma incompleta la correlación electrónica. Por ello, dichos efectos deben incluirse en tratamientos post-HF.

3.1.2 Método perturbativo Moller-Plesset (MP)

La energía HF dista mucho de ser la energía real del sistema (no relativista), ya que no incluye la energía de correlación. Una metodología eficiente propuesta para la inclusión de la correlación electrónica es el método Möller-Plesset-n (MP-n), conocido como teoría de perturbaciones a orden n^3 . Basado en la teoría de Rayleigh-Schrödinger⁴, consiste esencialmente en tomar la solución HF como punto de partida para un desarrollo en serie de componentes perturbativos de la energía del sistema y su función de onda.³

3.1-12

$$E = E^{(0)} + \lambda E^{(1)} + \lambda^2 E^{(2)} + \dots$$

$$|\Phi\rangle = |\Psi^{(0)}\rangle + \lambda|\Psi^{(1)}\rangle + \lambda^2|\Psi^{(2)}\rangle + \dots$$

donde $E^{(0)}$ y $|\Psi^{(0)}\rangle$ son la energía y la función de onda HF respectivamente, y los siguientes términos designan las sucesivas correcciones a ambos.

La convergencia de la serie no está a priori asegurada. Por otra parte, la energía total del sistema suele tener un carácter oscilante, aunque en principio se espera que dichas oscilaciones disminuyan al aumentar el orden de aproximación. Se trata de un método consistente con el tamaño (es decir, la energía del sistema es la suma de energías de los sistemas independientes), pero no es variacional, con lo cual es posible obtener energías por debajo de la energía exacta.

3.1.3 Métodos basados en el funcional de la densidad electrónica (DFT)

Estos métodos se basan en la relación existente entre la energía electrónica total y la densidad electrónica de un sistema. Los fundamentos teóricos están basados en los dos teoremas de Hohenberg-Kohn⁵.

- 1) El estado fundamental de un sistema multielectrónico está unívocamente definido por la densidad electrónica. Esto reduce el problema de muchos cuerpos de N electrones con 3N coordenadas espaciales a uno que depende sólo de 3.
- 2) Define un funcional para el sistema y demuestra que la densidad electrónica correcta para el estado fundamental lo minimiza.

El funcional de energía se escribe como la suma de dos términos:

$$E[\rho(\mathbf{r})] = \int V_{\text{ext}}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} + F[\rho(\mathbf{r})]$$

donde $\rho(\mathbf{r})$ es la densidad electrónica del sistema, V_{ext} es la interacción de los electrones con un potencial externo (normalmente la interacción electrostática con los núcleos) y $F[\rho(\mathbf{r})]$ es la suma de la energía cinética de los electrones así como las interacciones interelectrónicas.

La menor energía del sistema corresponde a la densidad electrónica del estado fundamental, y por tanto se usa el principio variacional para resolver la ec.3.1-144. El problema, no obstante, radica en saber la forma del funcional $F[\rho(\mathbf{r})]$. Kohn y Sham sugirieron (1965)⁶ la siguiente aproximación para este término:

3.1-15

$$E[\rho] = T[\rho] + E_{\text{el}}[\rho] + E_{\text{xc}}[\rho]$$

donde el primer término corresponde a la energía cinética, el segundo a la repulsión coulombica electrón-electrón y el tercero contiene las contribuciones de las integrales de intercambio y correlación.

Sustituyendo la ec. 3.1-155 en la ec. 3.1-144 se resuelve el sistema de manera autoconsistente, empezando por una aproximación a la densidad real y trabajando iterativamente según el principio variacional. Existen distintas aproximaciones para el tratamiento del funcional, entre las cuales podemos mencionar:

- a) Local (LDA): Depende solamente del valor de la densidad electrónica en cada el punto del espacio y no en derivadas de la densidad u orbitales
- b) Gradiente generalizado (GGA): Es una aproximación local pero, a diferencia de LDA, toma en cuenta también el gradiente de la densidad en el punto del espacio donde se evalúa la densidad.
- c) Meta Gradiente generalizado (meta-GGA): Es un paso adicional a GGA dado que depende del valor, el gradiente y el Laplaciano (segunda derivada) de la densidad.
- d) Híbrido: Incorpora una la expresión exacta para el intercambio de la teoría de HF y agrega intercambio y correlación de otras fuentes. Éstas pueden ser ab initio o empíricas.

En esta tesis los cálculos DFT realizados se basan en el funcional B3LYP, que es un híbrido del funcional de Becke (E_x^{GGA})⁷ de 3 parámetros para el término de intercambio y el funcional de Lee-Yang-Parr (E_c^{GGA})⁸ para el término de correlación (ecuación 3.1-16).

$$E_{xc}^{B3LYP} = E_{xc}^{LDA} + a_0(E_x^{HF} - E_{xc}^{LDA}) + a_x(E_x^{GGA} - E_x^{LDA}) + a_c(E_c^{GGA} - E_c^{LDA})$$

donde $a_0 = 0.20$, $a_x = 0.72$, $a_c = 0.81$ son parámetros ajustados empíricamente mediante energías de atomización, afinidades protónicas y energías atómicas totales.

3.1.4 Métodos semiempíricos

El desarrollo inicial de los métodos semiempíricos estuvo ligado al coste computacional de cálculos ab initio para sistemas grandes o para sets de muchas moléculas. Aunque el poder de cálculo ha crecido exponencialmente en los últimos 20 años, el desarrollo y refinamiento de estos métodos no ha cesado, dado que se mantienen una relación aceptable entre la calidad de los resultados y el coste de cálculo.

Los métodos semiempíricos más recientes (MNDO⁹; AM1¹⁰, PM3^{11,12} y PM5, RM1¹³, PM6¹⁴) comprenden unas características fundamentales que los caracterizan:

a) son métodos AVE (All Valence Electron), lo que significa que sólo consideran los electrones de valencia que se mueven en el campo creado por un core fijo de núcleos y electrones internos. Así, se considera un hamiltoniano electrónico reducido, H_v tomando sólo los términos correspondientes a los electrones de valencia:

$$H_v = T_c + V_{cn} + H_{cc} = -\sum_{i=1}^{N_v} \frac{1}{2} \nabla_i^2 - \sum_{i=1}^{N_v} \sum_{A=1}^M \frac{Z_A^{\text{core}}}{r_{iA}} + \sum_{i=1}^{N_v} \sum_{j=1}^{N_v} \frac{1}{r_{ij}}$$

- b) están basados en la aproximación NDDO (Neglect of Diatomic Differential Overlap), que considera nula el recubrimiento diferencial entre orbitales situados en átomos diferentes. Ello conlleva una drástica disminución en el número de integrales que deben calcularse
- c) utilizan una base mínima de orbitales de Slater.
- d) la energía total de la molécula E_T es la suma de las energías electrónicas y las repulsiones de los cores.
- e) diversos términos de la matriz de Fock y las repulsiones E_{AB}^{core} no son evaluados analíticamente, sino que se determinan a partir de datos experimentales o de expresiones semiempíricas que contienen parámetros cuyo valor se ajusta a

datos experimentales. Esto persigue la idea de compensar las limitaciones de utilizar una función de onda monodeterminantal y los errores debidos a la aproximación NDDO.

Aparte de diferencias en la parametrización, los distintos Hamiltonianos semiempíricos se diferencian por la modificación de la función de repulsión de core, usada para evaluar E_{AB}^{core} , agregando gaussianas atractivas en el caso de los métodos AM1 y PMx con objeto de mejorar la descripción de los puentes de hidrógeno.

Método RM1

Dado que un método semiempírico en general depende de valores experimentales, la precisión del mismo se sustenta no sólo en el formalismo y aproximaciones realizadas, sino también en su parametrización. En esta línea, recientemente Rocha et al.¹³ introdujeron con el nombre de RM1 una reparametrización del método AM1 con el objeto de mejorar su capacidad predictiva para las propiedades moleculares, reajustando sus 191 parámetros con un set de 1736 compuestos de interés orgánico y bioquímico.

Tal como se hizo con AM1, la parametrización se realizó en tres etapas: i) átomos centrales de la química orgánica: H, C, N, O, ii) átomos relacionados con la vida: S, P, y iii) átomos útiles en fármacos: halógenos. Los resultados obtenidos se ven en la Tabla 3.1-1, donde se compara la performance con otros métodos semiempíricos similares.

Tabla 3.1-1 - Errores medios para diversas propiedades con AM1, PM3, PM5 y RM1

Properties	AM1	PM3	PM5	RM1	N^a
Enthalpies of formation (kcal · mol ⁻¹)	9.06	5.98	5.41	5.04	986
Dipole moment (D)	0.26	0.29	0.33	0.23	59
Ionization potential (eV)	0.48	0.55	0.44	0.42	102
Bond length (Å)	0.027	0.023	0.030	0.020	618
Angles (degree)	6.06	7.33	10.50	7.05	737

a. número de datos usados en la comparación

Salvo para ángulos (donde AM1 es levemente superior), RM1 resulta más preciso que los otros métodos semiempíricos. La entalpía de formación posee prácticamente la mitad del error que poseía AM1 y los errores asociados a distancias, potenciales de ionización y momentos dipolares revierten la tendencia alcista de AM1-PM3-PM5. Este último es de central importancia en el cálculo de propiedades en solución acuosa, punto clave en el desarrollo del presente trabajo.

A lo largo del desarrollo de esta tesis J. Stewart publicó un nuevo método denominado PM6¹⁴ que ha mostrado muy buenos resultados, inclusive para sistemas grandes como proteínas.

3.1.5 Derivación de cargas parciales

En muchos casos es de particular interés conocer la carga parcial de los átomos del sistema. Si bien la densidad electrónica global es un observable, el modo de particionarla en átomos es arbitraria, por lo cual se han desarrollado varios métodos¹⁵. En esta tesis se ha usado dos metodologías.

Mulliken

El análisis de Mulliken se basa en asignar directamente la densidad electrónica de los elementos diagonales al átomo sobre el que residen y luego la mitad de los términos no diagonales sobre los átomos involucrados (ec.3.1-18).

3.1-18

$$q_A = Z_A - \sum_{\substack{\mu=1; \\ \mu \text{ en } A}}^K P_{\mu\mu} - \sum_{\substack{\mu=1; \\ \mu \text{ en } A}}^K \sum_{\substack{v=1; \\ v \neq \mu}}^K P_{\mu\nu} S_{\mu\nu}$$

donde q es la carga asociada al átomo A , Z su carga nuclear, mientras que P y S designan los elementos de la densidad electrónica y de la matriz de solapamiento.

Mulliken propuso desdoblar cada población de solapamiento por igual entre las dos funciones de base utilizadas, por lo que este tipo de análisis es muy dependiente de la base utilizada. Asimismo, la asignación de la mitad de la población de solapamiento en el análisis de Mulliken es arbitraria.

Natural Bond Orbitals

A raíz de estas limitaciones surgieron métodos que realizan un tratamiento más riguroso como el análisis de población natural (NPA)¹⁶. Este método se basa en la construcción de orbitales atómicos naturales ortonormales (NAO) a partir de las funciones de base atómicas. A partir de estos NAO se construyen unos orbitales de enlace naturales (NBO) que se clasifican en: orbitales internos (core), pares libres o de enlace. Con los NBO se pueden derivar cargas netas de un modo mucho más consistente y menos dependiente de la base utilizada.

3.2. Métodos basados en la mecánica clásica

En el marco de la mecánica clásica las moléculas pueden describirse como un conjunto de masas cargadas centradas en los núcleos atómicos y unidas mediante muelles. El conjunto de fuerzas a las que se encuentran sometidas dichas partículas permite examinar de forma más eficiente el espacio configuracional de sistemas con un gran número de átomos, siempre que las propiedades de interés no supongan una redistribución importante en la densidad electrónica, como son los procesos asociados a formación y rotura de enlaces¹⁷.

La aplicación de la mecánica clásica asume dos principios¹⁸: (1) la descripción de las distintas contribuciones energéticas mediante potenciales efectivos aditivos, y (2) la transferabilidad de dichos potenciales. En base al primer principio, se asume que la energía molecular puede expresarse como una suma de potenciales que representan fuerzas físicas simples. Las contribuciones energéticas se clasifican en dos grupos: (1) enlazantes, cuando describen las longitudes y ángulos de enlace y las torsiones de los diedros, y (2) los términos no enlazantes, que generalmente se limitan a las interacciones electrostática y de van der Waals^{17,18}. Así, la energía suele expresarse como

$$E_{total} = \left(\sum_{\substack{\text{enlaces} \\ i}} E_{\text{enlace}}(b_i) + \sum_{\substack{\text{ángulos} \\ i}} E_{\text{ángulos}}(\theta_i) + \sum_{\substack{\text{diedros} \\ i}} E_{\text{died.}}(\chi_i) \right)_{\text{enlazados}} + \left(\sum_{\substack{\text{no} \\ \text{enlazados} \\ (i,j), i < j}} E_r(r_{ij}) \right)$$

En base al segundo principio, se asume que los parámetros derivados de datos experimentales o cálculos mecánico-cuánticos de estructuras representativas pueden utilizarse para ajustar los potenciales efectivos que describirán las macromoléculas^{17,18}.

3.2.1 Campos de fuerzas

El conjunto de parámetros y ecuaciones utilizados para describir la energía potencial de un sistema mediante la mecánica molecular (clásica) se conoce con el nombre de campo de fuerzas^{17,18}. Existen varios tipos de campos de fuerzas que se diferencian básicamente por el número de términos utilizados para calcular la energía del sistema, así como en la parametrización de dichas contribuciones. Aquéllos orientados a aplicaciones biomoleculares suelen utilizar potenciales armónicos para describir los enlaces y ángulos, series de Fourier para las torsiones, y las interacciones entre pares no enlazados se describen mediante una función del tipo Lennard-Jones (van der Waals) y una función coulombica (electrostático).

A continuación se explicará brevemente las características del campo de fuerza AMBER¹⁹, que ha sido empleado en esta tesis.

Términos enlazantes

La ley de Hooke para un oscilador armónico (ecuación 3.2-2) es el modelo más simple que puede utilizarse para describir las deformaciones de los enlaces covalentes asumiendo que el desplazamiento de la posición de equilibrio es pequeño.

$$E_{vib} = \sum_{\text{enl}} K_{vib} (r - r_{eq})^2$$

donde r es la distancia de enlace, r_{eq} la distancia de referencia (normalmente la de equilibrio) y K_{vib} es la constante de fuerza asociada.

A su vez, dicho potencial también puede aplicarse para describir las deformaciones en los ángulos de enlace.

3.2-3

$$E_{ang} = \sum_{enl} K_{ang} (\theta - \theta_{eq})^2$$

siendo K_{ang} la constante de fuerza asociada y θ_{eq} el valor de referencia para el ángulo entre tres átomos unidos por enlaces covalentes.

La rotación alrededor de un enlace implica el movimiento de los grupos adyacentes a los átomos que se encuentran formando dicho enlace, que podrán adoptar conformaciones eclipsadas o alternadas confiriendo periodicidad al término de torsión. Para expresar la periodicidad del perfil de energía de una torsión se utilizan expresiones periódicas en serie de Fourier, que pueden depender sólo de los dos átomos centrales implicados en el diedro o bien ser específico de todos los diedros que definen la torsión.

3.2-4

$$E_{tor} = \sum_{tor} \sum_n \frac{V_n}{2} (1 + \cos(n\phi - \phi_0))$$

donde V_n representa la barrera de torsión asociada a cada término de la función, n la periodicidad del término, ϕ el ángulo diedro y ϕ_0 el ángulo de fase.

Finalmente, es necesario añadir una función para describir las torsiones impropias, a fin de mantener la planaridad o la quiralidad de ciertos grupos. Este término suele describirse mediante una expresión similar a la torsión convencional:

3.2-5

$$E_{tor} = \sum_{tor} \sum_n \frac{V_n}{2} (1 + \cos(2\xi_n - 180))$$

donde V_n representa la constante de fuerza de cada término de la función, n la periodicidad del término y ξ el ángulo diedro.

Términos no-enlazantes

La forma más sencilla de describir la distribución de carga en una molécula es utilizar cargas puntuales fraccionales centradas en los núcleos de los átomos¹⁷. En consecuencia, la interacción electrostática viene representada por la ley de Coulomb:

3.2-6

$$E_{ele} = \sum_i \sum_{j>i} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}$$

siendo q_i y q_j las cargas puntuales de cada átomo, r_{ij} la distancia entre átomos y ϵ la constante dieléctrica del medio.

En la mayoría de los campos de fuerza, las cargas puntuales fraccionales se obtienen mediante ajuste al potencial electrostático molecular (MEP). En el campo de fuerzas de AMBER, las cargas se derivan con el método RESP²⁰ (Restrained Electrostatic Potential) donde se introducen restricciones hiperbólicas sobre la carga de los átomos pesados, a fin de evitar que el ajuste no restringido al potencial conduzca a soluciones donde las cargas de los átomos internos adopten valores sin sentido químico. En este trabajo se utilizó el campo de fuerzas ff99SB, cuyas cargas parciales fueron parametrizadas a nivel HF/6-31G*²¹

Para describir el potencial generado por las fuerzas de dispersión-repulsión se utiliza una función del tipo Lennard-Jones 12-6^{17,18}. Dicha expresión incluye las interacciones atractivas a corta distancia, mientras que el término repulsivo evita la fusión de átomos.

3.2-7

$$E_{vdw} = \sum_i \sum_{j>i} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right)$$

donde A_{ij} y B_{ij} son constantes para cada par de átomos relacionados con sus radios de van der Waals y durezas, y r_{ij} es la distancia entre ellos.

Parametrización de los campos de fuerzas

La clave para que un campo de fuerzas sea fiable es sin duda la calidad de los parámetros que se utilizan. Para obtener los términos de enlace se pueden emplear tanto técnicas experimentales como la espectroscopia infrarroja o Raman para los enlaces y ángulos y la de RMN para las torsiones. También es posible ajustar los términos de enlace a cálculos cuánticos de alto nivel. Respecto a los términos no enlazantes el ajuste es más complejo, en algunos campos de fuerzas se ajustan las funciones de van der Waals y electrostáticas simultáneamente a fin de garantizar que el campo de fuerzas resultante sea capaz de reproducir propiedades físicas macroscópicamente de sistemas (densidad, calor de vaporización, compresibilidad, etc). La mayoría de los campos de fuerzas, no obstante, optan por una parametrización separada donde primero se ajustan las cargas atómicas a partir de datos mecanocuánticos y luego se refinan los parámetros de van der Waals sea a partir de datos de empaquetamiento cristalino, sea a partir del ajuste a energías de interacción experimentales o mecanocuánticas.

Es preciso resaltar una vez más que la parametrización se realiza siempre sobre un sistema modelo y que la transferabilidad de los parámetros al estudio de otros sistemas no está garantizada. Es por ello que al haber parametrizado una nueva molécula o parte de un sistema se debe verificar la calidad de los parámetros por comparación con observables experimentales.

3.2.2 Métodos basados en campos de fuerzas

Dos técnicas derivadas de los campos de fuerza con especial relevancia en estudios computacionales de sistemas biomoleculares son la mecánica molecular (MM) y la dinámica molecular (MD). La diferencia esencial estriba en que MD incluye el tiempo como variable. Con MM se obtiene la energía y geometría de un sistema en un mínimo (si se optimiza) o en un punto (cálculo puntual) de la superficie de energía potencial, mientras que con MD se recogen configuraciones del sistema a medida que éste evoluciona con el tiempo simulado en un determinado colectivo y sujeto a condiciones de contorno preestablecidas.

Mecánica Molecular

La energía potencial de una molécula es función de las coordenadas cartesianas asociadas a la posición de cada átomo en el espacio. Así, se puede asociar cada estructura a un punto en una hipersuperficie de energía potencial. En dicha hipersuperficie pueden existir muchos puntos de mínima energía que representan mínimos locales, siendo el de menor energía el mínimo global^{17,18}.

Existen diversas clases de métodos para identificar los mínimos en la superficie de energía potencial, que difieren en la utilización de la información contenida en el gradiente y/o la hessiana de cara a trazar el camino hacia el mínimo de energía. En mecánica clásica, las ecuaciones del campo de fuerzas son fácilmente derivables, lo que permite realizar dicha minimización de manera analítica (Figura 3.2-1).

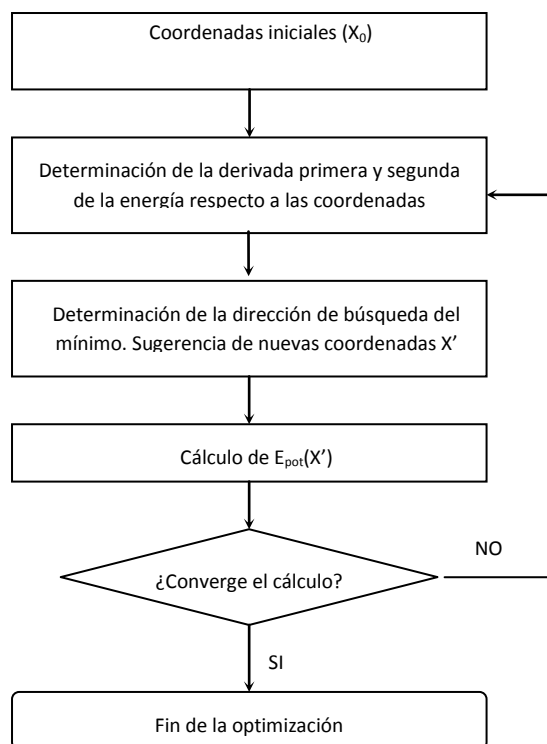


Figura 3.2-1 - Algoritmo de optimización en Mecánica Molecular

Los métodos más utilizados para la minimización son steepest descent y conjugate gradient. El primero utiliza una dirección de búsqueda basada en la dirección de máximo gradiente, siendo idóneo cuando la posición de partida está alejada del mínimo. El segundo se basa en moverse en una dirección que conjuga la del gradiente con las

direcciones previas de búsqueda, siendo más eficiente cuando la búsqueda está cercana al mínimo¹⁷. En la práctica, suele realizarse una combinación de ambos, usando primero steepest descent y conjugate gradients después.

Para sistemas pequeños es posible aplicar los métodos que calculan explícitamente segundas derivadas, tales como el método Newton-Raphson o métodos quasi-newtonianos¹⁸. Todos ellos son eficaces cuando la función de energía potencial depende de un número pequeño de variables, pues se debe calcular la hessiana del sistema a cada paso, por lo que los requisitos computacionales son mucho más grandes.

Dinámica molecular

Es un método determinístico basado en la evolución temporal de un sistema sujeto a condiciones de contorno. Esta técnica proporciona información estructural, termodinámica o cinética de procesos biomoleculares, como cambios conformacionales asociados a unión de ligandos, plegamiento de proteínas, transporte de iones o formación de agregados proteicos, entre otros^{22,23}.

A partir de una estructura inicial y un conjunto inicial de velocidades, se obtiene una trayectoria (conjunto de estructuras ordenado emporalmente) mediante la integración sucesiva de las ecuaciones de Newton. A su vez, dicha trayectoria viene determinada por las condiciones de simulación, que reflejan la elección de un determinado colectivo, como puede ser el isobárico-isotérmico, para lo cual se reajustan la posición y velocidad de las partículas que componen el sistema (Figura 3.2-2).

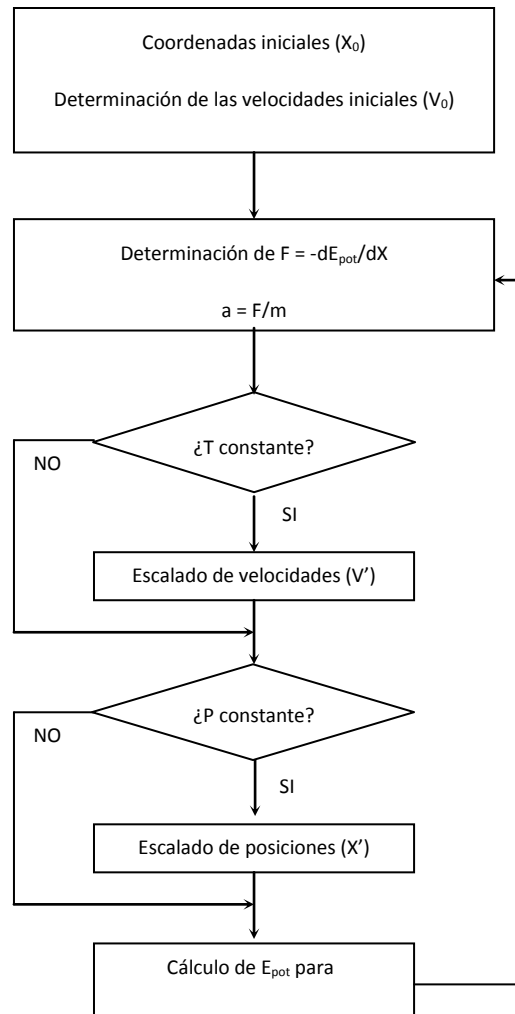


Figura 3.2-2 - Algoritmo de Dinámica Molecular

En sistemas formados por múltiples partículas bajo la influencia un potencial continuo, la fuerza que actúa sobre una partícula será función de las posiciones del resto de partículas, y cambiará al variar su posición así como la de cualquier otra partícula del sistema. Éste fenómeno, conocido como problema de muchos cuerpos, es debido a que el movimiento de las partículas está acoplado¹⁷. En consecuencia, la integración de las ecuaciones del movimiento debe realizarse de forma diferencial por métodos de diferencias finitas donde la etapa de integración debe ser siempre menor que el movimiento más rápido del sistema (véase Tabla 3.2-1)

Tabla 3.2-1 - Relación entre los intervalos de tiempo de fenómenos característicos de moléculas y sistemas biomoleculares y los métodos teóricos existentes.

evento	extensión espacial (nm)	amplitud (nm)	tiempo (s)	simulación apropiada
vibración de la distancia de enlace	0.2-0.5	0.001-0.01	10^{-14} - 10^{-13}	Métodos QM
vibración elástica de un dominio globular	1.0-2.0	0.005-0.05	10^{-12} - 10^{-11}	DM convencional
rotación de cadenas expuestas a solvente	0.5-1.0	0.5-1.0	10^{-11} - 10^{-10}	DM convencional
oscilación torsional de grupos internos	0.5-1.0	0.05	10^{-11} - 10^{-19}	DM convencional
hinge bending (movimiento relativo de dominios globulares)	1.0-2.0	0.1-0.5	10^{-11} - 10^{-7}	Dinámica de Langevin, métodos de amplio muestreo?
rotación de cadenas laterales internas	0.5	0.5	10^{-4} -1	métodos de amplio muestreo?
transiciones alostéricas	0.5-0.4	0.1-0.5	10^{-5} -1	métodos de amplio muestreo?
desnaturalización local	0.5-1.0	0.5-1.0	10^{-5} - 10^1	métodos de amplio muestreo?
movimientos de loops	1.0-5.0	1.0-5.0	10^{-9} - 10^{-5}	Dinámica Browniana
movimientos de cuerpo rígido (hélices)		1.0-5.0	10^{-9} - 10^{-6}	métodos de amplio muestreo?
transiciones hélice-enrollado		>5.0	10^{-7} - 10^4	métodos de amplio muestreo?
asociación de proteínas	>>1.0			Dinámica Browniana

La elección de δt es crucial, pues al no recalcularse las fuerzas durante este intervalo de tiempo el sistema tiende a seguir el estado de movimiento actual, lo que puede llevar a inestabilidades cuando el paso de integración es demasiado grande. No obstante, si se escoge una etapa demasiado pequeña, la exploración del espacio conformacional quedará muy limitada, siendo necesarios muchos más pasos de integración, convirtiendo el procedimiento en ineficaz computacionalmente. En la práctica suele aceptarse tiempo de integración que sea 1/10 del movimiento más rápido del sistema. Para los sistemas macromoleculares usados en esta tesis, el movimiento más rápido corresponde a la vibración de un enlace C-H, que es de unos 10 fs. Por ello, el paso de integración usado es 1 fs. Una de las opciones más utilizadas para aumentar el paso de integración es aplicar al sistema restricciones. El algoritmo más extendido para aplicar restricciones al sistema es SHAKE²⁴. Este método fija las distancias de enlace entre átomos quedando como únicas posibilidades de movimiento las torsiones y ángulos que

tienen una frecuencia de vibración menor (2×10^{-14} s para los átomos ligeros), lo cual permite duplicar el paso de integración.

Condiciones de simulación

El punto inicial de cualquier simulación es una configuración inicial, que puede provenir de datos experimentales de Rayos-X, Resonancia Magnética Nuclear o de un modelo teórico. Asimismo, precisa un conjunto de velocidades iniciales, que se asignarán al azar siguiendo una distribución de tipo Maxwell-Boltzmann, tal que la energía cinética asociada se corresponda a la energía térmica para la temperatura de trabajo. Finalmente, requiere fijar las condiciones de simulación: el número de partículas (N), el volumen (V), la temperatura (T), la presión (P) o la energía total del sistema (E). En función de la combinación de los mismos se distingue entre varios colectivos: i) microcanónico (NVE), ii) isotérmico-isobárico (NPT), o iii) canónico (NVT), siendo NPT y NVT los más usados en dinámica molecular de proteínas.

Al trabajar bajo condiciones isotérmicas-isobáricas se tendrá que controlar tanto la temperatura como la presión. En el caso de la temperatura existen diferentes opciones para su control. La primera de ellas es el re-escalado de las velocidades mediante la aplicación de un factor cada determinado tiempo de simulación. Otra alternativa es mantener la temperatura acoplada a un baño térmico externo fijado a la temperatura deseada. Dichos baños actúan como fuente de energía térmica, transfiriendo energía térmica al o desde sistema en función de las necesidades. Las velocidades son re-escaladas en función de la diferencia de temperatura entre el baño y el sistema, pero permitiendo la relajación térmica. Los métodos más conocidos son el de Berendsen²⁵ y el de Nosé-Hoover²⁶. Métodos similares son empleados para mantener la presión. Típicamente se calcula el virial para determinar la presión interna, y ésta se acopla a un pistón que garantiza equilibrio con la presión externa, aumentando o disminuyendo para ello el volumen de sistema simulado.

La imposibilidad de simular sistemas infinitos junto con el uso de potenciales continuos plantea diferentes problemas al llevar a cabo una simulación. Los sistemas están necesariamente confinados en un espacio finito, por lo que las partículas ubicadas en los límites del sistema interactúan con las paredes de éste. Para sistemas pequeños esto significa que la mayor parte de las partículas del sistema se encuentran bajo la

influencia de las paredes. Si se quieren obtener propiedades macroscópicas a partir de sistemas modelo, necesariamente debe eliminarse de alguna manera el efecto de las paredes^{17,18}. Por otro lado, dado que cada partícula interactúa con todas las otras, sería necesario calcular $N(N-1)$ pares de interacciones no enlazantes a cada paso, aumentando de forma considerable el tiempo de cálculo necesario.

El primer problema puede minimizarse si se utilizan condiciones periódicas de contorno (Figura 3.2-3), por las cuales se confina las partículas del sistema en una caja que se replica en todas las direcciones del espacio. Durante la simulación, si una partícula abandona la caja central, su imagen entrará en la caja central por el lado opuesto, manteniéndose por lo tanto constante el número de partículas en dicha caja.

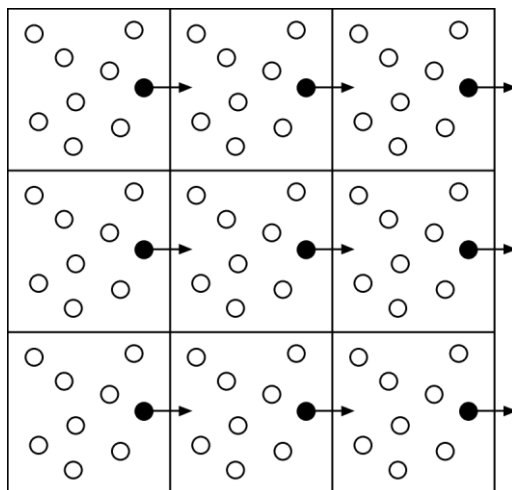


Figura 3.2-3 - Representación de las condiciones periódicas de contorno

Para el segundo problema basta con tener en cuenta la forma del potencial que se esté utilizando. El término de van der Waals, usualmente modelado como un potencial Lennard-Jones, decae rápidamente con la distancia de tal modo que a una distancia interatómica de 2.5σ (siendo σ la distancia entre átomos para la cual el potencial toma valor nulo) su valor es el 1% del que toma a la distancia de equilibrio, por lo que a distancias suficientemente grandes puede despreciarse. Así, puede restringirse el cálculo de las interacciones no enlazantes cortando a una distancia dada el potencial (cut-off) y usando el método de la mínima imagen. De esta manera las interacciones de esta clase se calculan entre una partícula y todas aquellas que se encuentren en un radio arbitrario a su alrededor escogiendo las imágenes de las partículas que están más cercanas.

Por otro lado, el potencial electrostático decae con la distancia como $1/r$, por lo que no puede despreciarse a distancias grandes tal como para van der Waals si se quieren evitar situaciones sin sentido físico en sistemas cargados como proteínas y ácidos nucleicos^{17,18}. El método más utilizado para resolver este tipo de problemas es el método de sumas de Ewald²⁷. En él se transforma el cálculo de la sumatoria de todas las interacciones electrostáticas posibles en otros dos términos de convergencia mucho más rápida: uno que engloba las interacciones a corta distancia (espacio directo), que son calculadas mediante un potencial coulombico modificado, y otro en el que se incluyen el resto (espacio recíproco), que se calcula por sumas vectoriales. Una variante más eficaz de este método es Particle Mesh Ewald²⁸, en la que se optimiza el cálculo del espacio recíproco mediante interpolación B-spline de las cargas puntuales a mallas tridimensionales. Dichas mallas son posteriormente tratadas mediante transformadas de Fourier, las cuales reducen notablemente el costo computacional.

Dinámica esencial

Dado que la flexibilidad estructural es clave en la función de las proteínas²⁹⁻³⁶, su estudio puede aportar información relevante de los procesos en que participa. Existen dos técnicas computacionales para extraer las deformaciones esenciales de un sistema de partículas: el análisis de Modos Normales (NM)³⁷⁻³⁹ y la Dinámica Esencial (ED)⁴⁰.

En el método ED, la matriz de covarianza se diagonaliza obteniéndose vectores propios que reflejan la naturaleza de los diferentes modos de deformación estructural, cuyos valores propios asociados dan una medida de la contribución de cada vector propio a la varianza global. Los valores propios son coordenadas colectivas que representan el desplazamiento a lo largo de un modo a una determinada temperatura y dan información sobre el tipo de movimiento de la biomolécula.

Cuando se dispone de diferentes sistemas sobre los cuales se requiere un estudio comparativo de la deformabilidad, la obtención de los modos esenciales posibilita una forma directa de comparación a través un índice de similitud^{41,42} que se obtiene por multiplicación escalar de los vectores (ec. 3.2-8).

$$\gamma_{AB} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n (v_i^A \cdot v_j^B)^2$$

donde n es el número mínimo de vectores propios necesarios para explicar un determinado valor (por ejemplo el 80% o 90%) de la varianza estructural.

El valor del índice se sitúa próximo a 0 cuando los movimientos son ortogonales y próximo a 1 cuando existe un grado elevado de semejanza. Si bien el índice de similitud señalado es una medida robusta para el estudio de las similitudes entre espacios conformacionales, asume que todos los modos contribuyen de igual manera al producto escalar, con lo que es imposible detectar permutaciones de los mismos. El uso de índices más complejos que tienen en cuenta los valores propios asociados a cada movimiento puede ser útil en estos casos (ec. 3.2-9)⁴³.

$$\xi_{AB} = \frac{2 \sum_{i=1}^{i=z} \sum_{j=1}^{j=z} (v_i^A \cdot v_j^B) \exp \left\{ -\frac{(\Delta x)^2}{\lambda_i^A} - \frac{(\Delta x)^2}{\lambda_j^B} \right\}}{\sum_{i=1}^{i=z} \exp \left\{ -\frac{(\Delta x)^2}{\lambda_i^A} \right\} \sum_{j=1}^{j=z} \exp \left\{ -\frac{(\Delta x)^2}{\lambda_j^B} \right\}} \left[\frac{\exp \left\{ -\frac{(\Delta x)^2}{\lambda_i^A} - \frac{(\Delta x)^2}{\lambda_j^B} \right\}}{\sum_{i=1}^{i=z} \exp \left\{ -\frac{(\Delta x)^2}{\lambda_i^A} \right\} \sum_{j=1}^{j=z} \exp \left\{ -\frac{(\Delta x)^2}{\lambda_j^B} \right\}} \right]^2$$

$$\xi_{AB} = \frac{\sum_{i=1}^{i=z} \left(\frac{\exp \left\{ -2 \frac{(\Delta x)^2}{\lambda_i^A} \right\}}{\left(\sum_{i=1}^{i=z} \exp \left\{ -\frac{(\Delta x)^2}{\lambda_i^A} \right\} \right)^2} \right)^2 + \sum_{j=1}^{j=z} \left(\frac{\exp \left\{ -2 \frac{(\Delta x)^2}{\lambda_j^B} \right\}}{\left(\sum_{j=1}^{j=z} \exp \left\{ -\frac{(\Delta x)^2}{\lambda_j^B} \right\} \right)^2} \right)^2}{\left(\sum_{i=1}^{i=z} \exp \left\{ -\frac{(\Delta x)^2}{\lambda_i^A} \right\} \right)^2 + \left(\sum_{j=1}^{j=z} \exp \left\{ -\frac{(\Delta x)^2}{\lambda_j^B} \right\} \right)^2}$$

donde λ_i es el valor propio (en Å^2) asociado al vector propio, con vector unitario v_i , y la suma puede ser extendida a todos ($z=m$) o a un número importante de vectores propios ($z=n$).

3.3. Métodos híbridos cuántico-clásico (QM-MM)

Existen casos donde interesa estudiar propiedades de un sistema grande sin obviar el tratamiento cuántico de la región de interés, como es por ejemplo una reacción química

catalizada por una enzima, procesos de excitación electrónica en fases condensadas o bien procesos con una notable polarización electrónica. En este caso existen esencialmente tres alternativas: i) recortar el sistema y estudiar sólo el sitio de interés, ii) usar un método de mecánica cuántica de bajo nivel como los métodos semiempíricos, o iii) realizar un cálculo híbrido que combine métodos cuánticos y clásicos

En esta sección se describe brevemente la tercera estrategia, que recae en el uso de métodos híbridos de mecánica cuántico-mecánica clásica^{44,45}. La parte cuántica es aquélla donde interesa que haya una correcta descripción electrónica, mientras que el resto del sistema se trata con métodos clásicos. La dificultad de estos métodos radica en el tratamiento de la frontera entre ambos subsistemas y del efecto de uno sobre el otro.

El Hamiltoniano total suele tomar la forma:

3.3-1

$$H = H_{MM} + H_{QM} + H_{QM-MM}$$

donde el primer y segundo término corresponde a la contribución del subsistema clásico y cuántico, mientras que el tercero da cuenta del acoplamiento entre ambos sistemas, que generalmente puede describirse como

3.3-2

$$H_{QM-MM} = \sum_{i=1}^A q_i \int \frac{\rho(\vec{r})}{|\vec{r} - \vec{r}_i|} d^3r + \sum_{i=1}^A \sum_{\alpha=1}^B \frac{q_i Z_\alpha}{|\vec{r} - \vec{R}_\alpha|} + H_{QM-MM}^{LJ}$$

donde el primer término designa la interacción de la contribución electrónica con las cargas clásicas, el segundo a la de los núcleos de ambos sistemas, y el tercero es la interacción de tipo Lennard-Jones entre las partículas de ambos subsistemas, siendo A es el número de núcleos de la región clásica de cargas q_i y B las partículas de la región cuántica con cargas Z.

3.4. Método de Metrópolis Monte Carlo

Esta metodología es una estrategia alternativa para la exploración del espacio conformacional de un sistema. A diferencia de MD, el tiempo no es una variable del sistema en MC. Las sucesivas configuraciones del sistema se eligen entonces mediante cambios aleatorios en las coordenadas del mismo. Cada nueva conformación generada

es comparada según su energía potencial con la del paso anterior. Si la nueva conformación del sistema es de menor energía se acepta. Si, por el contrario, es de mayor energía, su aceptación depende de si x (ec3.4-1) es mayor que un número al azar entre 0 y 1. El algoritmo de Metrópolis⁴⁶ se esquematiza en la Figura 3.4-1.

3.4-1

$$x = e^{-\frac{(E - E_0)}{k_b T}}$$

donde k_b es la constante de Boltzmann.

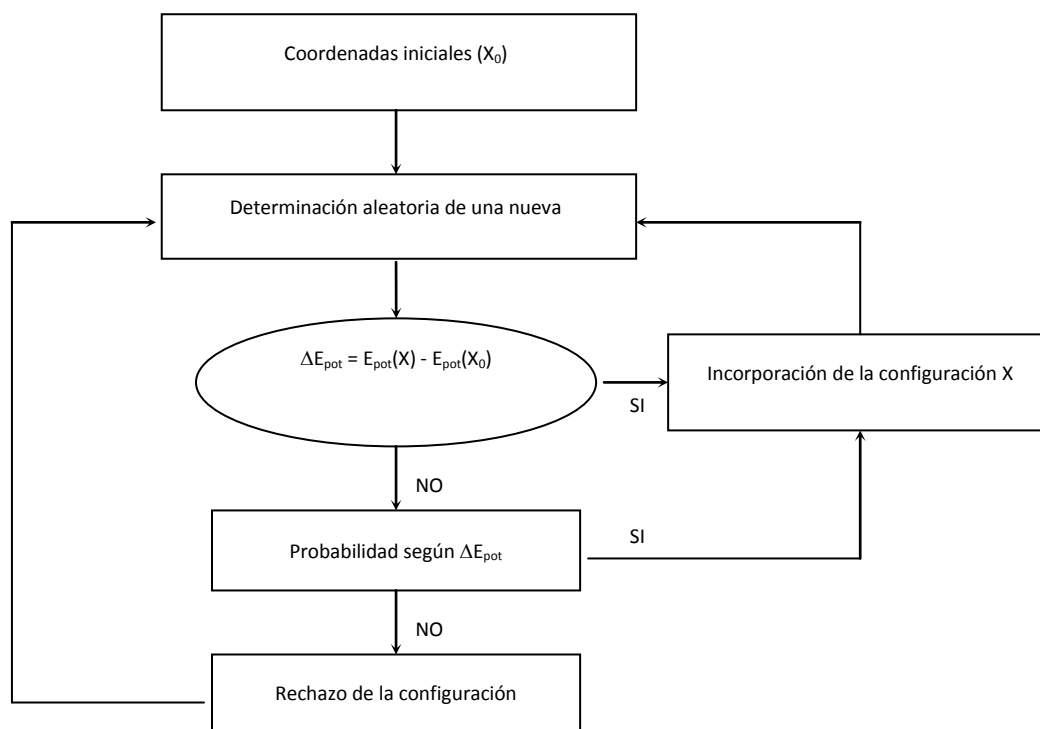


Figura 3.4-1 - Algoritmo básico de Metrópolis.

Ergodicidad

Uno de los objetivos centrales de la simulación molecular es el de obtener resultados que se puedan comparar con el experimento. En general, el puente entre las simulaciones y los observables es la termodinámica estadística. En este contexto y desde un punto de vista estricto, un muestreo en el espacio de las fases es válido si es ergódico, es decir, que todo punto accesible del espacio configuracional puede ser

alcanzado en un número finito de pasos desde cualquier otro punto. Otro modo de enunciarlo es que el promedio temporal de las propiedades de un sistema es igual al promedio sobre el colectivo. Esto se cumple en el límite de simulación infinita pero no siempre en las simulaciones reales. Más aún, el rendimiento ergódico decae a medida que el sistema crece en el número de componentes dado que los grados de libertad se multiplican. Por esta razón es importante ser cautos a la hora de asumir la validez de los muestreos y la correspondencia entre muestreos hechos con técnicas distintas. Es importante tener una clara noción del tiempo o cantidad de muestro necesario en cada caso para obtener resultados confiables.

3.5. Perfiles de energía libre asociados a la migración de ligandos en proteínas

La estimación del cambio de energía libre de un proceso requiere la evaluación precisa de la energía del sistema, así como la realización de un muestreo lo más completo posible. Es decir, se debe calcular tanto la energía como la entropía con el mejor compromiso entre calidad y eficiencia. Puesto que el poder de cálculo es un recurso limitado, existe siempre el problema de que al mejorar la descripción energética se pierde la capacidad de cálculo del término entrópico y viceversa.

Un sistema puede encontrarse en diferentes estados estables que corresponden a mínimos de energía, pero el camino de un mínimo a otro puede estar impedido por una barrera energética alta que dificulta el paso de un estado a otro a lo largo de una simulación por MD. Así, puede suceder que el sistema se encuentre atrapado en un mínimo de energía potencial y que nunca visite el otro estado de interés en tiempos de simulación asequibles. Por ello, se han desarrollado diversas estrategias para poder evaluar el perfil de energía libre asociado a diversos procesos, como Umbrella Sampling y Thermodynamic Integration.

Dado que en la presente memoria se realizaron cálculos de energía libre asociados a la migración de ligandos bimoleculares en proteínas se centrarán las siguientes secciones en una descripción de los principios de los métodos utilizados: i) Steered Molecular Dynamics / Jarzynski, y ii) Implicit ligand Sampling. Por último, se incluye en esta sección Fpocket dado que se utiliza con los mismos fines (estudio de migración

de ligandos pequeños en proteínas), aunque no permite estimar energías libres, sino que se basa en criterios geométricos.

3.5.1 Método de Steered Molecular Dynamics acoplado a la ecuación de Jarzynski

El método consiste en aplicar al sistema una fuerza que guíe el proceso a lo largo de la coordenada de interés. Evidentemente al aplicar una fuerza el sistema se hallará fuera del equilibrio, obteniendo de este modo el trabajo irreversible asociado. Esto permite acelerar el muestreo que describe al proceso a costa de que la trayectoria esté fuera del equilibrio.

Para un proceso que viene descrito por el cambio en un parámetro λ desde λ_0 a tiempo 0 hasta λ_t a tiempo t , el trabajo promedio realizado en el sistema no puede ser inferior a la diferencia de energía libre de los estados correspondientes a λ_0 y λ_t (eq. 3.5-1). Esto es consecuencia de la irreversibilidad de los trabajos realizados en tiempos finitos. Por lo tanto, el trabajo asociado a un proceso de no equilibrio proporciona una cota superior a la diferencia de energía libre. Jarzynski demostró que si se toman todas las simulaciones fuera del equilibrio que llevan al sistema de λ_0 a λ_t se obtiene la energía libre como un promedio de Boltzman de los trabajos irreversibles asociados⁴⁷ (ec. 3.5-2). Es decir, es una técnica que permite obtener propiedades de equilibrio a partir de simulaciones o experimentos fuera del equilibrio

3.5-1

$$\Delta F = F(\lambda_0) - F(\lambda_t) \leq \langle w \rangle$$

3.5-2

$$\exp[-\Delta A(\lambda)/k_B T] = \langle \exp[-W(\lambda)/k_B T] \rangle$$

donde $A(\lambda)$ representa la energía libre, $W(\lambda)$ es el trabajo externo realizado sobre el sistema para ir desde λ_0 a λ_t , y k_B es la constante de Boltzmann.

La perturbación añadida al potencial original es dependiente del tiempo y mueve el sistema a lo largo de la coordenada de reacción. La implementación utilizada en esta memoria para el estudio de migración de ligandos bimoleculares dentro de la matriz proteica modifica el potencial de la forma

$$V'(r) = V(r) + k[r - (\lambda_0 + v\Delta t)]^2$$

donde k es la constante de fuerza con la que forzamos la posición del sistema (por ejemplo la partícula en un canal dentro de la matriz proteica), λ_0 la distancia inicial, v la velocidad con la que se mueve la partícula y Δt el incremento de tiempo entre cada paso.

El problema de éste método es el coste computacional que implica, dado que en el término exponencial de la ecuación de Jarzynski los trabajos pequeños son los que tendrán mayor contribución al hacer el promedio, y éstos son difíciles de obtener en las condiciones de simulación. Por lo tanto, es necesario realizar un número elevado de simulaciones independientes para poder obtener un promedio convergido²⁹.

3.5.2 Implicit Ligand Sampling (ILS)

El enfoque de esta metodología está basado en el cálculo del potencial de fuerza media (Potential of Mean Force, PMF) que corresponde a la colocación de un pequeño ligando (O_2 , CO) en cada punto del interior de la matriz de una proteína. Este PMF está asociado a la energía libre necesaria para incorporar el ligando en una posición particular y, por lo tanto, describe cuáles son las regiones más y menos favorables para la presencia del ligando.

ILS no trata al ligando explícitamente sino de modo implícito. Esto significa que la presencia del ligando en la matriz proteica se trata como una perturbación. La metodología se sustenta en el hecho de que ligando pequeños sin carga o dipolo importante interactúan débilmente con la proteína y prácticamente no modifican la dinámica de ésta. Basado en esta hipótesis se utiliza una MD realizada en ausencia del ligando y luego se calcula la interacción del ligando como una perturbación en cada posición en la proteína y para cada foto de la MD. Esto permite obtener una estimación de la energía libre en todos los puntos del espacio simultáneamente en un solo cálculo y con una sola MD con suficiente muestreo. En este sentido es un método relativamente económico en términos de poder de cálculo.

El PMF o la energía libre de colocar al ligando en un punto r , $G(r)$, se obtiene mediante la ecuación 3.5-4 y puede ser asociada directamente con la probabilidad de encontrar al ligando en esa posición.

3.5-4

$$G(r) = -k_b T \ln \sum_M \sum_C e^{\frac{-\Delta E(q,r)}{MC}}$$

donde M es el número de fotos de la MD, C distintas orientaciones del ligando y $\Delta E(q,r)$ es la energía de interacción entre la proteína en la conformación q y el ligando en la posición r . La energía de interacción se trata como un potencial de Lennard-Jones^{48,49}.

En la práctica el sistema se divide en celdas que forman una red tridimensional cuya resolución se debe escoger según la precisión que se requiera. Normalmente la resolución está en celdas de entre 0.5 y 1 Å. Cuanto más fina sea la grilla más suaves serán las isosuperficies de energía libre que permiten obtener sitios de unión y canales a lo largo de la matriz proteica. Otro parámetro que debe escogerse es el número de estructuras que se utilizarán para obtener la energía libre en cada punto. Para obtener resultados razonables se utilizan entre 1000 y 10000. Según el ligando bimolecular para el cual se quiera obtener la energía libre, se deben especificar los parámetros asociados al mismo: la distancia entre los átomos que lo forman y los valores para σ y ϵ , parámetros del potencial de LJ. En la presente memoria los cálculos por ILS se realizaron con el programa VMD^{49,50}.

3.5.3 FPocket

Esta metodología, a diferencia de las otras presentadas en esta sección, cae en la categoría de métodos geométricos dado que no se basa en un modelo de energía^{51,52}. Es rápido y permite obtener un mapa tridimensional de cavidades y túneles tal como lo hace ILS, pero basado en las características estructurales del sistema.

Primeramente se realiza una descomposición de Voronoi (Voronoi tessellation) usando los átomos como puntos en el espacio. Los diagramas de Voronoi tienden a estar involucrados en situaciones donde el espacio debe ser particionado en “esferas de

influencia” y parece gobernar muchos procesos de crecimiento y distribución en la naturaleza. Una vez obtenidos los nodos Voronoi, fpocket los filtra de acuerdo a un radio mínimo y uno máximo de las esferas alfa asociadas. Una esfera alfa es aquella centrada en un nodo Voronoi en cuyo borde se encuentran cuatro átomos y ninguno en su interior (Figura 3.5-1). Luego se realiza un clusterizado (agrupado) para agrupar esferas alfa cercanas y formar así un pocket (set de esferas alfa). A cada pocket se le da un puntaje y se establece así un ranking de acuerdo al puntaje de cada uno.

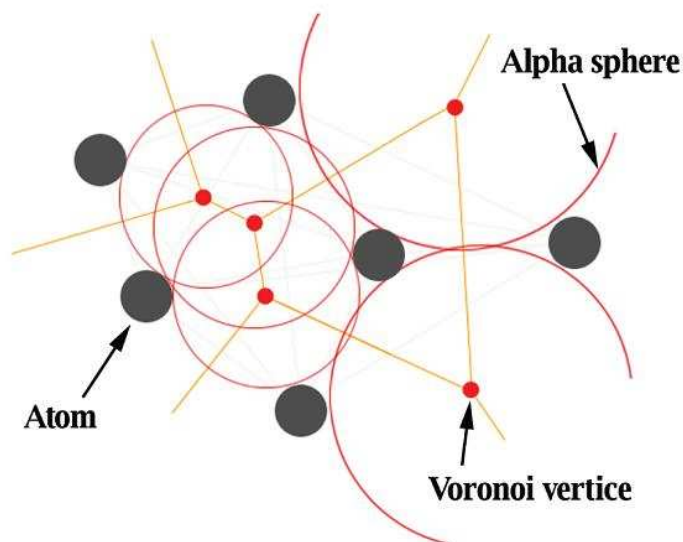


Figura 3.5-1 - Representación de esfera alfa en 2D

Este método puede adaptarse al análisis del comportamiento dinámico de las cavidades y túneles mediante el examen de una colección de estructuras extraídas de MD (mdpocket)³⁴. Así, ello permite calcular una densidad de cavidades para cada punto del espacio. Aquellas zonas con alta densidad corresponden a cavidades conservadas a lo largo de la MD, mientras que densidades bajas denotan cavidades transitorias.

3.6. Solvatación

El medio juega un papel clave en la determinación de propiedades y de la reactividad de sustancias en fases condensadas. Para poder describir el proceso de inserción de una molécula en un medio condensado, es necesario precisar qué entendemos por solvatación y dotar al proceso de una base termodinámica.

3.6.1 Descripción termodinámica

La solvatación se define como el trabajo mediante el cual el soluto se transfiere de manera reversible desde una posición fija en fase gas, a la misma posición en solución a temperatura y presión constantes y concentración generalmente $1M^{53}$. Dado que el proceso termodinámico se define a presión y temperatura constantes, el potencial termodinámico necesario para describir el proceso es la energía libre de Gibbs G , de ahora en adelante energía libre de solvatación. Por ello al trabajo reversible necesario para transferir el soluto en las condiciones anteriores, se le conoce como energía libre de solvatación ΔG_{sol} . La energía libre estándar ΔG_{sol}^0 se puede determinar a partir de la relación de concentraciones de equilibrio del soluto en fase gas y solución mediante:

3.6-1

$$\Delta G_{sol}^0(A) = \lim_{[A]_{sol} \rightarrow 0} \left\{ -RT \ln \left(\frac{[A]_{sol}}{[A]_{gas}} |_{eq} \right) \right\}$$

En esta sección describiremos brevemente el formalismo del modelo continuo IEF-MST, que está basado en el modelo PCM^{54,55}.

3.6.1 Modelos de solvatación continuos

En todos estos modelos la energía libre de solvatación se descompone en tres contribuciones, cada una de ellas asociada a una fase del proceso de solvatación.

3.6-2

$$\Delta G_{sol} = \Delta G_{cav} + \Delta G_{vdw} + \Delta G_{ele}$$

donde ΔG_{cav} es el trabajo reversible necesario para crear una cavidad en el solvente suficientemente grande para alojar el soluto, ΔG_{vdw} es el trabajo reversible necesario para activar las interacciones de dispersión-repulsión entre soluto y solvente, y ΔG_{ele} es el trabajo reversible para generar la distribución de carga del soluto dentro del solvente.

Las dos primeras componentes son especialmente importantes para estudiar la solvatación en solventes poco polares y constituyen lo que suele conocerse por

contribución no electrostática. La última es particularmente importante para solutos polares o cargados en solventes polares debido a la fuerte polarización del solvente, que se modeliza por un medio continuo y uniforme de constante dieléctrica en los modelos de solvente implícito (continuo).

Contribución electrostática a la energía libre de solvatación

Cuando se introduce el soluto en el solvente, la distribución de carga del soluto perturba al solvente, que reacciona a su vez creando un campo eléctrico opuesto al generado por el soluto denominado campo de reacción. En el marco de la teoría de respuesta lineal, el tratamiento cuántico del proceso se lleva a cabo modelizando la respuesta del solvente mediante un operador perturbativo añadido al Hamiltoniano en vacío del soluto (ecuación 3.6-3)

3.6-3

$$(\mathbf{H}_0 + \mathbf{V}_R) |\Psi_{\text{sol}}\rangle = E |\Psi_{\text{sol}}\rangle$$

donde \mathbf{H}_0 es el Hamiltoniano en vacío del soluto y \mathbf{V}_R es el operador perturbación que da cuenta del campo de reacción del solvente. F

Finalmente la contribución electrostática a la energía libre de solvatación se obtiene mediante la diferencia⁵⁶:

3.6-4

$$\Delta G_{\text{e}} = \langle \Psi_{\text{s}} | (\mathbf{H}_0 + \mathbf{V}_R) | \Psi_{\text{s}} \rangle - \frac{1}{2} \langle \Psi_{\text{s}} | \mathbf{V}_R | \Psi_{\text{s}} \rangle - \langle \Psi_0 | \mathbf{H}_0 | \Psi_0 \rangle$$

donde el factor 1/2 es debido a que para polarizar el solvente, se asume que solo se emplea la mitad de la energía de interacción soluto-solvente.

En el modelo PCM, desarrollado por Miertus, Scrocco y Tomasi^{54,55}, el campo de reacción se modeliza por un conjunto de cargas puntuales no polarizables repartidas por la superficie del soluto. Esta superficie se divide en fragmentos suficientemente

pequeños (teseras) como para que la densidad superficial de cada uno α_k se pueda considerar constante. En este caso, el operador perturbativo es simplemente:

3.6-5

$$V_R = \sum_j \frac{q_j}{|\mathbf{r} - \mathbf{r}_j|}$$

donde q_j es el conjunto de cargas correspondientes a cada tesera.

Para la obtención de las cargas q_j se utiliza el formalismo IEF (Integral Equation Formalism) que, respecto al modelo original MST(PCM), presenta la ventaja de un mejor tratamiento de la renormalización de carga, contribuyendo a minimizar su impacto en el término electrostático. La aproximación IEF, basada en el formalismo de ecuaciones integrales y en el uso de funciones de Green, presenta la ventaja de estar menos afectada por la densidad de carga que se extiende más allá de la cavidad, permitiendo así estudiar medios dieléctricos homogéneos isótropos y anisótropos, así como soluciones iónicas. Así, la determinación de las cargas superficiales tiene lugar mediante la resolución de un sistema de ecuaciones lineales:

3.6-6

$$A\sigma = -g$$

3.6-7

$$A = (2\pi - D_e)S_i + S_e(2\pi + D_i^\dagger)$$

3.6-8

$$g = [(2\pi - D_e) - S_e S_i^{-1}(2\pi - D_i)]V_M$$

donde σ es la densidad superficial de carga, los subíndices i y e designan las zonas interior y exterior de la cavidad respectivamente, la matriz D contiene las componentes normales del campo electrostático que las cargas superficiales ejercen entre sí, y por último la matriz S contiene el potencial de reacción que se ejercen las cargas entre sí.

Contribución de cavitación a la energía libre de solvatación

Este término tiene en cuenta el trabajo necesario para crear en el solvente una cavidad suficientemente grande para introducir el soluto. Su evaluación se ha efectuado usando el tratamiento de partículas escaladas de Pierotti⁵⁷. En ésta, se trata el solvente como un fluido de esferas duras, y la energía de cavitación se obtiene como un desarrollo en serie de potencias de la distancia de contacto atómica soluto-solvente.

3.6-9

$$\Delta G_{cav} = K_0 + K_1 R_{MS} + K_2 R_{MS}^2 + K_3 R_{MS}^3$$

donde las K_x dependen de la presión y la temperatura y R_{MS} es la suma de los radios del soluto y el solvente respectivamente.

Este modelo funciona bastante bien para solventes y solutos pequeños y esféricos. Sin embargo para casos no esféricos se hace necesario definir radios efectivos para las moléculas de solvente. Por ello, Claverie sugirió el uso de una corrección en función de la exposición de cada átomo al solvente⁵⁸:

3.6-10

$$\Delta G_{cav} = \sum_i \frac{S_i}{4\pi R_i^2} \Delta G_{cav}(\mathbf{R}_i)$$

donde S_i y R_i son el área y el radio respectivos del átomo i -ésimo.

Contribución de dispersión-repulsión a la energía libre de solvatación.

Este término generalmente suele expresarse como una suma de contribuciones atómicas, las cuales son proporcionales a la superficie expuesta del átomo.

3.6-11

$$\Delta G_{vdw} = \sum_i \gamma_i A_i$$

donde las constantes de proporcionalidad γ_i se denominan tensiones superficiales.

Dichas tensiones se parametrizan a partir de un conjunto de datos experimentales, de forma que el modelo reproduzca lo mejor posible las energías libres de solvatación de los compuestos parametrizados. Estos conjuntos de parametrización se escogen de forma que estén representados varios grupos funcionales en un solvente dado, con la esperanza de que los parámetros obtenidos sean transferibles.

3.7. Descomposición estructural de modos normales (NSD)

Un factor que afecta a la reactividad del grupo hemo es su distorsión. Jentzen et al.⁵⁹ propusieron una clasificación sistemática para la estructura de una hemo distorsionado respecto a uno de referencia de estructura D_{4h} ^{60,61} basada en los modos normales (Figura 3.7-1) más relevantes fueran del plano (saddling, ruffling, doming, X-waving, Y-waving and propellering) y en el plano (meso-stretching, N-pyrrole stretching, pyrrole translation (X,Y), breathing and pyrrole rotation). Se ha demostrado que diferentes familias de hemoproteínas conservan distorsiones similares⁵⁹.

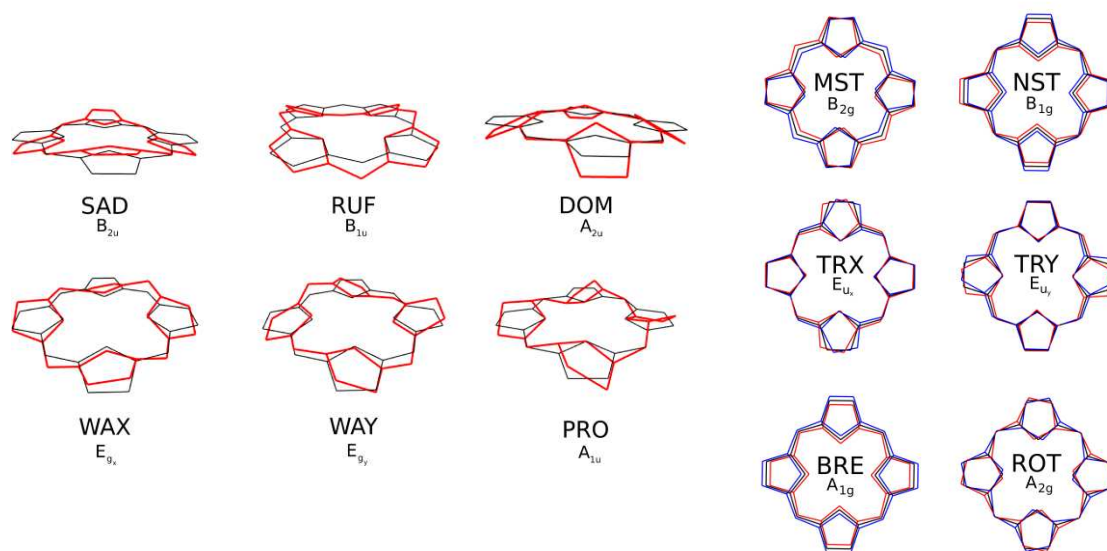


Figura 3.7-1 - Distorsiones del hemo en NSD. Izquierda: Fuera del plano (rojo); doming (*dom*), ruffling (*ruf*), saddling (*sad*), waving (*wax*, *way*) y propellering (*pro*). Derecha: En el plano, positivas (rojo) y negativas (azul): meso-stretching (*mst*), N-pyrrole stretching (*nst*), pyrrole translation (*trx*, *try*), breathing (BRE) y pyrrole rotation (*rot*). La referencia D_{4h} en negro

En el marco de NSD la distorsión de un anillo porfirínico (\mathbf{D}_{obs} ; ec. 3.7-1) se cuantifica según los desplazamientos observados para los átomos de C y N en la estructura del macrociclo respecto a una referencia \mathbf{D}_{4h} (\mathbf{S}_{ref} ; ec. 3.7-1). Este \mathbf{D}_{obs} puede ser escrito como una combinación lineal de los desplazamientos de los modos normales ($\hat{\mathbf{D}}_i$; ec. 3.7-2).

3.7-1

$$\mathbf{D}_{\text{obs}} = \mathbf{S}_{\text{obs}} - \mathbf{S}_{\text{ref}}$$

3.7-2

$$\mathbf{D}_{\text{obs}} = \sum_i d_i \hat{\mathbf{D}}_i$$

donde d_i es el coeficiente asociado con el desplazamiento del modo normal $\hat{\mathbf{D}}_i$.

Esta base mínima de 12 coordenadas colectivas ortonormales está formada por los modos de menor frecuencia y permite describir cualquier hemo distorsionado de un modo natural y sin la necesidad de utilizar los 66 grados de libertad internos. Utilizar una base mínima significa una importante reducción del espacio de las variables con un costo pequeño de disminución de exactitud en la descripción de la distorsión

3.8. Bibliografía

- (1) Feynman, R. Lectures on Physics, 1963.
- (2) Slater, J. C. Quantum Theory of Matter; McGraw-Hill: New York, 1968.
- (3) Möller, C., M. S. Plesset, M. S. Phys. Rev. **1934**, 46, 618.
- (4) Bransden, B. H., Joachain, C. J. Introduction to Quantum Mechanics; Addison Wesley Longman Limited, 1997.
- (5) Hohenberg, P. K., W. Phys. Rev. B **1964**, 136, B864.
- (6) Kohn, W. S., L. J. . Phys. Rev. **1965**, 140, 1133.
- (7) Becke, A. D. Phys. Rev. A **1988**, 38, 3098.
- (8) Lee, C. T. Y., W. T.; Parr, R. G. Phys. Rev. B **1988**, 37, 785.
- (9) Dewar, M. J. S. T., W. J Am Chem Soc **1977**, 99, 4899.
- (10) Dewar, M. J. S., Zoebisch, E. G., Healy, E. F. and Stewart, J. J. P., . J Am Chem Soc **1985**, 107, 3902.
- (11) Stewart, J. J. P. J. Comput. Chem. **1989**, 10, 209.
- (12) Stewart, J. J. P. J. Comput. Chem. **1989**, 10, 221.
- (13) Rocha, G. B. F., R. O.; Simas, A. M.; Stewart, J. J. P. J. Comput. Chem. **2006**, 27, 1101.
- (14) Stewart, J. J. P. J. Mol. Model. **2007**, 13, 1173.
- (15) Cramer, J. C. Essentials of Computational Chemistry: Theories and Models, 2nd ed.; Wiley, 2004.
- (16) Carpentry, J. E. W., F. J. Mol. Struct (Theochem) **1988**, 169, 41.
- (17) Leach, A. R. Molecular Modelling: Principles and applications, 2nd edition ed.; Pearson Education (Prentice Hall): Essex, 2001.
- (18) Schlick, T. Molecular Modeling and Simulation. An interdisciplinary Guide; Springer-Verlag New York, Inc: New York, 2002.
- (19) Weiner, P. K., Kollman, P. A. J. Comput. Chem. **1981**, 2, 287.
- (20) Bayly, C. I., Cieplak, P., Cornell, D. W., Kollman, P. A. J. Phys. Chem. **1993**, 97, 10269.
- (21) Hornak, V., Okur, A., Rizzo, R. C., Simmerling, C. Proteins **2006**, 65, 712.
- (22) Stweart, A. A., McCammon, J. A. Chem Rev. **2006**, 106, 1589.
- (23) Karplus, M., McCammon, A. Nature Structural Biology **2002**, 9, 646.
- (24) Ryckaert, J. P., Cicotti, G., Berendsen, H. J. C. J. Chomp. Phys. **1977**, 23, 327.
- (25) Berendsen, H. J. C., Postma, J. P. M., vzn Gunsteren, W. F., Di Nola, A., Haak, J. R. J. Chem. Phys. **1984**, 81, 3684.
- (26) Hoover, W. G. Phys Rev A **1985**, 31, 1695.
- (27) Ewald, P. P. Ann. Phys. **1921**, 64, 253.
- (28) Darden, T. L., York, D., Pedersen, L. J. Chem. Phys. **1993**, 98, 10089.
- (29) Daniel, R. M., Dunn, R. V., Finney, J. L., Smith, J. C., . Annu. Rev. Biophys. Biomol. Struct. **2002**, 32, 69.
- (30) Eisenmesser, E. Z., Bosco, D. A., Akke, M., Kern, D. Science **2002**, 295, 1520.
- (31) Hinsen, K., Thomas, A., Field, J. Proteins **1999**, 34, 369.
- (32) Luo, J., Bruice, T.C. P Natl Acad Sci USA **2004**, 101, 13152.
- (33) Ma, J., Karplus, M. P Natl Acad Sci USA **1998**, 95, 8502.
- (34) Remy, Y., Wilson, I. A., Michnik, S. W. Science **1999**, 283, 990.
- (35) Sacquin-Mora, S., Lavery, R. Biophys. J. **2006**, 90, 2706.

- (36) Yang, L.-W., Bahar, I. *Structure* **2005**, 893.
- (37) Go, N., Noguti, T., Nishikawa, T. *P Natl Acad Sci USA* **1983**, 80, 3696.
- (38) Levitt, M., Sander, C., Stern, P.S. *J Mol Biol* **1985**, 181, 423.
- (39) Cui, Q., Bahar, I., eds. *Normal Mode Analysis. Theory and applications to biological and chemical systems*; Chapman & Hall/CRC: London.
- (40) Amadei, A., Linssen, A. B., Berendsen, H. J. *Proteins* **1993**, 17, 412.
- (41) Hess, B. *Phys. Rev. E Stat. Nonlin. Soft. Matter. Phys.* **2000**, 8438.
- (42) Hess, B. *Phys. Rev. E Stat. Nonlin. Soft. Matter. Phys.* **2002**, 65.
- (43) Perez, A., Blas, J. R., Rueda, M., López-Bes, J. M., de La Cruz, X., Luque, F. J., Orozco, M. J. *Chem. Theor. Comp.* **2005**, 1, 790.
- (44) Monard, G., Mertz, K. M. *Acc. Chem Res.* **1999**, 32, 904.
- (45) Gao, J., Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, 53, 457.
- (46) Metropolis, N. R., A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. J. *Chem. Phys.* **1953**, 21, 1087.
- (47) Jarzynski, C. *Phys. Rev. Lett.* **1997**, 78, 2690.
- (48) Cohen, J.; Arkhipov, A.; Braun, R.; Schulten, K. *Biophys. J.* **2006**, 91, 1844.
- (49) Cohen, J.; Olsen, K. W.; Schulten, K. Finding Gas Migration Pathways in Proteins Using Implicit Ligand Sampling. In *Methods in Enzymology*; Poole, R. K., Ed., 2008; Vol. 437; pp 439.
- (50) Humphrey, W., Dalke, A. and Schulten, K. J. *Molec. Graphics* **1996**, 14, 33.
- (51) Le Guilloux, V.; Schmidtke, P.; Tuffery, P. *BMC Bioinformatics* **2009**, 10, 168.
- (52) Schmidtke, P.; Le Guilloux, V.; Maupetit, J.; Tuffery, P. *Nucleic Acids Res* **2010**.
- (53) Ben-Naim, A. *Solvation Thermodynamics*; Plenum: New York, 1987.
- (54) Miertus, S., Scroco, E; Tomasi, J. J. *Chem. Phys.* **1981**, 55, 117.
- (55) Miertus, S., Tomasi, J. J. *Chem. Phys.* **1982**, 65, 239.
- (56) Yomosa, S. J. *Phys. Soc. Japan* **1973**, 35, 1738.
- (57) Pierotti, R. A. *Chem. Rev.* **1976**, 76, 717.
- (58) Claverie, P., Daudey, J.P. , Langlet, J., Pullman, B., Piazzola, D., Huron, M.J. J. *Phys. Chem.* **1978**, 82, 405.
- (59) Jentzen, W. S., M. C.; Hobbs, J. D.; Song, X. Z.; Ema, T.; Nelson, N. Y.; Medforth, C. J.; Smith, K. M.; Veyrat, M.; Mazzanti, M.; Ramasseul, R.; Marchon, J. C.; Takeuchi, T.; Goddard, W. A. III; Shelnut, J. A. *J Am Chem Soc* **1995**, 117, 11085.
- (60) Jentzen, W. S., X.Z.; Shelnut, J. A. *J. Phys. Chem. B* **1997**, 101, 1684.
- (61) Shelnut, J. A.; Song, X. Z.; Ma, J. G., Jia, S. L.; Jentzen, W.; Medforth, C. J. . *Chem. Soc. Rev.* **1998**, 27, 31.

Capítulo 4

Metodología Multinivel

Objetivos

El desarrollo de la metodología Multinivel y su aplicación serán presentadas en tres partes:

- Extensión del modelo de solvatación continuo MST al hamiltoniano semiempírico RM1
- Desarrollo de la metodología Multinivel Monte Carlo-RM1/MST acoplada a cálculos ab initio
- Aplicación de la metodología a histamina

4.1. Extensión del modelo de solvatación MST al hamiltoniano semiempírico RM1

Los trabajos previos de parametrización de distintos hamiltonianos al modelo de solvatación MST llevados adelante en nuestro grupo¹ utilizan la estrategia de cavidad dual. En ella los componentes no electrostáticos de la energía libre de solvatación (dispersión-repulsión) son determinados usando una superficie molecular construida a partir de radios atómicos de van der Waals, mientras el término electrostático se determina escalando la superficie expuesta a solvente. Este factor (κ) generalmente se ajusta a partir de comparación de la respuesta electrostática determinada por cálculos cuánticos Self Consistent Reaction Field (QM SCRF) MST y la componente electrostática obtenida por simulaciones clásicas discretas (dinámica molecular o monte carlo).

El factor de escalado depende del solvente y los radios adoptados en parametrizaciones previas² se han mantenido en este trabajo (Tabla 4.1-1). Por último, se ha utilizado el procedimiento GEPOL para determinar la cavidad que separa soluto y solvente de cara a determinar las distintas contribuciones a la energía libre.³

Tabla 4.1-1. Parámetros definidos para cálculos MST/RM1 en agua, octanol, cloroformo y tetracloruro de carbono.

Factor de escalado (κ)			
Agua	Octanol	Cloroformo	Cloruro de carbono
1.25	1.50	1.60	1.80
Radio atómico (R) en Å			
General		Específico para agua ^b	
H ^a	1.20 (0.80)	N in -N-	
C	1.50		
N	1.50	N in -N-	1.40
		N in -NH ₂	1.55
O	1.40	O in -OH	1.45
F	1.35		
P	1.80		
S	1.75	S in -SH	1.85
Cl	1.80		
Br	1.95		

^a Entre paréntesis el radio para un átomo de hidrógeno unido a heteroátomo.

^c Radio modificado para grupos químicos específicos en solución acuosa

En línea con los trabajos previos, la extensión del modelo MST al hamiltoniano RM1 siguió la misma estrategia dual, completándose una parametrización para compuestos neutros en los cuatro solventes habituales (agua, octanol, cloroformo y tetracloruro de carbono) y para iones en agua. Los cálculos RM1 fueron llevados a cabo usando una versión del MOPAC 6.0 modificada localmente.

4.1.1 Compuestos neutros

Un total de 81, 66, 57 y 50 compuestos conteniendo grupos funcionales (bio)orgánicos prototipo han sido considerados para llevar adelante la parametrización

del modelo MST/RM1 para agua, octanol, cloroformo y tetracloruro de carbono, respectivamente. La validez estadística del modelo fue valorada a posteriori por comparación de las energías de solvatación calculadas y experimentales para un conjunto de prueba de compuestos.

La parametrización del modelo MST para el hamiltoniano RM1 se llevó adelante intentando reducir al mínimo el número de parámetros ajustables para una mayor transferibilidad y con el objeto de evitar artefactos por “sobreentrenamiento”. La estrategia se limita al ajuste de las tensiones superficiales (γ , ver Métodos) por comparación entre energías libres de solvatación experimentales y calculadas. Se realiza una regresión lineal múltiple para obtener la componente de van der Waals de la energía libre de solvatación sustrayendo los componentes de cavitación y electrostático a los valores experimentales, tal como se expresa en la ecuación 4.1-1

4.1-1

$$\Delta G_{\text{vw}} = \Delta G_{\text{sol}}^{\text{expt}} - \Delta G_{\text{ele}} - \Delta G_{\text{cav}}$$

La Tabla 4.1-2 muestra las energías libres de solvatación en los cuatro solventes para los compuestos del set de entrenamiento y la Tabla 4.1-3 los parámetros estadísticos de la comparación de las energías libres de solvatación experimentales y teóricas (para más información, véase la Información Suplementaria de la referencia⁴)

Tabla 4.1-2. Energía libre de solvatación (kcal/mol) determinada por cálculos MST/RM1 para el set de entrenamiento en agua, octanol, cloroformo y tetracloruro de carbono (calculada/experimental).

Compuesto	Agua	Octanol	Cloroformo	Tetracloruro de carbono
C, H				
Etano	1.4/1.8	-0.9/-0.6		
Neopentano		-1.7/-1.7		
Ciclohexano	1.9/1.2	-3.3/-3.7		
n-Hexano		-3.1/-3.3		
n-Heptano		-3.6/-4.1		
n-Octano		-4.2/-4.6		
Etileno	0.1/1.3	-1.8/-0.3		
2-Metilpropeno	0.5/1.2	-2.6/-2.0		-2.7/-2.6

1-Buteno	0.7/1.4	-2.6/-1.9		-2.9/-2.5
E-2-Penteno	0.8/1.3			-3.5/-3.5
Benzeno	-1.4/-0.9	-4.3/-3.7	-5.1/-4.6	-4.6/-4.5
Tolueno	-1.3/-0.9	-4.8/-4.6	-5.5/-5.5	-5.0/-5.1
m-Xileno	-1.2/-0.8		-6.0/-5.9	-5.4/-5.7
C,H,O				
Agua	-6.5/-6.3	-5.6/-4.4		
Metanol	-5.4/-5.1	-4.4/-3.9	-3.5/-3.4	-2.6/-2.2
Etanol	-5.2/-5.0	-4.7/-4.4	-4.0/-3.9	-3.1/-3.1
n-Propanol	-4.9/-4.8	-5.2/-5.0	-4.6/-4.3	-3.7/-3.7
2-Propanol	-4.2/-4.8		-4.6/-4.3	-3.6/-3.2
n-Butanol	-4.6/-4.7	-5.8/-5.7	-5.3/-5.3	-4.3/-4.2
n-Pentanol	-4.5/-4.5	-6.4/-6.4	-6.0/-5.9	-5.0/-5.0
n-Hexanol	-4.4/-4.4	-6.9/-7.1	-6.6/-6.7	-5.6/-5.7
n-Heptanol	-4.2/-4.2	-7.6/-7.8	-7.3/-7.5	-6.2/-6.5
n-Octanol		-8.1/-8.1		
Fenol	-6.1/-6.6	-7.4/-8.7	-7.0/-7.1	-6.0/-6.1
o-Metilfenol	-5.2/-5.9		-7.2/-7.6	-6.3/-6.8
p-Metilfenol	-5.6/-6.1		-7.3/-7.6	-6.4/-6.1
Dimetileter	-1.3/-1.9	-2.1/-2.1		
Dietileter	-1.6/-1.6		-4.3/-4.3	
Tetrahidrofurano	-2.5/-3.7	-3.9/-3.9		
Anisol	-2.4/-1.0	-5.5/-5.5	-6.5/-6.2	-5.7/-5.5
Acetaldeído	-3.6/-3.5		-4.0/-3.7	
Benzaldeído	-4.1/-4.0	-6.5/-6.1	-7.3/-7.1	-6.4/-6.1
p-hidroxibenzaldeído	-9.1/-10.5		-9.2/-10.4	-7.8/-8.2
Acetona	-3.8/-3.9	-4.1/-3.2	-4.6/-4.3	-3.4/-3.5
Butanona	-3.4/-3.6		-5.1/-5.4	-4.0/-4.1
Acetofenona	-4.6/-4.6		-7.8/-8.4	-6.8/-6.4
Ácido acético	-7.5/-6.7	-6.1/-6.4	-5.4/-4.5	-4.2/-3.4
Ácido propiónico	-6.8/-6.5	-6.4/-6.9	-5.8/-5.2	-4.7/-4.3
Ácido butírico	-6.5/-6.4	-6.9/-7.6	-6.5/-6.0	-5.3/-5.0
Metiletanoato	-3.1/-3.3	-3.9/-3.5	-4.9/-4.9	-3.8/-3.7
Etiletanoato	-3.1/-3.1		-5.5/-5.5	-4.4/-4.4
Metilpropanoato	-2.8/-2.9		-5.4/-5.5	-4.3/-4.4
Metilbenzoato	-3.6/-4.3	-6.9/-7.3	-8.1/-7.8	-7.1/-7.2
C,H,N				
Amoníaco	-3.5/-4.3			
Metilamina	-4.9/-4.6	-3.9/-3.8	-3.5/-3.2	-2.4/-2.5
Etilamina	-4.6/-4.5	-4.4/-4.1	-4.1/-4.0	-3.0/-2.8
n-Propilamina	-4.6/-4.4	-5.0/-4.8	-4.8/-4.7	-3.6/-3.6
n-Butilamina	-4.4/-4.3	-5.5/-5.4	-5.4/-5.1	-4.2/-4.2
Dimetilamina	-3.5/-4.3		-3.5/-3.7	-2.5/-2.8
Dietilamina	-4.1/-4.1	-4.6/-4.8	-4.9/-5.2	-3.8/-4.1
Trimetilamina	-2.2/-3.2	-2.5/-3.6	-3.9/-4.0	-2.6/-3.1
Piperidina	-4.7/-5.1	-5.5/-5.3		

Anilina	-5.8/-5.0		-7.0/-6.8	-5.9/-5.7
Acetonitrilo	-4.0/-3.9	-3.4/-3.2	-4.8/-4.4	
Propanonitrilo	-3.9/-3.9	-3.9/-3.7		
Butanonitrilo	-3.7/-3.6	-4.4/-4.3		
Benzonitrilo	-4.7/-4.1	-6.4/-6.1	-8.0/-7.2	-6.8/-6.3
Piridina	-4.3/-4.7	-4.7/-5.3	-6.2/-6.7	-4.9/-5.0
2-Metilpiridina	-4.2/-4.6		-6.6/-7.0	
4-Metilpiridina	-4.4/-4.9		-6.7/-7.5	
C,H,N,O				
Acetamida	-10.1/-9.7	-7.4/-8.0	-6.5/-7.1	
Nitroetano	-4.3/-3.7	-4.4/-3.9		
Nitropropano	-3.6/-3.3			-4.9/-4.5
Nitrobenzeno	-4.2/-4.1	-6.5/-6.6	-8.5/-7.8	-7.0/-6.9
m-Nitrofenol			-10.1/-10.5	-8.3/-8.8
p-Nitrofenol			-10.5/-11.0	-8.5/-9.4
C,H,S				
Sulfuro de hidrógeno	-1.2/-0.7			
Metanotiol	-1.2/-1.2			
1-Propanotiol		-3.8/-3.5		
Tiofenol	-2.0/-2.6	-5.7/-6.0	-7.6/-7.6	
Dimetiltioeter	-1.4/-1.5	-3.3/-4.2		
Dietiltioeter	-2.0/-1.4	-4.6/-4.1	-5.5/-6.4	
Tioanisol	-2.2/-2.7	-6.5/-6.5	-7.5/-6.0	-5.7/-5.7
Tiofeno		-4.3/-3.9	-5.2/-5.8	
C,H,Halógenos				
Flurometano	-0.4/-0.2			
Tetrafluorometano	4.3/3.2	1.7/1.5		
1,1-Difluoroetano	-0.4/-0.1	-1.6/-1.1		
2,2,2-Trifluoroetano	-6.2/-4.3	-5.0/-4.8	-3.2/-3.0	
Fluorobenzeno		-4.1/-3.9	-4.3/-4.3	-3.6/-3.6
Cloroformo		-3.9/-3.8		
Cloroetano		-3.5/-2.6		
Clorobenzeno	-1.3/-1.1		-5.8/-5.5	-5.4/-5.2
p-Diclorobenzeno	-0.6/-1.0	-6.0/-5.7	-6.4/-6.3	-6.2/-6.3
Fluorotriclorometano		-2.2/-2.6	-2.7/-2.6	
Difluorodiclorometano	3.8/1.7	-0.9/-1.3	-1.2/-1.6	
Diclorometano	-2.0/-1.4			
Bromometano	-1.5/-0.8	-2.9/-2.4		
Bromoetano	-1.8/-0.7	-3.6/-2.9		
Bromobenzeno	-1.7/-1.5	-5.7/-5.5		-6.3/-6.3
p-Dibromobenzeno	-1.3/-2.4	-6.8/-7.5		
1-Bromo-2-Cloroetano	-1.6/-2.0			
p-Bromofenol	-6.9/-7.1		-8.6/-8.6	-7.9/-7.9
C, H, P				

Fosfina 0.6/0.6

Las desviaciones medias con signo (msd) presentadas en la Tabla 4.1-3 son cercanas a 0 para los cuatro solventes y el coeficiente de escalado de la ecuación de regresión entre valor experimental y teórico (c) cercano a la unidad. Esto indica una ausencia de error sistemático para una serie extensa y variada de compuestos neutros. La desviación cuadrática media (rmsd) entre energías de solvatación experimentales y teóricas es de 0.6, 0.5, 0.5 y 0.3 kcal/mol para agua, octanol, cloroformo y tetracloruro de carbono, respectivamente, lo que está muy próximo a las incertidumbres obtenidas con los modelos MST/AM1 y MST/PM3 y para las versiones ab initio del método.

Tabla 4.1-3. Parámetros estadísticos para el ajuste entre energías de solvatación experimentales y teóricas derivadas de cálculos MST/RM1.

Parámetro ^a	Agua	Octanol	Cloroformo	Tetracloruro de carbono
Set de entrenamiento				
n	81	66	57	50
msd	0.0	0.1	0.0	0.0
rmsd	0.6	0.5	0.5	0.3
c	0.99	1.00	1.00	1.00
r	0.99	0.99	1.00	1.00
F	3194	5853	10247	15442
Set de validación				
n	32	24	16	12
msd	0.1	0.0	0.0	-0.1
rmsd	0.7	0.7	0.8	0.6
C	0.97	1.01	0.99	1.02
R	0.99	0.99	0.99	0.99
F	2068	1572	798	851

^a número de compuestos (n); desviación media (msd; kcal/mol); desviación cuadrática media (rmsd; kcal/mol); factor de escalado (c), coeficiente de correlación de

Pearson (r) and valor F.

Si se analizan los distintos componentes de la energía de solvatación (véase la Información Suplementaria de la referencia⁴) se aprecia que la componente electrostática da cuenta en promedio del 132%, 39%, 24% y 11% del total de la energía de solvatación en agua, octanol, cloroformo y tetracloruro de carbono. Consistentemente, el cambio en el momento dipolar inducido al solvatar alcanza alrededor del 40%, 17%, 13% and 6%, lo que está de acuerdo con el cambio del momento dipolar inducido por el solvente predicho por versiones previas del algoritmo MST y de cálculos QM/MM discretos.^{5,6}

Para validar la parametrización del método MST/RM1, se han calculado las energías de solvatación de un conjunto adicional de 84 compuestos (agua: 32; octanol: 24; cloroformo: 16; tetracloruro de carbono: 12; Tabla 4.1-4) que denominamos set de validación. Evidentemente, ninguna de estas moléculas fue considerada en la parametrización. La fiabilidad del modelo MST/RM1 fuera del conjunto de entrenamiento puede ser verificada por la excelente concordancia entre las estimaciones experimentales y teóricas (Tabla 4.1-3 y Figura 4.1-1)

Tabla 4.1-4. Energía libre de solvatación (kcal/mol) determinada por cálculos MST/RM1 para el set de validación en agua, octanol, cloroformo y tetracloruro de carbono (calculada/experimental).

Compuesto	Agua	Octanol	Cloroformo	Tetracloruro de carbono
C, H				
Propano	1.9/2.0	-1.4/-1.3		
Hexano	2.7/2.5			
Ciclopenteno	-0.3/0.6			
Etilbenzeno	-1.1/-0.8	-5.3/-5.1	-6.1/-5.8	-5.5/-5.7
p-Xileno	-1.2/-0.8	-5.3/-5.2		
Naftaleno	-2.5/-2.4	-7.5/-7.0	-7.7/-7.9	-7.0/-7.6
Ciclohexano			-4.0/-4.5	
C, H, O				
2-Metil-2-propanol	-4.0/-4.5	-5.0/-4.8	-4.9/-4.5	-3.8/-3.4
Dioxano	-2.7/-1.8	-4.1/-4.9	-4.9/-6.2	-4.0/-5.0
Metilpropilether	-1.3/-1.7	-3.3/-3.6		

Propanal	-3.7/-3.4	-4.1/-4.1		
Ciclopentanone	-4.2/-4.7	-5.4/-5.0		-4.6/-5.3
Metilmetanoato	-3.0/-2.8	-3.3/-2.8		
2-Propen-1-ol	-5.9/-6.8	-6.0/-5.3	-5.0/-4.3	
p-Cresol		-7.8/-8.8	-7.3/-7.6	
Ácido Butírico		-6.9/-7.6	-6.5/-6.0	-5.3/-4.8
Acetaldehído			-4.0/-3.7	
Ácido Propiónico			-5.8/-5.4	
Benzaldehído				-6.4/-6.1
Butanona				-4.0/-4.1
Metiletanoato				-3.8/-3.8
C, H, N				
Pirrolidina	-5.1/-5.5			
Piperazina	-7.7/-7.4	-7.1/-5.8		
p-Metilnilina	-6.0/-5.6			
Metilpirazina	-4.8/-5.6			
Metiladenina	-12.9/-13.6			
Aminoanilina	-10.9/-9.9			
Anilina		-7.2/-6.7		
2-Metilpiridina		-5.0/-6.1		
4-Metilpiridina		-5.2/-6.6		
Ammonia			-3.1/-2.4	-2.2/-1.1
Piperidina			-5.3/-6.4	
Trimetilamina				-2.6/-3.1
C, H, N, O				
Benzamida	-11.8/-10.9			
Urea	-15.5/-13.8			
Nitrometano	-4.5/-4.0			
Morfolina	-5.5/-7.2		-5.6/-6.7	
Metiltimina	-9.8/-10.4			
Nitropropano		-4.7/-4.4		
C, H, S				
Etanotiol	-1.5/-1.3			
Dimetilsulfuida	-1.4/-1.8			
Dietiltioeter		-4.6/-4.1	-5.5/-6.4	
Tioanisol			-7.5/-7.5	-5.7/-5.7
C, H, Halógenos				
Fluorobenzeno	-0.8/-0.8			
Diclorometano	-1.9/-1.4			
o-Clorotolueno	-0.8/-1.2			
Dibromometano	-1.7/-0.7			
Clorofluorometano	-2.3/-0.8			
Tetrafluorobenzeno		1.7/1.5		
2,2,2-Trifluoroetanol				-5.0/-4.8
Colorobenzeno		-5.3/-5.0		

p-Dicolorobenzeno	-6.0/-5.7	
p-Bromofenol	-9.2/-10.6	
Bromobenzeno		-6.7/-6.1
C, H, P		
Trietilfosfato	-7.9/-7.8	

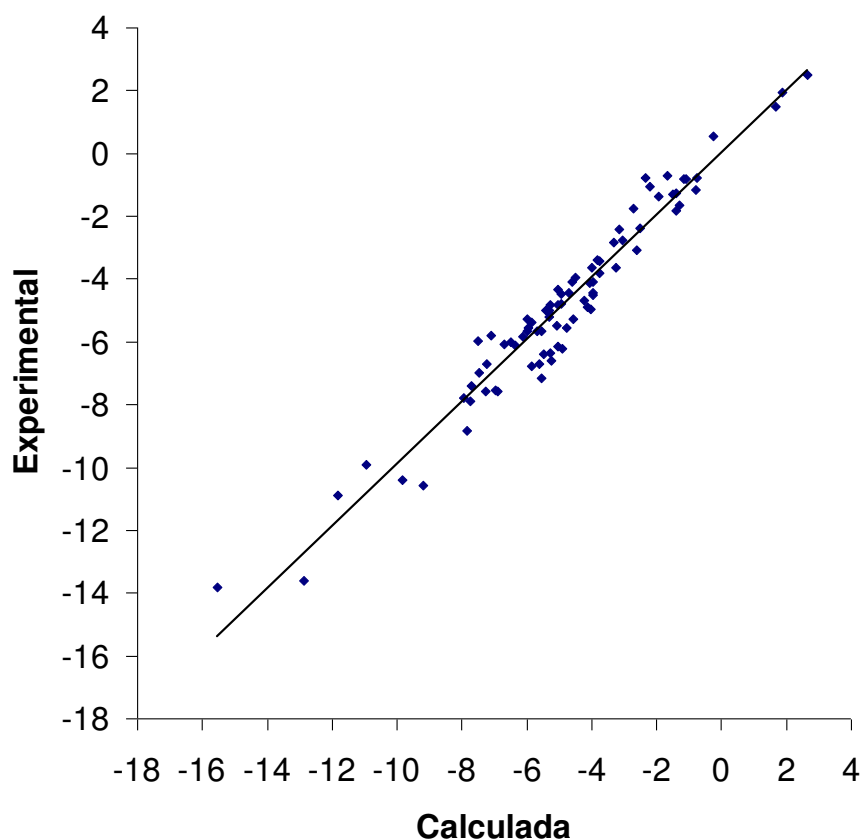


Figura 4.1-1. Energías libres de solvatación experimentales y calculadas determinadas para el set de validación de compuestos en agua, octanol, cloroformo y tetracloruro de carbono

La comparación de las tensiones superficiales optimizadas para el modelo MST/RM1 en los cuatro solventes (Tabla 4.1-5) con aquellas optimizadas para los modelos MST/AM1 y MST/PM3 revela que existe una importante correspondencia, lo que es esperable por la similitud entre los tres hamiltonianos. Las diferencias más importantes se encuentran en los tipos de átomo C y N lo que puede ser atribuido -al menos en parte- al hecho de que ambos, MST/AM1 y MST/PM3, requirieron de la inclusión de un término de corrección para los grupos nitrilo, lo que no fue necesario

para MST/RM1

Tabla 4.1-5. Tensiones superficiales atómicas optimizadas (¶ ¶ para MST/RM1 (MST/AM1 / MST/PM3) en agua, octanol, cloroformo y tetracloruro de carbono.

γ	Agua	Octanol	Cloroformo	Tetracloruro de carbono
C	-0.169 (-0.217/-0.288)	-0.248 (-0.280/-0.280)	-0.161 (-0.239/-0.242)	-0.180 (-0.220/-0.220)
CH	-0.133 (-0.124/-0.125)	-0.197 (-0.192/-0.193)	-0.158 (-0.151/-0.154)	-0.153 (-0.150/-0.151)
CH2	-0.130 (-0.130/-0.127)	-0.194 (-0.194/-0.192)	-0.148 (-0.150/-0.150)	-0.143 (-0.144/-0.144)
CH3	-0.120 (-0.114/-0.120)	-0.182 (-0.178/-0.178)	-0.143 (-0.140/-0.143)	-0.135 (-0.134/-0.135)
N	-0.073 (-0.128/-0.101)	-0.110 (-0.245/-0.204)	-0.182 (-0.256/-0.206)	-0.151 (-0.206/-0.190)
NH	-0.130 (-0.115/-0.145)	-0.227 (-0.325/-0.311)	-0.179 (-0.194/-0.185)	-0.174 (-0.174/-0.175)
O	-0.060 (-0.013/-0.031)	-0.150 (-0.138/-0.105)	-0.154 (-0.134/-0.116)	-0.152 (-0.139/-0.133)
OH	-0.174 (-0.152/-0.159)	-0.258 (-0.273/-0.265)	-0.186 (-0.179/-0.173)	-0.189 (-0.183/-0.182)
F	-0.048 (-0.047/-0.049)	-0.124 (-0.124/-0.118)	-0.074 (-0.077/-0.076)	-0.051 (-0.052/-0.048)
P	-0.074	----	----	----
PH	-0.074	----	----	----
S	-0.104 (-0.079/-0.058)	-0.191 (-0.205/-0.173)	-0.169 (-0.180/-0.156)	-0.120 (-0.119/-0.109)
SH	-0.093 (-0.064/-0.053)	-0.176 (-0.176/-0.157)	-0.188 (-0.187/-0.172)	----
Cl	-0.094 (-0.118/-0.110)	-0.163 (-0.167/-0.175)	-0.129 (-0.128/-0.128)	-0.126 (-0.125/-0.125)
Br	-0.102 (-0.115/-0.108)	-0.163 (-0.175/-0.173)	-0.141 (-0.135/-0.135)	-0.139 (-0.134/-0.135)

Tomando los datos para los cuatro solventes en conjunto, los valores de desviación

cuadrática media (rmsd) entre ΔG_{ele} determinados para RM1, AM1 and PM3 están próximos a 0.5 kcal/mol (Tabla 4.1-6). Sin embargo, el rmsd varía de 0.1 kcal/mol en tetracloruro de carbono a alrededor de 0.9 kcal/mol en agua, lo que es esperable dada la reducción del factor de escalado κ que modula la frontera soluto/solvente, lo que a su vez resulta en grandes diferencias en el término ΔG_{ele} entre los formalismos semiempíricos.

Tabla 4.1-6. Parámetros estadísticos de la comparación entre la componente electrostática de la energía de solvatación en agua, octanol, cloroformo y tetracloruro de carbono derivados de cálculos MST/RM1, MST/AM1 y MST/PM3.

Parámetro ^a	msd	rmsd	C	r	F
Todos los valores					
RM1 vs AM1	0.1	0.5	0.92	0.99	8736
RM1 vs PM3	0.2	0.6	0.93	0.98	5062
AM1 vs PM3	0.1	0.8	0.91	0.97	5319
Agua					
RM1 vs AM1	0.5	0.8	0.91	0.97	1215
RM1 vs PM3	0.3	0.9	0.94	0.94	547
AM1 vs PM3	-0.2	0.8	1.03	0.95	645
Octanol					
RM1 vs AM1	-0.1	0.3	1.06	0.97	919
RM1 vs PM3	0.2	0.5	0.89	0.94	489
AM1 vs PM3	0.3	0.5	0.84	0.97	830
Cloroformo					
RM1 vs AM1	0.0	0.2	1.02	0.98	1107
RM1 vs PM3	0.1	0.3	0.90	0.94	455
AM1 vs PM3	0.1	0.3	0.88	0.96	571
Tetracloruro de carbono					
RM1 vs AM1	0.0	0.1	0.96	0.98	1240
RM1 vs PM3	0.0	0.1	0.89	0.95	452
AM1 vs PM3	0.0	0.1	0.92	0.96	621

^a desviación media (msd; kcal/mol); desviación cuadrática media (rmsd; kcal/mol); factor de escalado (κ), coeficiente de correlación de Pearson (r) and valor F.

Si se obtienen las energías de solvatación MST/RM1 asumiendo la transferibilidad de las ΔG_{vw} MST/AM1 o MST/PM3, la desviación cuadrática media es de 0.5 y 0.3 kcal/mol para cloroformo y tetracloruro de carbono, lo que es comparable a los valores obtenidos para el modelo optimizado MST/RM1 (Tabla 4.1-3). Sin embargo, resultan mayores a 0.5 kcal/mol que aquellas con el modelo optimizado MST/RM1 para los solventes más polares (agua y octanol). Aunque esta no es una diferencia drástica, representa un empeoramiento en la exactitud química, confirmando la necesidad de realizar una parametrización específica para el hamiltoniano RM1.

4.1.2 Compuestos iónicos en agua

La extensión del modelo MST al hamiltoniano RM1 para iones en agua fue llevada a cabo tomando como referencia experimental los 97 compuestos iónicos de la reciente compilación de Kelly et al.⁷, corregida por un factor de 1.9 cal/mol, que da cuenta de la diferencia entre el valor teórico determinado por Zhan y Dixon⁸ para la energía libre de hidratación del protón (usada en ⁹ para obtener las energías experimentales de hidratación de iones) y los valores experimentales más precisos obtenidos por Tissandier et al.¹⁰

De acuerdo a nuestros estudios ab initio previos la cavidad del soluto en agua debía ser modificada reduciendo el factor de escalado κ alrededor de un 10% (i.e., $\kappa = 1.12$) para los átomos (pesados e hidrógenos unidos) que llevan la carga formal.^{11,12} Desafortunadamente se halló que esta estrategia no era adecuada a nivel semiempírico. Por ejemplo, mientras la energía libre de solvatación para CH_3OH_2^+ era predicha como -92.8 kcal/mol, lo que es consistente con un valor experimental de -92.9 kcal/mol, la calculada y experimental para las versiones protonadas de CH_3NH_2 (-81.2 y -75.9 kcal/mol), $\text{O}(\text{CH}_3)_2$ (-71.4 y -79.9 kcal/mol), y CH_3COCH_3 (-66.6 y -76.9 kcal/mol) difieren en 5.3, 8.5 y 10.3 kcal/mol, respectivamente. De manera similar, las energías de hidratación para CH_3O^- and CH_3S^- , que eran predichas como -90.6 y -92.3 kcal/mol, difieren con el valor experimental (-95.1 y -74.1 kcal/mol, respectivamente) en -4.5 y

18.2 kcal/mol, indicando una moderada subestimación para CH_3O^- pero una gran sobreestimación para CH_3S^- .

Estos resultados rechazan fuertemente la posibilidad de utilizar un factor de escalado universal que module la cavidad electrostática asociada a los átomos que llevan la carga formal. Por el contrario, la magnitud de este factor parece ser muy sensible a la naturaleza del átomo pesado y al entorno químico. Este comportamiento no es del todo inesperado, dado que estudios previos habían señalado que AM1 y PM3 predicen afinidades protónicas y basicidades con una dispersión no sistemática de los valores experimentales,^{13,9,14,15} sugiriendo que la capacidad predictiva de aquellos métodos semiempíricos podría ser mejorada utilizando un escalado basado en el entorno químico⁹.

Basándonos en estas consideraciones, realizamos una exploración sistemática para determinar el factor de escalado óptimo para una serie de cationes y aniones monovalentes.

La Tabla 4.1-7 reporta los valores calculados y experimentales para la energía libre de solvatación obtenidos para una serie de 47 cationes, agrupados en 3 categorías (conjuntos C1-C3) dependiendo del valor del factor de escalado. Los datos en la Tabla 4.1-7 indican que existe una relación inversa entre la magnitud del factor de escalado y el número de hidrógenos unidos al átomo protonable. De este modo, para éteres y tioéteres, compuestos con carbonilo y nitrilo, aminas secundarias aromáticas y aminas terciarias o heterociclos como la piridina (conjunto C1), un escalado de 0.99 da el mejor acuerdo entre energías libres de solvatación experimentales y calculadas. Para alcoholes, aminas secundarias y compuestos como el pirrol (conjunto C2), el factor aumenta a 1.15 y, finalmente, para aminas primarias alifáticas o aromáticas no sustituidas (conjunto C3) alcanza el valor de 1.24.

Tabla 4.1-7. Energías de solvatación calculadas y experimentales (kcal/mol) de cationes monovalentes obtenidas adoptando un factor de escalado dependiente del entorno químico para definir la frontera en la cavidad electrostática entre soluto y solvente.

	Calculado	Experimental
Conjunto C1 ($\kappa=0.99$)		
Dimetiléter	-78	-80

Dietiléter	-64	-72
Acetona	-72	-77
Acetofenona	-59	-65
Etiletanoato	-68	-69
Acetamida	-80	-74
Formamida	-88	-83
Trimetilamina	-61	-61
Trietilamina	-49	-55
Acetonitrile	-75	-75
N,N-dietilanilina	-51	-54
N,N-dimetilanilina	-61	-57
Piridina	-63	-61
p-Metil-N,Ndimetilanilina	-59	-56
Quinolina	-59	-56
Dimetiltioéter	-62	-65
Conjunto C2 ($\kappa=1.15$)		
Metanol	-91	-93
Etanol	-83	-88
Azacicloheptano	-58	-63
Azetidina	-66	-68
Aziridina	-68	-71
Dimetilamina	-68	-69
Dietilamina	-59	-64
Morfolina	-75	-70
N-etilanilina	-62	-62
N-metilanilina	-66	-63
Piperazina	-68	-66
Piperidina	-66	-64
Pirrol	-68	-62
Pirrolidina	-64	-66
Conjunto C3 ($\kappa=1.24$)		
1-Aminonaftaleno	-67	-68

t-Butilamina	-61	-67
Metilamina	-75	-76
Propilamina	-67	-72
Butilamina	-74	-72
Alilamina	-75	-72
Ciclohexanamina	-61	-69
Isopropilamina	-65	-70
m-Aminoanilina	-72	-66
m-Cloroanilina	-77	-75
m-Metilanilina	-70	-70
o-Metilanilina	-68	-70
p-Cloroanilina	-78	-74
p-Metilanilina	-68	-70
p-Metoxianilina	-75	-71
p-Nitroanilina	-88	-76
Anilina	-71	-72
msd	-0.2	
rmsd	4.3	
C	1.00	
R	1.00	
F	11876	

Análogamente, factor de escalado óptimo para los aniones muestra una fuerte dependencia con la naturaleza del tipo de átomo, abarcando desde 1,09 para alcoholóxidos y fenóxidos a 1.24 para aniones carboxilato e, inclusive, a un aumento de la cavidad (1.34) para aniones metilsulfato y tioalato.

Como simple regla empírica, se puede decir que el factor de escalado debe aumentarse cuando la carga positiva generada por protonación puede ser repartida en un número mayor de átomos de hidrógeno (cationes) o a medida que la carga negativa puede deslocalizarse en un número mayor de átomos (aniones).

Tabla 4.1-8. Energías de solvatación calculadas y experimentales (kcal/mol) de aniones monovalentes obtenidas adoptando un factor de escalado dependiente del entorno químico para definir la frontera en la cavidad electrostática entre soluto y solvente

	Calculada	Experimental
Conjunto A1 ($\kappa=1.09$)		
Metanol	-94	-95
Etanol	-90	-91
1,2-Etandiol	-91	-85
2,2,2-Trifluoroetanol	-64	-78
Metoxyetanol	-90	-89
Propanol	-90	-88
2-Propanol	-89	-86
2-Propen-1-ol	-87	-86
t-Butanol	-87	-82
Hexafluoropropan-2-ol	-53	-65
2-Butanol	-86	-84
Benzilalcohol	-83	-85
Fenol	-72	-72
m-Hidroxifenol	-74	-74
m-Metilfenol	-73	-71
m-Nitrofenol	-60	-62
o-Chlorofenol	-68	-66
o-Metilfenol	-70	-70
o-Nitrofenol	-62	-60
p-Chlorofenol	-65	-66
p-Hidroxifenol	-75	-78
p-Metilfenol	-72	-72
p-Nitrofenol	-59	-58
Difenilamina	-66	-54
Conjunto A2 ($\kappa=1.18$)		
Anilina	-75	-63
p-Nitroanilina	-58	-57

Acetamida	-81	-80
Propanenitrilo	-68	-66
3-Pentanona	-69	-74
Acetaldehído	-74	-76
Dimetilsulfóxido	-66	-68
Acetona	-74	-76
Nitrometana	-70	-76
Formaldehído	-74	-76
Acetonitrilo	-71	-72
Etiletanoato	-70	-66
Conjunto A3 ($\kappa=1.24$)		
Ácido Fórmico	-77	-76
Ácido Acético	-75	-78
Ácido Acrílico	-75	-74
Ácido Benzoico	-75	-71
Ácido Cloroacético	-69	-70
Ácido Dicloroacético	-64	-62
Ácido Hexanoico	-76	-74
Ácido Pirúvico	-69	-68
Ácido Propiónico	-74	-76
Ácido Trifluoroacético	-55	-59
Conjunto A4 ($\kappa=1.34$)		
Ácido Metanesulfónico	-66	-72
Etanotiol	-73	-72
Propanotiol	-73	-70
Tiofenol	-66	-63
Metanotiol	-73	-74
Msd	0.0	
Rmsd	4.4	
C	1.00	
R	1.00	

F	14296	
---	-------	--

El uso de un escalado dependiente del entorno químico para grupos ionizables resulta una alternativa prometedora para el cálculo de energías libres de solvatación a nivel semiempírico. Sin embargo, la implementación de esta estrategia computacional puede ser cuestionada por la incertidumbre intrínseca asociada a los valores experimentales de las energías de solvatación (alrededor de 3 kcal/mol⁷) y por el limitado número de compuestos cargados que den cuenta de toda la variedad de grupos funcionales. En este contexto puede resultar una alternativa la calibración de factores de escalado dependientes del entorno químico a partir de cálculos de energía de solvatación más sofisticadas como *ab initio*.

4.2. Estrategia multinivel

El objetivo central de esta parte del trabajo es el desarrollo de una metodología eficaz para la exploración conformacional de compuestos de potencial desarrollo como fármacos. En particular, nuestro objetivo se orienta a la determinación del coste asociado a la selección de la conformación bioactiva de un fármaco, es decir, aquella conformación que debe adoptar para poder ser reconocido e interactuar en el centro activo de su receptor macromolecular.

En base a dicho objetivo, el trabajo que se presenta a continuación está concebido como una técnica competitiva de cara a la exploración del espacio conformacional de moléculas bioactivas en solución. En un principio, dicho objetivo representa un esfuerzo computacional ingente, por cuanto conlleva la exploración del espacio conformacional accesible para la molécula objeto de interés, así como del conjunto de configuraciones que puede adoptar las moléculas de solvente. No obstante, la metodología que se presenta a continuación, si bien supone un coste computacional significativo, está diseñada con la finalidad de poder convertirse en una herramienta aplicable a fármacos o compuestos análogos. Por ello, el esfuerzo computacional se concentra íntegramente en la exploración del espacio conformacional de la molécula de interés, mientras que el solvente se trata con el modelo continuo MST parametrizado para el hamiltoniano RM1 presentado en la sección anterior.

La metodología adoptada puede describirse como un método de exploración multinivel, que combina diversas estrategias computacionales con objeto de poder realizar la exploración conformacional de moléculas flexibles en solución. Dada la naturaleza multinivel, éste es el nombre que hemos elegido para la metodología. En nuestra opinión, los resultados que se discutirán a continuación y la base teórica de la metodología apoyan la viabilidad de esta estrategia.

4.2.1 Descripción de la metodología *Multinivel*

Background teórico

Si se desean estudiar los estados estables de un sistema dado la termodinámica nos dice que la magnitud de interés es la energía libre. En general, la diferencia de energía libre de los estados que un sistema puede visitar da una medida de la población relativa de dicho estados. En el caso más simple de un sistema que puede existir en sólo dos estados A y B, la diferencia de energía libre entre éstos (ΔG_{AB}) nos dirá cuál es más estable y en qué medida.

Esto puede aplicarse al estudio de las preferencias conformacionales de una molécula. En el contexto de la predominant state approximation¹⁶ interesa obtener la energía libre asociada a cada conformación. Esta aproximación se basa en la suposición intuitiva de que la contribución más importante a la energía libre total se halla en los mínimos de energía y sus alrededores, puesto que el factor de Boltzman es allí mayor. Si estamos interesados en la energía libre total del sistema la integral configuracional se aproxima por la suma de las energías libres asociadas a cada mínimo o pozo en la superficie de energía potencial (SEP). Cuanto más suave es esta SEP o cuantos más grados de libertad tenga la molécula, mayor número de pozos habrá. En esta tesis no estamos interesados en la energía libre total sino en las diferencias de energía libre de los pozos de la SEP (es decir, conformaciones estables) en el marco de la predominant state approximation.

La ventaja de esta aproximación es que el esfuerzo computacional puede concentrarse sólo en aquellas áreas del espacio conformacional donde existen pozos en la SEP, evitando muestro inútil. En este trabajo, el muestreo del espacio conformacional de una molécula se lleva adelante variando sólo las torsiones activas (enlaces con libre rotación). Los grados de libertad llamados duros (estiramiento-estrechamiento de enlace y ángulo) no contribuyen apreciablemente a la flexibilidad molecular por lo que se asume que pueden no tenerse en cuenta y las energías libres relativas obtenerse sólo integrando los grados de libertad blandos, es decir, las torsiones activas. Dicho de otro modo, las estimaciones de diferencias en entropía entre pozos de la SEP no se verán afectadas apreciablemente si los grados de libertad duros no se consideran^{16,17}. En este trabajo las variaciones en ángulo se incluyen de un modo promedio dado que se optimizan en cada paso de la simulación. Los enlaces, por el contrario, se dejan fijos a lo largo de la simulación para evitar transferencias protónicas que cambiarían la naturaleza de la molécula en estudio.

En la predominant state approximation la energía libre asociada a un mínimo de la

SEP puede ser escrito como¹⁶:

4.2-1

$$G_j = -RT \ln \left[\frac{v_j}{N_j} \sum_{i=1}^{N_j} e^{-E_i/RT} \right]$$

donde v_j es el volumen del espacio conformacional muestreado para el pozo j , N_j es el número de estados conformacionales para este pozo y es E_i la energía de cada estado perteneciente al pozo j .

Metodología

Como se mencionó anteriormente la predominant state approximation es útil si el muestreo se concentra en las zonas de la SEP que corresponden a conformaciones estables. En este trabajo esto se logra mediante una exploración por Metropolis Monte Carlo¹⁸ variando las torsiones activas de la molécula bajo estudio.

La energía de cada conformación se calcula con el hamiltoniano RM1 recientemente parametrizado para el modelo de solvatación continuo MST (RM1/MST en adelante) que incluye el efecto del solvente en un modo promediado ($H_j + G_j^{RM1/MST}$). La energía libre asociada a cada pozo de la SEP se estima a nivel semiempírico RM1 por medio de la ecuación 4.2-1.

Pozos que tienen contribuciones no despreciables al espacio conformacional total explorado por una molécula difieren generalmente en menos de 2 kcal/mol. Esto significa que es necesario alcanzar una exactitud “<1 kcal/mol” cuando se desean estudiar las preferencias conformacionales de una molécula. En este sentido RM1 no es lo suficientemente preciso para describir las diferencias de energía libre entre pozos, lo que queda de manifiesto en los cálculos realizados para la presente memoria. Por esta razón hemos ido un paso más allá en la precisión de la estimación de las energías libres relativas de las conformaciones de una molécula lo que da como resultado la estrategia Multinivel que aquí se presenta. La idea central es la de recalcular las energías de las estructuras de mínima energía de cada pozo de la SEP con un nivel de teoría más robusto. Para ello, a partir de la exploración MC-RM1/MST se construye la SEP como función de los grados de libertad explorados y se calcula la energía libre de cada pozo

como se describe a continuación.

Partiendo de la ecuación 4.2-1 obtenemos la ecuación 4.2-2 siguiendo la deducción siguiente:

Si sumamos y restamos la energía del mínimo del pozo j (E_j^{\min}) al exponente de la exponencial llegamos a:

$$G_j = -RT \ln \left[\frac{v_j}{N_j} \sum_{i=1}^{N_j} e^{-(E_i - E_j^{\min} + E_j^{\min})/RT} \right]$$

$$G_j = -RT \ln \left[\frac{v_j}{N_j} \right] + E_j^{\min} - RT \ln \left[\frac{v_j}{N_j} \sum_{i=1}^{N_j} e^{-(E_i - E_j^{\min})/RT} \right]$$

Dado que el número de estructuras N_j es proporcional al volumen del pozo v_j , $\frac{v_j}{N_j}$ es una constante que depende de cuan fina sea la grilla que describe la SEP. Como interesan las diferencias de energía libre el primer término se cancela. Por esta razón en lo que sigue se omite. En este punto es donde entra la estrategia Multinivel. La energía efectiva E_j^{\min} (que a nivel semiempírico es $H_j^{\min} + G_j^{\min, RM1/MST}$) se recalcula a un nivel de teoría más robusto, como $E_j^{\min, high} + G_j^{\min, high/MST} + ZPE$. Los términos de la suma son: la energía en vacío a alto nivel de teoría, la energía libre de solvatación a alto nivel y la energía de punto cero.

$$G_j^{MULTI} = E_j^{\min, high} + G_j^{\min, high/MST} + ZPE + -RT \ln \left[\frac{v_j}{N_j} \sum_{i=1}^{N_j} e^{-(E_i - E_j^{\min})/RT} \right]$$

4.2-2

$$G_j^{MULTI} = E_j^{\min, high} + G_j^{\min, high/MST} + ZPE + G_j^{\min, local, RM1}$$

La estrategia Multinivel hace uso de una exploración conformacional “poco costosa” a nivel RM1 para cada pozo j obteniendo $G_j^{\min, local, RM1}$ y la costosa y precisa estimación a

alto nivel de teoría de la energía de la estructura que corresponde al mínimo del pozo ($E_j^{\text{min,high}} + G_j^{\text{min,high/MST}} + \text{ZPE}$). De este modo se logra un balance costo/calidad adecuado para una correcta estimación de los términos entrópicos y entálpicos asociados a las distintas conformaciones que adopta una molécula tipo fármaco.

La estimación de la energía libre asociada a cada pozo en el marco de la predominant state approximation y la estrategia Multinivel asume que la forma del pozo está bien descrita a nivel RM1 mientras que la diferencia entre pozos debe ser corregida. Es decir, la distribución de energías dentro del pozo está bien descrita a nivel RM1 pero la posición de los pozos en la SEP se corrige a un nivel más alto. Esto se presenta esquemáticamente en la figura Figura 4.2-1.

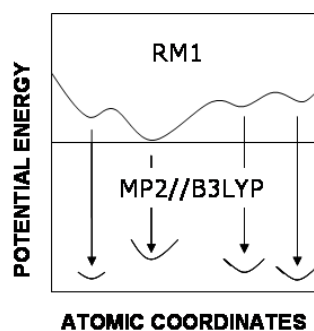


Figura 4.2-1. Representación esquemática de la estrategia Multinivel

En la presente implementación de la metodología Multinivel los cálculos a alto nivel de teoría se llevaron adelante a nivel MP2/aug-cc-pVDZ//B3LYP/6-31G*-MST corrigiendo la energía de punto cero (ZPE) a nivel B3LYP/6-31G*-MST. La energía de solvatación de la estructura de mínima energía del pozo se calcula al mismo nivel de la optimización: B3LYP/6-31G*-MST level.

Los niveles de teoría escogidos se basan en trabajos previos y estudios realizados específicamente para el presente trabajo. El nivel de teoría usado para la optimización de geometría (B3LYP/6-31G*-MST) es adecuado para comparar estructuras similares pero se producen errores no despreciables si hay un cambio importante en los contactos intramoleculares¹⁹. Es por ello que se debe recurrir a niveles de teoría aún más robustos para describir correctamente las preferencias conformacionales de una molécula tipo fármaco. Una comparación exhaustiva acerca del rendimiento de una gran variedad de funcionales DFT y hamiltoniano MP2 combinados con diferentes bases fue llevado

adelante por Riley et al.²⁰. De acuerdo a sus resultados MP2/aug-cc-pVDZ da los errores más bajos en varias propiedades entre las cuales se encuentra “conformational energy”, es decir, la diferencia de energía entre conformeros de una misma molécula. Basado en esto y nuestros propios análisis en términos de costo/calidad se eligió MP2/aug-cc-pVDZ para el cálculo puntual de alto nivel. Un ejemplo a modo ilustrativo se muestra en la Tabla 4.2-1

Tabla 4.2-1. Tiempo de CPU para cálculo puntual y error medio en la energía conformacional a distintos niveles de teoría

	B3LYP/6-31+g*	MP2/6-31+g*	MP2/aug-cc-pVDZ	MP2/aug-cc-pVTZ
Tiempo de CPU (min) ^a	2 (0.3)	2 (0.4)	36 (1)	570 (33)
Error en la energía conformacional (%) ^b	22.1	16.8	6.8	8.9

^a Tiempo para un cálculo puntual para histamine neutral y 1,2-dicloroetano (ente aréntesis)

^bRiley et al., JTCT, 07²⁰

Una vez se ha obtenido la energía libre asociada a cada pozo, la población relativa de cada conformación puede ser estimada según:

4.2-3

$$\text{pop}_j = \frac{e^{-G_j/RT}}{\sum_i e^{-G_i/RT}}$$

Protocolo

En esta sección se describirá el protocolo completo utilizado para obtener los diferentes componentes de G_j^{MULTI} .

El primer paso es la exploración conformacional por medio de la técnica de Metropolis Monte Carlo utilizando para la energía la reciente implementación del hamiltoniano RM1 y el modelo de solvatación continuo MST: MC-RM1/MST. Esto fue implementado en una versión modificada del MOPAC 6.0. La exploración conformacional se realiza variando las torsiones activas de la molécula en estudio. Los grados de libertad duros (enlace, ángulo) no se exploran. Lo mismo sucede con los grados de libertad del solvente que están integrados en el modelo de solvatación continuo. El resultado de la exploración conformacional se proyecta sobre una grilla n-dimensional (siendo n el número de torsiones activas exploradas) construyendo así una SEP (n+1)-dimensional. Luego se identifican j pozos en la SEP obtenida y se calcula $G_j^{\text{local,RM1}}$ para cada uno. Las j estructuras que corresponden a los mínimos de energía de cada uno de los j pozos se optimizan a nivel B3LYP/6-31G*-MST y se obtiene $G_j^{\text{min,high/MST}}$. Sobre las j estructuras optimizadas se realiza un cálculo puntual a nivel MP2/aug-cc-pVDZ y un cálculo de frecuencias al mismo nivel de la optimización para obtener $E_j^{\text{min,high}}$ y ZPE, respectivamente. El protocolo completo se esquematiza en la Figura 4.2-2.

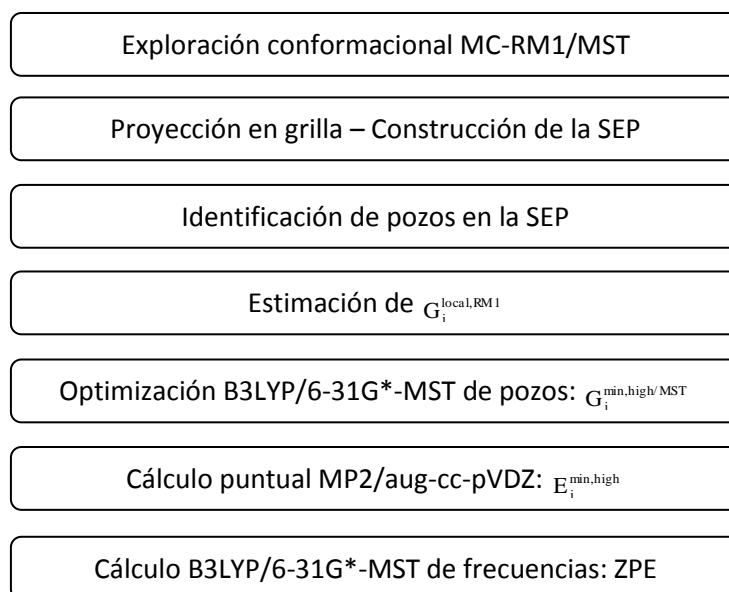


Figura 4.2-2. Protocolo *Multinivel*

Una vez que la exploración conformacional se ha completado, la identificación de los pozos en la SEP se logra mediante el algoritmo que se describe a continuación y se esquematiza en la Figura 4.2-3.

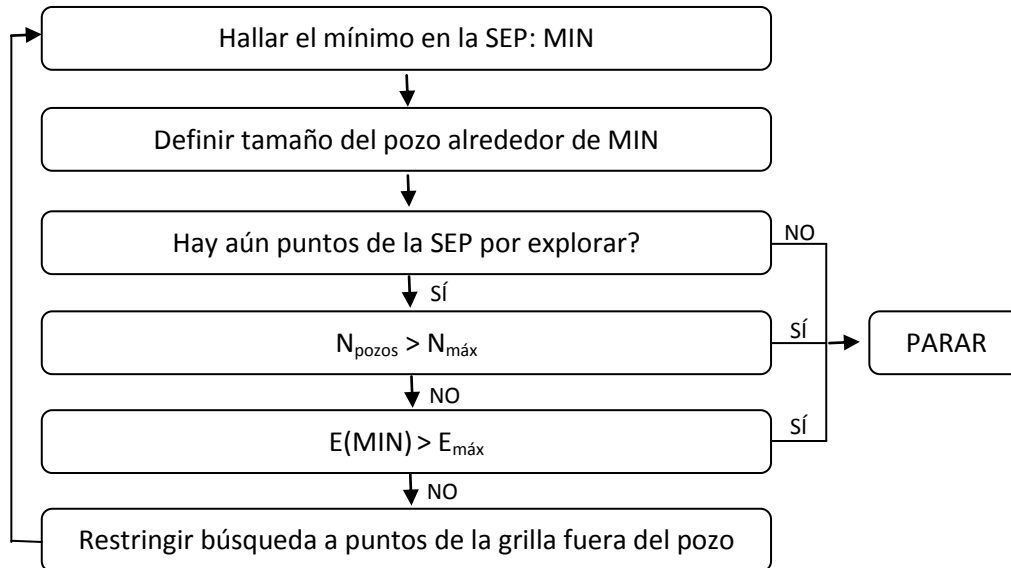


Figura 4.2-3. Representación esquemática de la identificación de pozos en la SEP

Primero se identifica el mínimo de energía en la SEP (MIN), representada mediante la grilla $(n+1)$ -dimensional, siendo n el número de torsiones activas exploradas. Una hipersuperficie cúbica n -dimensional se define alrededor de MIN que delimita la extensión del pozo. El algoritmo permite que esta hipersuperficie crezca en las n -direcciones hasta que se entra en otro pozo. Es decir, la energía deja de subir y comienza a descender. Una vez este nuevo pozo ha sido identificado el algoritmo realiza 3 chequeos antes de buscar un nuevo pozo:

1. ¿Existen aún puntos de la grilla que no pertenecen a ningún pozo encontrado hasta ahora?
2. ¿El número máximo de pozos a encontrar $N_{máx}$ ha sido alcanzado?
3. ¿La energía del último pozo es mayor que $E_{máx}$, la energía máxima definida por el usuario?

Si la condición 1 es falsa el algoritmo para porque significa que los pozos

encontrados ya cubren por completo el espacio conformacional. Si, por el contrario, es falsa, se chequea que las condiciones 2 y 3 se cumplan. Si alguna de éstas es verdadera el algoritmo para. $N_{m\acute{a}x}$ es el máximo número de pozos a hallar y $E_{m\acute{a}x}$ es la máxima energía que un pozo puede tener respecto al mínimo global. $N_{m\acute{a}x}$ y $E_{m\acute{a}x}$ son valores definidos por el usuario. Si ninguna de estas condiciones se cumple el proceso de búsqueda de mínimo comienza nuevamente, buscando un nuevo mínimo global en la grilla dentro de aquellos puntos de la misma que no pertenecen a ningún pozo hallado hasta este punto.

4.2.2 Exploración conformacional de moléculas flexibles: histamina.

La viabilidad de la metodología Multinivel se ha calibrado estudiando el espacio conformacional de la histamina neutra (Figura 4.2-4).

Dicha molécula ha sido escogida por diversas razones. Por un lado, es una biomolécula relevante biológica y farmacológicamente, pues interviene en una amplia variedad de procesos fisiológicos y patológicos. La dilucidación de las funciones biológicas de la histamina ha supuesto un objetivo constante dentro de la investigación farmacológica dado su rol en diversas manifestaciones clínicas, como la secreción de ácido gástrico o bien reacciones alérgicas e inflamatorias y, en general, su papel como neurotransmisor.

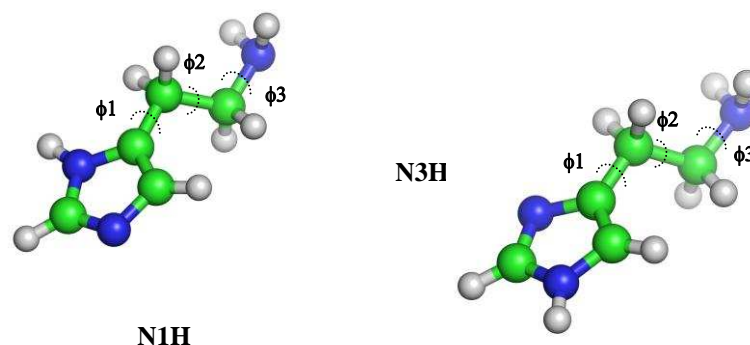


Figura 4.2-4. Torsiones activas para los tautómeros N1H y N3H de la histamina neutra

Por otro lado, desde el punto de vista estrictamente químico, es un compuesto altamente versátil, pues presenta un interesante y complejo balance poblacional en función de las distintas formas iónicas, tautoméricas y conformacionales, que la convierten en un prototipo ideal para poder explorar el espacio conformacional con la metodología *Multinivel* propuesta en este trabajo.

La histamina es una molécula hidrofílica, muy polar, que posee una gran capacidad para establecer enlaces puente de hidrógeno. Existen dos centros básicos: el nitrógeno tipo piridina del anillo heterocíclico y el nitrógeno tipo amina de la cadena lateral. Sus valores de pK_a son 5.8 y 9.4, respectivamente. Por tanto, puede existir en diversas formas iónicas (neutra, catión y dicatión), cuya población dependerá del pH del medio. En particular, a pH fisiológico la histamina existirá principalmente en forma monocatiónica, debido a la protonación del grupo amino terminal.

Además de sus propiedades iónicas, la histamina está sujeta a un equilibrio tautomérico prototrópico entre los átomos de nitrógeno del anillo imidazólico tanto en la forma neutra como en la monocatiónica. Según el átomo de nitrógeno del anillo que una al hidrógeno, denominaremos a los tautómeros N1H y N3H.

Finalmente, debe considerar el hecho de que posee una elevada flexibilidad conformacional debida a los enlaces simples C-C que definen 3 torsiones activas (ϕ_1 , ϕ_2 y ϕ_3 en la Figura 4.2-4). Según el valor de ϕ_2 denominaremos a los conformeros trans ($\phi_2 = 180$) o gauche ($\phi_2 = 60$ o 300).

Las características estructurales expuestas de la histamina ponen de manifiesto la complejidad que supone su estudio, ya que en solución existirá como una mezcla de muchas especies.

Estudios previos sobre las preferencias conformacionales de la histamina

Las preferencias conformacionales, tautoméricas y iónicas de la histamina han sido fruto de numerosos estudios experimentales y teóricos, pues son claves a la hora de entender su rol fisiológico. Dado el delicado equilibrio que existe entre las diversas formas de la histamina, pequeños errores teóricos o experimentales pueden conducir a conclusiones distintas, como se observa en los distintos estudios que se indican a continuación para las propiedades tautoméricas y conformacionales.

En base a estudios basados en determinaciones de pK_a de metil y bencil derivados

de histamina sustituidos en N1 y N3²¹, Ganellin determinó que aproximadamente el 77% de la forma monocatiónica existe como tautómero N3-H en solución acuosa a 25° C.

Estudios de resonancia magnética nuclear (RMN) de ¹³C indican un porcentaje de monocatión como tautómero N3-H comprendido entre 80% y 90 %^{22,23}, aunque para la forma neutra también se ha indicado un predominio de dicho tautómero.

La determinación de la estructura cristalina de la histamina neutra²⁴ y de la sal de bromuro monohidratada del monocatión²⁵ revelan que el primero cristaliza como tautómero N1-H en tanto que el segundo lo hace como N3-H. Ambas lo hacen en conformación trans.

Cálculos mecanocuánticos²⁶⁻²⁸ indican que la preferencia tautomérica de la especie N3H con respecto al tautómero N1H en la histamina monocatiónica se reduce drásticamente como consecuencia de la desprotonación del monocatión. La consideración del efecto del solvente no introduce cambios significativos con respecto a los resultados en fase gas²⁹.

Estudios de RMN^{30,31} y espectroscopía de infrarrojo³² indican que en solución tanto el dicatión como el monocatión y la forma neutra se hallan como confórmeros trans ($\phi_2 = 180$) y gauche ($\phi_2 = 60$ o 300), existiendo para dichas especies iónicas aproximadamente la misma proporción de ambos confórmeros.

El valor de ϕ_1 depende según la especie iónica considerada. Datos cristalográficos³³ dan un amplio rango de valores para dicho ángulo cuando se considera la especie dicatiónica (desde 4 a 83 grados), siendo mucho más restringido para el monocatión²⁵ y la forma neutra²⁴, pues los valores hallados son 90 y 66 grados, respectivamente.

Mediante experimentos de chorro de expansión libre, Vogelsanger et al.³⁴ estudiaron el espectro rotacional de la histamina; en combinación con cálculos teóricos identificaron 20 posibles tautómeros/confórmeros. De ellos, cuatro fueron identificados como los más estables, todos ellos gauche, uno como tautómero N1-H y los otros tres como N3H.

En base a cálculos ab initio y simulaciones Monte Carlo, Nagy et al.³⁵ estudiaron las preferencias tautoméricas y conformacionales de la histamina. En fase gas las formas más estables para la forma neutra resultaron ser gauche, estabilizadas por puentes de hidrógeno intramoleculares, siendo el tautómero N1H más estable que el N3H. Para el monocatión la forma gauche del tautómero N3H resultó ser la más estable. Los

resultados en solución acuosa para la forma neutra arrojan una relación trans/gauche de 87/13. Sin embargo a pH fisiológico, donde el monocatión es la forma mayoritaria, dicha relación cambia a 36/64.

Hernández-Laguna et al.³⁶ estudiaron la superficie de energía potencial del monocatión por métodos RHF y MP2. La forma gauche del tautómero N3H resultó ser la más estable, estando desfavorecida la forma trans. Karpinska et al.³⁷, en base a estudios teóricos a nivel HF y MP2, hallaron para el monocatión al tautómero N3H como el más estable y para la forma neutra el N3H en la conformación gauche.

El estudio teórico más reciente ha sido el de Ramírez et al.³⁸, quienes realizaron un análisis estructural y vibracional por espectroscopía Raman e Infrarrojo combinados con cálculos DFT de la forma neutra de la histamina. Los resultados reflejan una estabilidad relativa en orden creciente según t1-t3-g1-g3, indicando con la letra el conformero (gauche/trans) y con el número el tautómero (N1/N3)

El conjunto de trabajos descritos pone de manifiesto que la dificultad de establecer de forma unívoca las preferencias conformacionales y tautoméricas de la histamina, representando su estudio un desafío para los métodos de exploración conformacional.

En este contexto se plantean los objetivos del presente trabajo, que utiliza a la histamina como molécula de prueba para la exploración conformacional, identificación de estructuras estables y población relativa de las mismas haciendo uso del método Multinivel.

Histamina neutra

Exploración MC-RM1-MST

Se llevaron a cabo simulaciones de MC-RM1/MST tanto en fase gas como en solución con la versión modificada de MOPAC 6.0 totalizando $4 \cdot 10^5$ pasos para ambos tautómeros de la forma neutra de la histamina (N1H y N3H). Se exploraron las 3 torsiones “activas” (Figura 4.2-4), relajando tanto ángulos como las restantes torsiones para cada estructura a lo largo de la exploración conformacional. Las distancias de enlace se mantuvieron fijas a lo largo de todas las simulaciones.

En la Figura 4.2-5 y Figura 4.2-6 se presentan los histogramas de distribución de

población de cada torsión activa en fase gas y solución para ambos tautómeros de la histamina neutra.

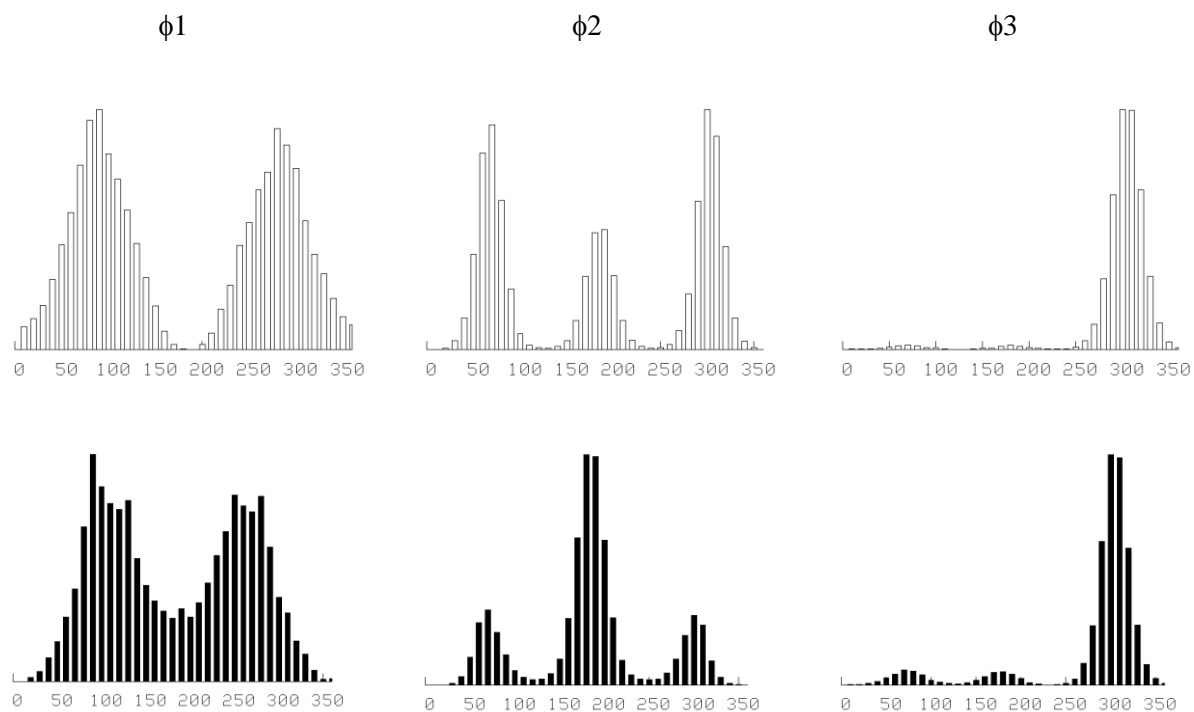


Figura 4.2-5. Histogramas para la población de las torsiones del tautómero N1H de la histamina neutra en fase gas (gris) y solución (negro)

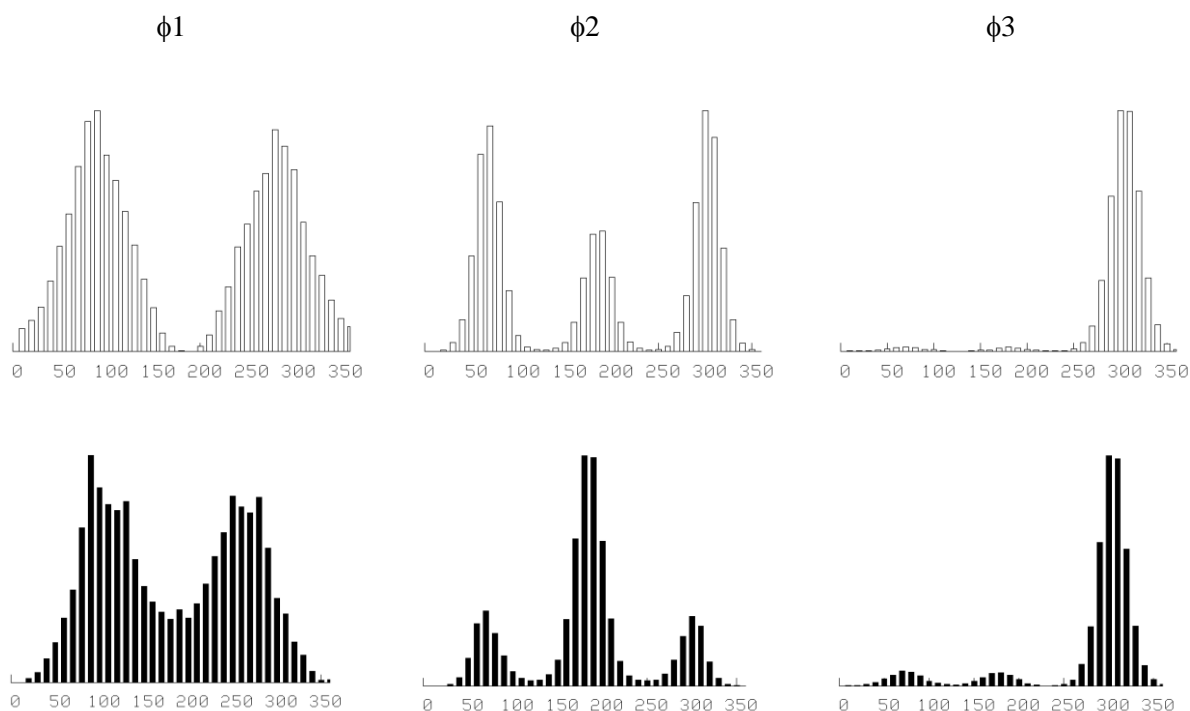


Figura 4.2-6. Histogramas para la población de las torsiones del tautómero

N3H de la histamina neutra en fase gas (gris) y solución (negro)

Como es de esperar la torsión ϕ_2 presenta 3 máximos en todos los casos, que corresponden al conformero gauche (alrededor de 60 y 300 grados) y trans (alrededor de 180 grados). Las formas gauche y trans del tautómero N1H las llamaremos g1 y t1, respectivamente, mientras que las del tautómero N3H serán g3 y t3.

En todos los casos existen poblaciones apreciables de ambos conformeros, aumentando la relación trans/gauche cuando se pasa de fase gas a solución. Este resultado puede explicarse en términos de una mejor energía de solvatación obtenida al separar los centros básicos de la molécula. En la Figura 4.2-7 se aprecian estos resultados en términos de la distancia del nitrógeno N1 al H del grupo amino. En fase gas las zonas más pobladas caen alrededor de distancias centradas en torno a 3.1 y 4.5 Å, correspondiendo a las distancias preferidas para las formas gauche (donde el N del amino apunta hacia N1) y trans, respectivamente. Al pasar de fase gas a solución, se observa una mayor población por encima de los 4 Å, poblándose especialmente en el rango de 4,5 a 5 Å, lo que demuestra la preferencia a separar el grupo amino del anillo,

rompiendo el puente de hidrógeno y logrando una mejor exposición al solvente.

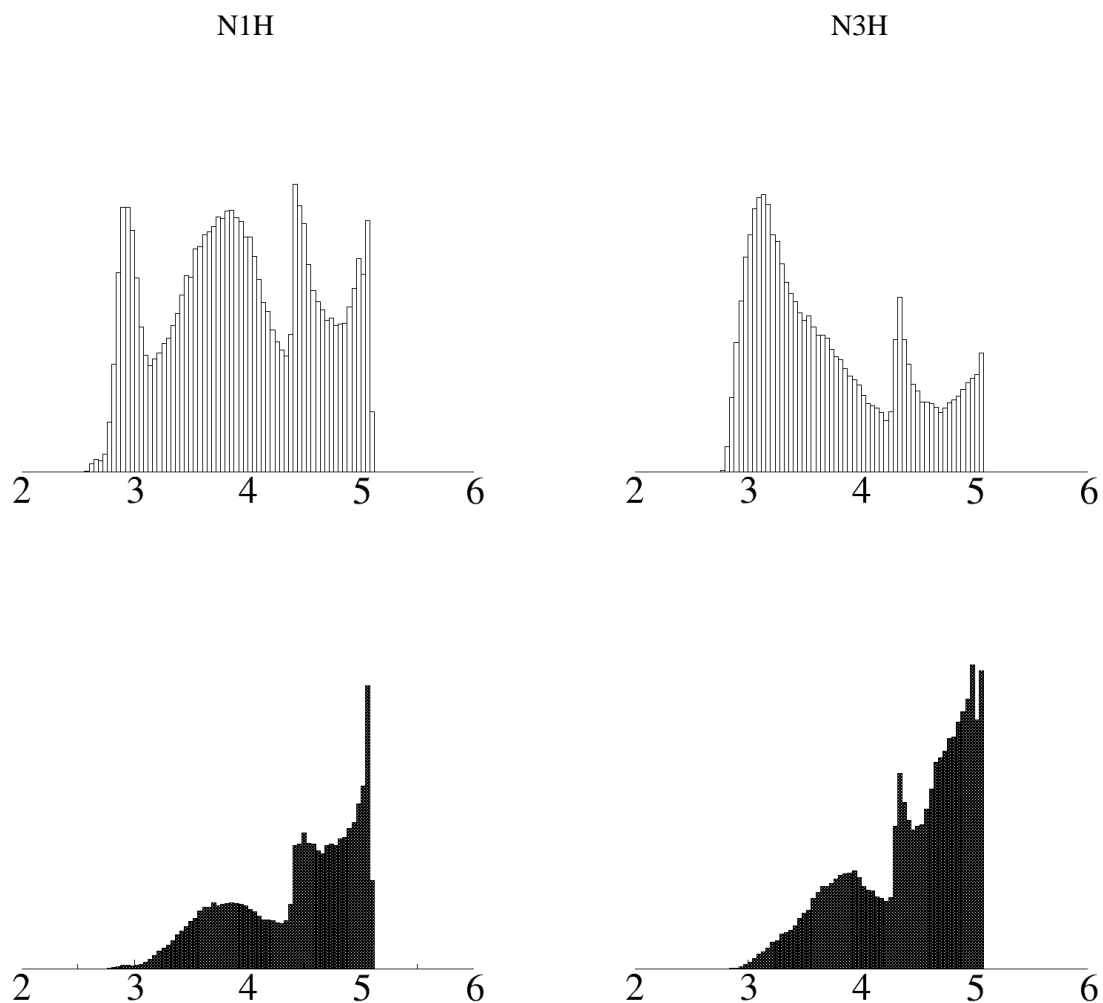


Figura 4.2-7. Histogramas para la distribución de la distancia del nitrógeno N1 al H del grupo amino en fase gas (gris) y solución (negro) para ambos tautómeros

Desde un punto de vista cuantitativo, los resultados obtenidos a nivel RM1-MST para las distintas poblaciones conformacionales son más cuestionables. Así, a partir de las energías efectivas RM1 ($H_{\text{gas}} + G_{\text{MST}}$), el tautómero N1H es 1 kcal/mol más estable que N3H en fase gas y 2 kcal/mol en solución. Sin embargo, muchos de los resultados obtenidos en estudios previos (véase Estudios previos sobre las preferencias conformacionales de la histamina al inicio de la sección 4.2.2) indican que el tautómero N3H es la especie predominante en solución. Esta discrepancia debe atribuirse a una limitación intrínseca del método RM1. Esto queda de manifiesto si comparamos las

estabilidades relativas de los tautómeros del 4(5)metilimidazol a nivel RM1 y distintos niveles de teoría más altos (Tabla 4.2-2)³⁹

Tabla 4.2-2. Entalpías relativas para la conversión tautomérica N(π)H a N(τ)H del 4(5)metilimidazol

Nivel de teoría	ΔH (kcal/mol)
RM1	0,5
HF/6-31G(d,p)	-0,3
HF/6-311+G(d,p)	-0,5
MP2/6-311+G(d,p)	-0,7
MP3/6-311+G(d,p)	-0,5
MP4/6-311+G(d,p)	-0,7

Cálculo multinivel

Siguiendo la estrategia Multinivel, pozos en la SEP definida en función de las 3 torsiones activas fueron identificados para ambos tautómeros en solución. Las Figura 4.2-8 y Figura 4.2-9 muestran los histogramas 3D (distribución de población alrededor de 2 torsiones activas) para ambos tautómeros. A su vez se muestran isocontornos proyectados sobre el plano XY que dan una idea clara de las diferentes posibles combinaciones de las 3 torsiones activas que definen los pozos. Por ejemplo, mientras el histograma 2D (Figura 4.2-6 y Figura 4.2-7) sólo muestra 3 máximos para la torsión ϕ_1 en solución, la Figura 4.2-8 y Figura 4.2-9 permiten ver diferencias cuando ϕ_2 está alrededor de 60, 180 o 300. Existe un desplazamiento en la posición de los picos para cada combinación estable de ϕ_1 y ϕ_2 .

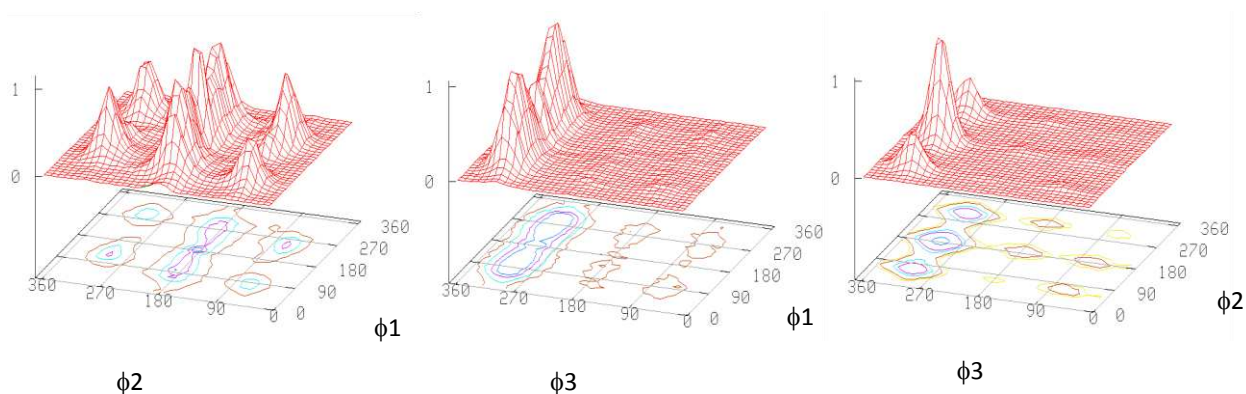


Figura 4.2-8. Histogramas 3D para la distribución de la población para el tautómero N1H de la histamina neutra en función de 2 de las 3 torsiones activas. Isocontornos proyectados sobre el plano XY.

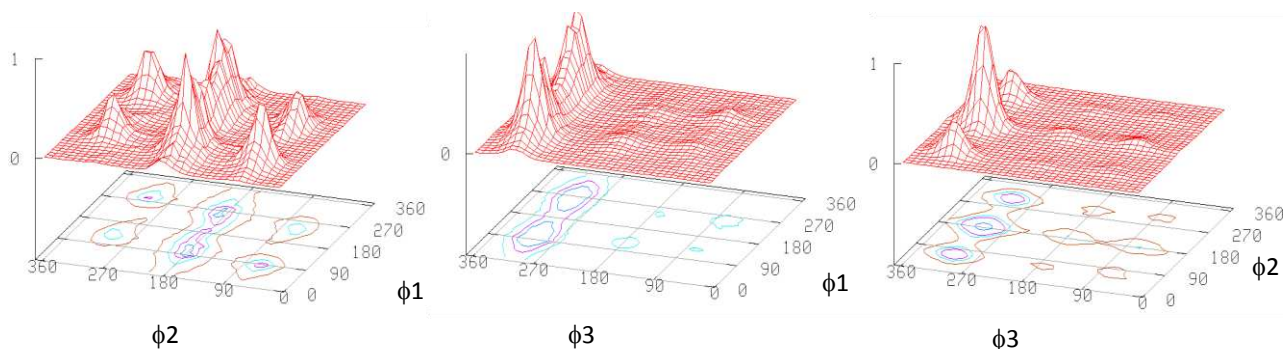


Figura 4.2-9. Histogramas 3D para la distribución de la población para el tautómero N1H de la histamina neutra en función de 2 de las 3 torsiones activas. Isocontornos proyectados sobre el plano XY.

Siguiendo la metodología Multinivel se optimizaron las estructuras de mínima energía de cada pozo y obtuvieron los distintos términos de la energía libre asociada a cada conformación (pozo). A partir de la ecuación 4.2-3 se obtuvieron las poblaciones relativas de cada conformero. Esto se presenta en la Tabla 4.2-3

Tabla 4.2-3. Población relativa de cada pozo en la SEP (confórmero) para la histamina neutra en solución según el método *Multinivel*

	N1H		N3H	
Trans	080.170.060	2%	080.170.290	5%
	080.170.180	2%	120.170.060	3%
	090.180.300	3%	120.180.170	7%
	170.170.060	1%	170.180.060	1%
	170.180.170	1%	180.170.290	1%
	170.180.300	1%	210.170.170	2%
	250.180.170	2%	240.170.060	7%
	260.170.060	2%	240.180.300	5%
	260.180.290	3%		
Gauche	030.290.170	5%	080.050.030	0%
	060.070.060	0%	080.050.300	4%
	070.070.170	2%	080.060.170	3%
	080.060.300	2%	100.270.160	0%
	110.290.170	0%	120.290.290	1%
	110.290.290	2%	120.300.070	1%
	140.290.070	1%	230.050.290	1%
	170.290.160	0%	230.060.160	1%
	180.060.070	0%	250.080.060	0%
240.060.300	2%	270.300.180	1%	

	240.070.170	2%	270.300.300	4%
	250.050.050	11%	280.290.050	4%
	270.300.300	3%		
	280.280.170	0%		
	290.290.070	2%		

La Tabla 4.2-3 muestra más pozos para el tautómero N1H puesto que existen menos combinaciones estables para las 3 torsiones activas para el tautómero N3H. Esto puede observarse comparando las gráficas izquierda y central de la Figura 4.2-8 y Figura 4.2-9. Allí se aprecian picos más definidos para ϕ_1 alrededor de 180 grados en N1H que en N3H, para el cual, en algunos casos, son inexistentes.

La mayor diferencia de energía entre pozos (sin tener en cuenta el menos estable) es de 3,1 kcal/mol y 1,9 kcal/mol si no se consideran los 5 conformeros menos estables. Estos resultados son una evidencia de la complejidad conformacional de esta a priori simple molécula dado que las preferencias conformacionales intrínsecas de cada tautómero no pueden ser descritas simplemente por un número reducido y arbitrario de conformeros. Por el contrario, el espacio conformacional completo debe ser explorado con el objeto de comprender las preferencias conformacionales.

Si se agrupan los conformeros de acuerdo al valor de ϕ_2 en t1, g1, t3 g3, se pueden derivar las preferencias conformacionales y tautoméricas de la histamina neutra en solución. A fines comparativos se presentan estos resultados en la Tabla 4.2-4 a distintos niveles de teoría para el cálculo puntual de la estructura de mínima energía de cada pozo. Se observa que el nivel de teoría utilizado es el que mejor rendimiento tiene en términos de costo/calidad, tal como se mencionó en la descripción de la metodología.

Tabla 4.2-4. Población relativa para los grupos t1, g1, t3 g3 para la histamina neutra en solución según el método *Multinivel*. Distintos niveles de teoría de alto nivel se presentan para el cálculo puntual de las geometrías de mínima energía

	B3LYP/6- 31+g*	MP2/6- 31+g*	MP2/aug-cc- pVDZ	MP2/aug-cc- pVTZ
g1	24%	54%	32%	31%
t1	24%	10%	16%	17%
g3	12%	16%	21%	21%
t3	40%	20%	31%	31%
trans/gauche	64/36	30/70	47/53	48/52
N1H/N3H	48/52	64/36	48/52	48/52

Los resultados obtenidos usando la estrategia Multinivel muestran que la solvatación, los puentes de hidrógeno y el equilibrio prototrópico no dirigen exclusivamente las preferencias conformacionales hacia ninguna de las posibles especies. Por el contrario, se observa un delicado balance entre estos factores que resulta en la ausencia de una especie dominante para la histamina neutra en solución. Esto es cierto tanto para el equilibrio trans/gauche como para el N1H/N3H en donde las poblaciones son comparables. En este sentido, es importante resaltar que los diferentes estudios previos no concuerdan en cuanto al orden de estabilidad de los distintos grupos (t1, g1, t, g3). Bonnet et al.²⁴ dan a la superfamilia t1 como la más estable mientras que Richards et al.²⁶, Weinstein et al.²⁷ y Topiol et al.²⁸ obtienen que el tautómero N3H es el preferido en solución para la histamina neutra. Byrn et al.³² realizaron un estudio conformacional exhaustivo en fase gas donde se concluyó que 4 conformeros gauche (1 g1 y 3 g3) eran los más estables. Nagy et al.³⁵ llevaron a cabo un estudio teórico y encontraron que la mezcla en equilibrio está compuesta de 83% t3, 12% g3, 4% t1 y 1% g1 mientras que el enfoque teórico-experimental de Ramirez et al.³⁸ dio una estabilidad descendiente según $g3 > g1 > t3 > t1$. Karpinska et al.³⁷, por otro lado, concluyeron que N1H y N3H son igualmente estables.

Estos resultados refuerzan la idea de que existe un balance sutil entre efectos que contribuyen a las preferencias tautoméricas y conformacionales. De hecho, la diferencia

de energía libre más alta entre los grupos obtenidos por la metodología Multinivel es sólo de 0,4 kcal/mol entre t1 y g1. Esto significa que pequeños errores en la estimación de las diferentes contribuciones a la energía libre pueden dar lugar a conclusiones completamente diferentes. La mayor diferencia de energía libre entre diferentes bases a nivel MP2 es de sólo 0,33 kcal/mol para el grupo g1 con 6-31+G* and aug-cc-VTZ. Es generalmente aceptado que el grupo t1 es el menos estable lo cual concuerda con nuestros resultados. Por otro lado, los resultados con DFT tienden a sobreestimar las formas trans posiblemente debido a la reportada tendencia a la subestimación de las interacciones débiles. Como se explicó, DFT es sólo utilizado para la optimización de geometría.

4.3. *Discusión*

Consideramos necesario resaltar que la mayor parte de los estudios que existen en la literatura derivan las preferencias conformacionales de la histamina (lo que es válido también en general para estudios conformacionales) de acuerdo al análisis de un número limitado de todos las posibles de conformeros que existen por combinación de distintos valores de las 3 torsiones activas. En el presente trabajo, por el contrario, realizamos una exploración conformacional completa para identificar todas las estructuras estables sin ninguna restricción a priori para el valor de las torsiones. Una vez identificamos los conformeros estables, agrupamos a posteriori las preferencias conformacionales de la histamina de un modo conveniente (t1, g1, t3, g3). Creemos que éste es un modo consistente y completo para estudiar las preferencias conformacionales de una molécula tipo fármaco dado que no se asume a priori que ninguna conformación sea la más estable de un grupo. En el particular caso de la histamina esto implica explorar todas las posibles combinaciones estables de ϕ_1 y ϕ_3 compatibles con un cierto valor del grupo definido por ϕ_2 .

Los resultados obtenidos nos alientan a continuar probando la estrategia multinivel con otras moléculas, cuyas preferencias conformacionales hayan sido previamente estudiadas y utilizarlo para la exploración conformacional de potenciales fármacos o moléculas tipo fármaco. En esta línea se está trabajando en 4 compuestos antirretrovirales del VIH estudiados por el grupo de Jorgensen⁴⁰

4.4. Bibliografía

- (1) Colominas, C.; Luque, F. J.; Orozco, M. J. *Phys. Chem. A* **1999**, 103, 6200.
- (2) Curutchet, C.; Salichs, A.; Barril, X.; Orozco, M.; Javier Luque, F. J. *Comput. Chem.* **2003**, 24, 32.
- (3) Pascual-Ahuir, J. L.; Andres, J.; Silla, E. *Chemical Physics Letters* **1990**, 169, 297.
- (4) Forti, F.; Barril, X.; Luque, F. J.; Orozco, M. J. *Comput. Chem.* **2008**, 29, 578.
- (5) Orozco, M., Luque F. J., Habibollah-zadeh, D., Gao, J. *J Chem Phys* **1995**, 102, 6145.
- (6) Luque, F. J.; Zhang, Y.; Aleman, C.; Bachs, M.; Gao, J.; Orozco, M. J. *Phys. Chem.* **1996**, 100, 4269.
- (7) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, 1, 1133.
- (8) Zhan, C.-G.; Dixon, D. A. *J. Phys. Chem. A* **2001**, 105, 11534.
- (9) Ozment, J. L.; Schmiedekamp, A. M. *Int J Quantum Chem* **1992**, 43, 783.
- (10) Tissandier, M. D.; Cowen, K. A.; Feng, W. Y.; Gundlach, E.; Cohen, M. H.; Earhart, A. D.; Coe, J. V.; Tuttle, T. R. *J. Phys. Chem. A* **1998**, 102, 7787.
- (11) Orozco, M.; Luque, F. J. *Chemical Physics* **1994**, 182, 237.
- (12) Curutchet, C.; Bidon-Chanal, A.; Soteras, I.; Orozco, M.; Luque, F. J. *J. Phys. Chem. B* **2005**, 109, 3565.
- (13) Dewar, M. J. S.; Dieter, K. M. *J Am Chem Soc* **1986**, 108, 8075.
- (14) Voets, R.; François, J. P.; Martin, J. M. L.; Mullens, J.; Yperman, J.; Van Poucke, L. C. *J. Comput. Chem.* **1989**, 10, 449.
- (15) Szafran, M.; Koput, J. *J. Comput. Chem.* **1991**, 12, 675.
- (16) Head, M. S.; Given, J. A.; Gilson, M. K. *The Journal of Physical Chemistry A* **1997**, 101, 1609.
- (17) Chang, C.-e. A.; Chen, W.; Gilson, M. K. *P Natl Acad Sci USA* **2007**, 104, 1534.
- (18) Nicholas, M., Arianna, W. R., Marshall, N. R., Augusta H. T., Edward T. *J. Chem Phys* **1953**, 21, 1087.
- (19) Butler, K. T.; Luque, F. J.; Barril, X. *J. Comput. Chem.* **2009**, 30, 601.
- (20) Riley, K. E.; Op't Holt, B. T.; Merz, K. M. *Journal of Chemical Theory and Computation* **2007**, 3, 407.
- (21) Ganellin, C. R. *J. Pharm. Pharmacol.* **1973**, 25, 787.
- (22) Reynolds, W. F., Tzeng, C.W. *Can J Biochem* **1977**, 55, 576.
- (23) Wasylishen, R. E., Tomlinson, G. *Can J Biochem* **1977**, 55, 579.
- (24) Bonnet, J. J., Ibers, J. A. *J Am Chem Soc* **1973**, 95, 4829.
- (25) Prouth, K., Critchley, S. R., Ganellin, C. R. *Acta Cryst* **1979**, 30, 2884.
- (26) Richards, W. G., Wallis, J., Ganellin, C. R. *Eur. J. Med Chem* **1979**, 14, 9.
- (27) Weinstein, H., Chou, D. , Johnson, C. L. , Kang, S., Green, J. P. *Mol. Pharmacol.* **1976**, 12, 738.
- (28) Topiol, S., Weintein, H., Osman, R. *J. Med. Chem.* **1984**, 27, 1531.

-
- (29) Centelles, E. Tesis de Licenciatura **1987**.
- (30) Ham, N. S., Casy, A. F., Ison, R. R. *J. Med. Chem.* **1973**, 16, 470.
- (31) Ganellin, C. R., Pepper, E. S., Port, G. N. J., Richards, W. G. *J. Med. Chem.* **1973**, 16, 610.
- (32) Byrn, S. R., Graber, C. W., Midiland, S. L. *J. Org. Chem.* **1976**, 41, 2283.
- (33) Ganellin, C. R., Parsons, M. E. *Pharmacology of Histamine Receptors*; Wright PSG: Bristol, 1982.
- (34) Vogelsanger, B., Godfrey, P. D., Brown R. D. . *J Am Chem Soc* **1991**, 113, 7864.
- (35) Nagy, P. I., Durant, G. J., Hoss, W. P., Smith, D. A. *J Am Chem Soc* **1994**, 116, 4898.
- (36) Hernández-Laguna, A., Cruz-Rodríguez, Z., Smweyer, Y. G., Arteca, G. A., Abboud, J. L. M., Tapia, O. J. *Mol. Struct (Theochem)* **1995**, 225, 77.
- (37) Karpinska, G., Dobrowski, J. C., Mazurek, A. P. *J. Mol. Struct (Theochem)* **1996**, 369, 137.
- (38) Ramírez, G. J., Tuñon, I., Colado, J. A., Silla, E. *J Am Chem Soc* **2003**, 125, 2328.
- (39) Luque, F. J., López-Bes, J. M, Cemeli, J., Aroztegui, M., Orozco, M. *Theor. Chem. Acc.* **1997**, 96, 105.
- (40) Tirado-Rives, J. J. W. L. *J Med Chem.* **2006**, 49, 5880.

Capítulo 5

Protoglobina de *Methanosarcina* *acetivorans*

Objetivos

Los estudios llevados a cabo para esta proteína han estado dirigidos a:

- Determinar la influencia de la distorsión del grupo hemo en la afinidad por O₂
- Caracterizar la dinámica de MaPgb en solución analizando el efecto de la dimerización y la unión del ligando y el impacto en la difusión de O₂.

5.1 Impacto de la distorsión del hemo en la afinidad por O₂

Para determinar el rol que juega la distorsión del hemo en la afinidad por O₂ en MaPgb, es útil realizar una clasificación sistemática de la distorsión en términos de la Descomposición Estructural en Coordenadas Normales (Normal-Coordinate Structural Decomposition; NSD), propuesta por Jentzen et al.¹⁻³, tal como se describe en el “Capítulo 3 – Métodos”. En aquel trabajo identificaron los seis modos normales fuera del plano (saddling, ruffling, doming, X-waving, Y-waving y propellering) –que abreviaremos sad, ruf, dom, wax, way y pro, respectivamente- y en el plano (meso-stretching, N-pyrrole stretching, pyrrole traslation (X,Y), breathing y pyrrole rotation) – que abreviaremos mst, nst, trx/try, bre y rot, respectivamente- más relevantes para la descripción de la distorsión del hemo respecto a una referencia de estructura D_{4h}.

5.1.1 Estructuras hexa y pentacoordinada y limitaciones del modelo

En la Tabla 5.1-1 se resumen las distorsiones fuera del plano (out-of-plane; OOP) y en el plano (in plane; IP) para un hemo modelo pentacoordinado (imidazol+anillo) y hexacoordinado (imidazol+anillo+O₂) según la descomposición en las coordenadas de base mínima del método NSD. Se observan diferencias importantes para ambas estructuras. La especie pentacoordinada posee dom negativo debido a la salida del Fe del plano porfirínico y bre negativo. La hexacoordinada, por otro lado, está esencialmente deformada en la coordenada dada por el modo ruf.

Tabla 5.1-1. Proyección de la distorsión en NSD para un hemo modelo

OOP	penta	hexa	IP	penta	Hexa
sad	-0.030	-0.010	mst	0.000	0.036
ruf	0.000	0.234	nst	0.041	0.000
dom	-0.231	-0.043	trx	0.001	0.001
wax	0.000	-0.004	try	0.000	-
way	0.001	-0.008	bre	-	-
pro	0.000	-0.000	rot	0.000	0.000

En ausencia de distorsión, la energía de unión está determinada por la posición del hierro respecto a los átomos de nitrógeno pirrólicos y respecto al imidazol. Cuando el hemo se encuentra enterrado dentro de la matriz proteica la posición del hierro no sólo depende de la distancia relativa a la histidina proximal, sino también de la distorsión del grupo prostético. A su vez, la posición de la histidina proximal depende de la proteína. En el presente trabajo el imidazol es un modelo para una histidina proximal ubicada en la posición óptima para cada estructura (hexa o pentacoordinada). Evidentemente esto no es siempre así para un par hemo+histidina proximal real dentro de la proteína. En una proteína, la histidina proximal puede encontrarse en una posición cercana a la óptima para penta o hexacoordinación, o bien en una posición intermedia, o puede suceder que no haya restricciones en la posición de la histidina. Los cálculos realizados asumen esta última situación, aunque es de esperar que la mayor parte de las proteínas impongan alguna restricción sobre la posición de la histidina proximal. Por lo tanto, la validez de los cálculos realizados se sustenta en la hipótesis de ausencia de restricciones en la relajación geométrica de la histidina proximal.

5.1.2 Efecto de las distorsiones fuera del plano (OOP)

En la figura Figura 5.1-1 se muestra el coste energético asociado a la distorsión a lo largo del modo ruf. En la figura superior se muestra el coste (ΔE) para las formas oxy y deoxy, y en la inferior la diferencia entre ambas ($\Delta(\Delta E)$). De modo equivalente, pero

con el ΔE representado en la figura interna se representan los otros 3 modos más blandos fuera del plano en la Figura 5.1-2. El cero de energía de los perfiles para las formas oxy y deoxy (ΔE) es el de la estructura de mínima energía. Haciendo la diferencia entre ambas curvas se obtiene una estimación del efecto de cada modo en la afinidad por O_2 respecto a un hemo perfectamente plano, que es tomado como referencia ideal ($\Delta(\Delta E)$).

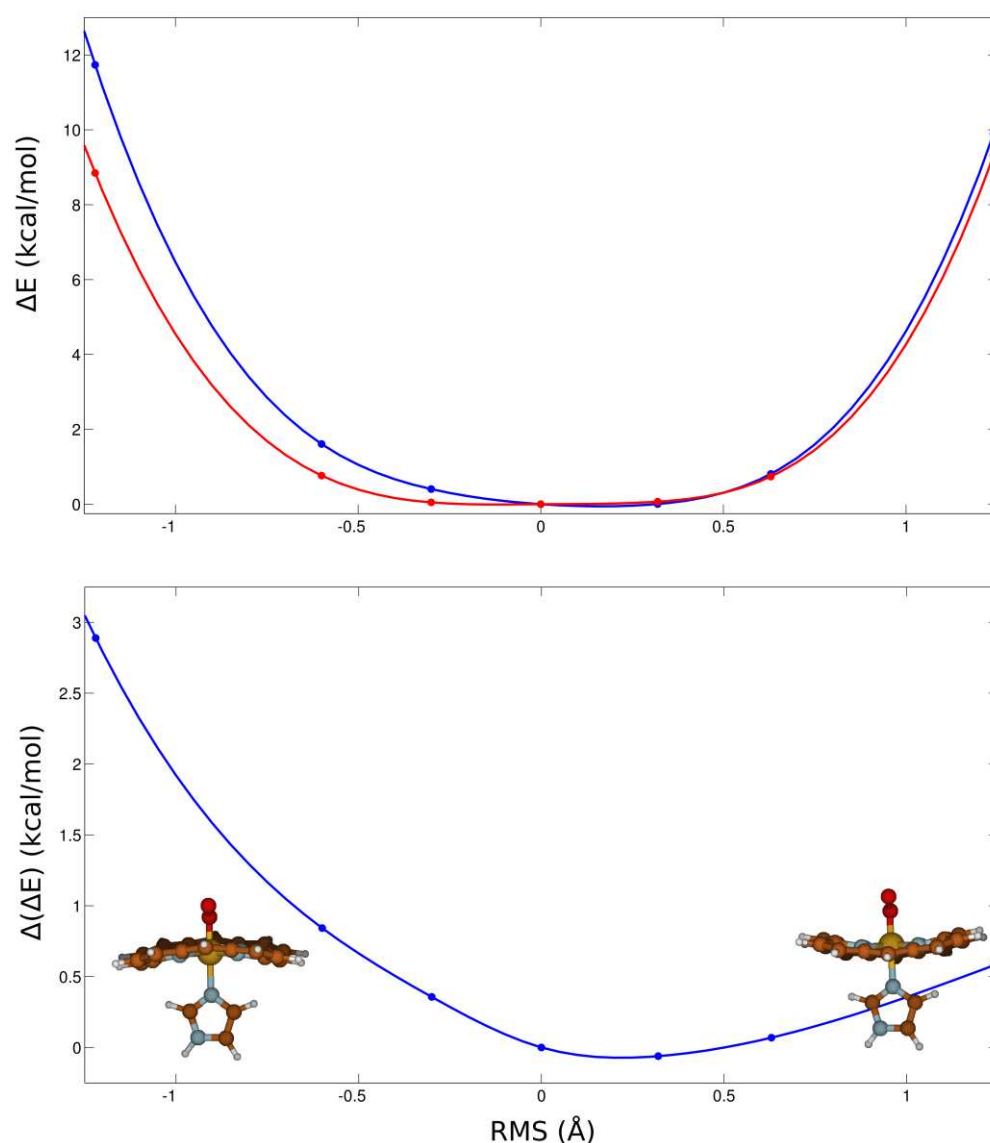


Figura 5.1-1. Efecto de la distorsión a lo largo del modo *ruf* en la afinidad por O_2 (ΔE) para la forma oxy (azul) y deoxy (rojo) –superior- y la diferencia entre ambas ($\Delta(\Delta E)$) –inferior--

La tendencia general que se observa es que todas las distorsiones fuera del plano disminuyen la afinidad por O₂. Esto es así dado que tanto distorsiones positivas como negativas en todos los casos muestran un $\Delta(\Delta E)$ positivo (a excepción de pequeñas desviaciones positivas de la planaridad para ruf y dom). En la Tabla 5.1-2 se presentan algunos valores que ilustran esta tendencia.

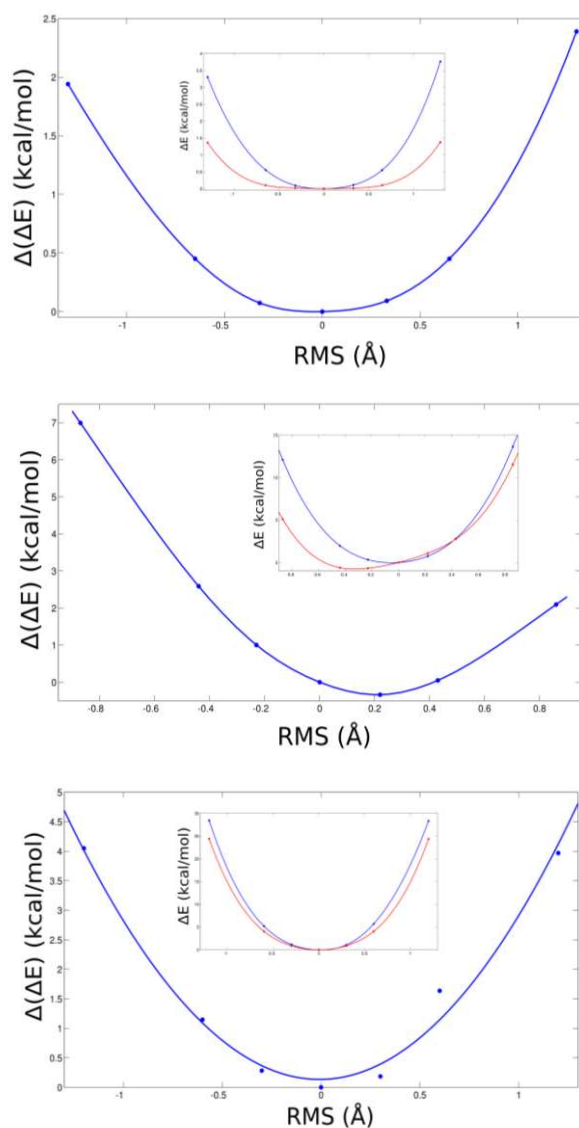


Figura 5.1-2. Efecto de la distorsión a lo largo de los modos *sad*, *dom*, *wax* en la afinidad por O₂ (ΔE) para la forma oxy (azul) y deoxy (rojo) –figuras internas- y la diferencia entre ambas ($\Delta(\Delta E)$) –figuras principales-

Tabla 5.1-2. Valores de $\Delta(\Delta E)$ (kcal/mol) para distorsiones (en Å) a lo largo de los 6 modos fuera del plano

RMSD (Å)	$\Delta(\Delta E_{O_2})$						
	-1.2	-0.6	-0.3	0.0	0.3	0.6	1.2
sad	1.9	0.5	0.1	0.0	0.1	0.5	2.4
ruf	2.9	0.8	0.4	0.0	-0.1	0.1	0.6
dom	7.0	2.6	1.0	0.0	-0.3	0.0	2.1
wax/way	3.6	0.7	0.2	0.0	0.3	1.2	3.5
pro	2.2	0.5	0.1	0.0	0.1	0.5	2.2

La asimetría observada en algunos modos puede ser atribuida a la posición del mínimo en las curvas para los estados oxy y deoxy, así como en la curvatura de los perfiles de energía alrededor del mínimo. Por otro lado, pese a la tendencia de todos los modos a disminuir la afinidad por O_2 , la magnitud del efecto es pequeña para distorsiones bajas. Esto sugiere que la contribución a la regulación de la afinidad por oxígeno en hemoproteínas no debe ser importante.

5.1.3 Efecto de las distorsiones en el plano (IP)

En la Figura 5.1-3 se muestra el coste energético asociado a la distorsión a lo largo del modo breathing (bre). Al igual que en la Figura 5.1-4, en la parte superior se muestra el coste (ΔE) para las formas oxy y deoxy, y en la inferior la diferencia entre ambas ($\Delta(\Delta E)$). De modo equivalente, pero con el ΔE representado en la figura pequeña, se representan otros dos de los modos en el plano: mst y nst.

La tendencia observada para los modos en el plano es la misma que para los fuera del plano: las distorsiones, tanto negativas como positivas, disminuyen la afinidad por O_2 . Sin embargo, para el caso de los modos en el plano existe una excepción: el modo bre. Para este desplazamiento normal, la posición del mínimo y la diferencia en la curvatura para las formas oxy y deoxy dan lugar a una tendencia distinta. Sorprendentemente, aunque para valores negativos de distorsión (expansión del hemo) la tendencia es la misma que para el resto, la tendencia se invierte hacia valores

positivos (compresión del hemo) de bre. Además, se observa que la afinidad por O₂ es mucho más sensible a la distorsión en bre que para los modos fuera del plano. Esto se ve claramente teniendo en cuenta que para sólo 0.5 Å en el RMSD $\Delta(\Delta E)$ alcanza valores cercanos a 12 kcal/mol. Este comportamiento diferenciado del modo bre se ve claramente reflejado en la Tabla 5.1-3

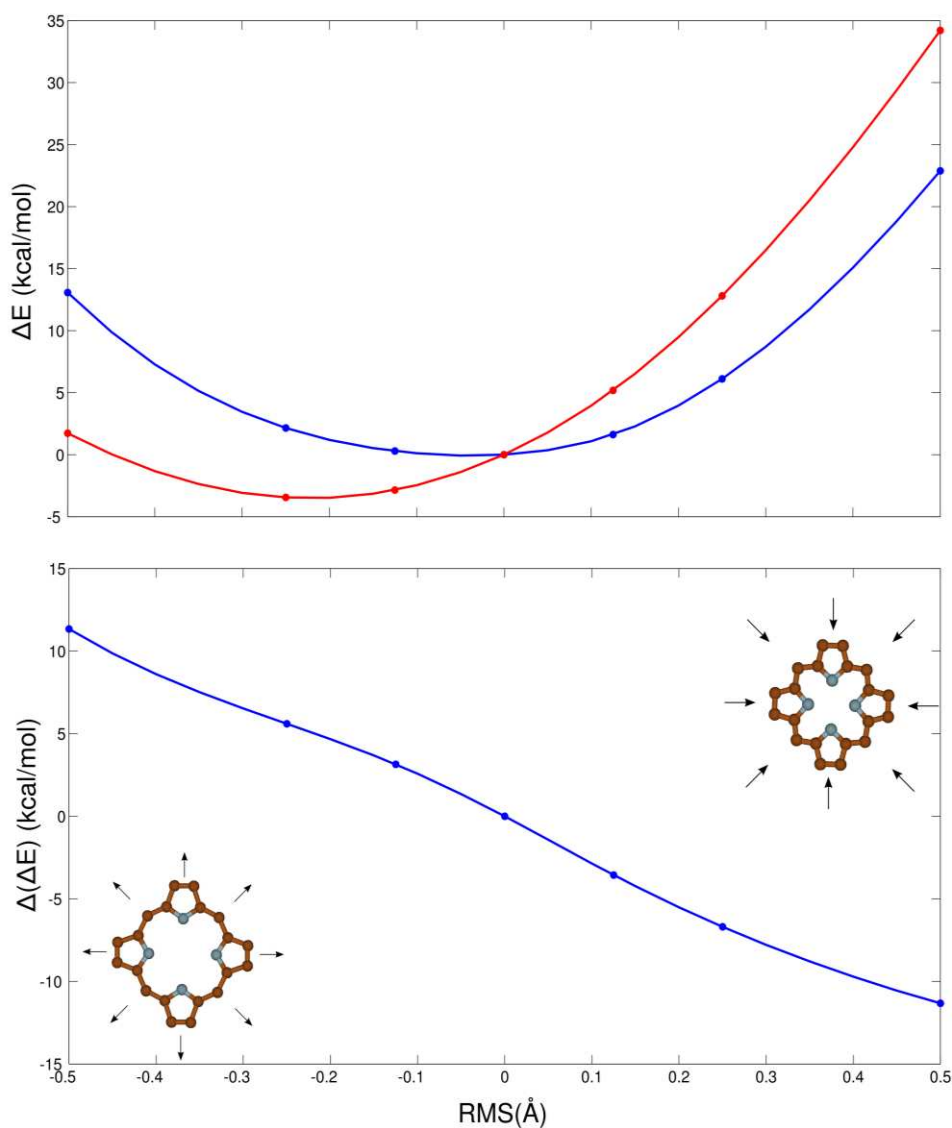


Figura 5.1-3. Efecto de la distorsión a lo largo del modo *bre* en la afinidad por O₂ (ΔE) para la forma oxy (azul) y deoxy (rojo) –superior- y la diferencia entre ambas ($\Delta(\Delta E)$) –inferior-

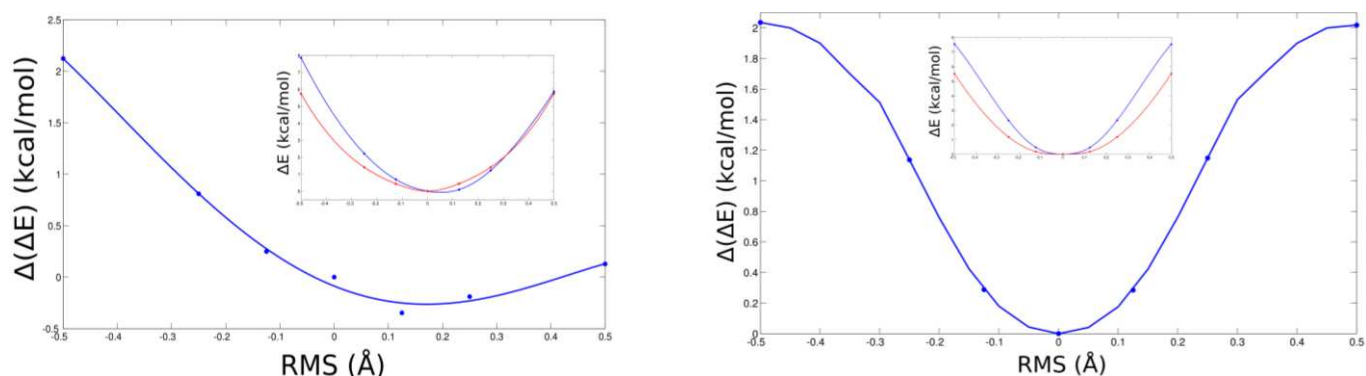


Figura 5.1-4. Efecto de la distorsión a lo largo de los modos *mst*, *nst* en la afinidad por O₂ ($\Delta(\Delta E)$) para la forma oxy (azul) y deoxy (rojo) –figuras internas- y la diferencia entre ambas ($\Delta(\Delta E)$) –figuras principales-

Tabla 5.1-3 - Valores de $\Delta(\Delta E)$ (kcal/mol) para distorsiones (en Å) a lo largo de los 6 modos en el plano. En negrita el aumento de afinidad para valores positivos de *bre*

RMS	$\Delta(\Delta E_{O_2})$						
	-0.5	-0.25	-0.125	0.0	0.125	0.25	0.5
<i>mst</i>	2.1	0.8	0.2	0.0	-0.3	-0.2	0.1
<i>nst</i>	2.0	1.1	0.3	0.0	0.3	1.1	2.0
<i>trx/try</i>	0.3	0.1	0.0	0.0	0.0	0.0	0.2
<i>bre</i>	11.3	5.6	3.1	0.0	-3.6	-6.7	-11.3
<i>rot</i>	2.0	0.6	0.2	0.0	0.0	0.2	1.3

5.1.4 Análisis de la alta afinidad de *MaPgb* por O₂

Las globinas que exhiben alta afinidad por O₂ (en particular baja k_{off}) usualmente presentan una red de puentes de hidrógeno entre el ligando y residuos presentes en la cavidad distal (Figura 5.1-5). Por lo tanto, al encontrarse una nueva globina con un valor para la k_{off} tan bajo como es el caso de la *MaPgb* (0,092-0,0094 s⁻¹), cabe pensar que deben existir residuos capaces de estabilizar al ligando por la formación de puentes de hidrógeno cuando éste se une al hemo. Sin embargo, tal como se observa en la Figura 5.1-5, el único residuo que sería capaz de formar un puente de hidrógeno con el

ligando para MaPgb es la Tyr(61)B10, tal como se ha visto en otras globinas⁴⁻⁷. En el cristal este residuo se halla lejos del ligando, formando un puente de hidrógeno con el carbonilo de LeuE4 (Leu86).

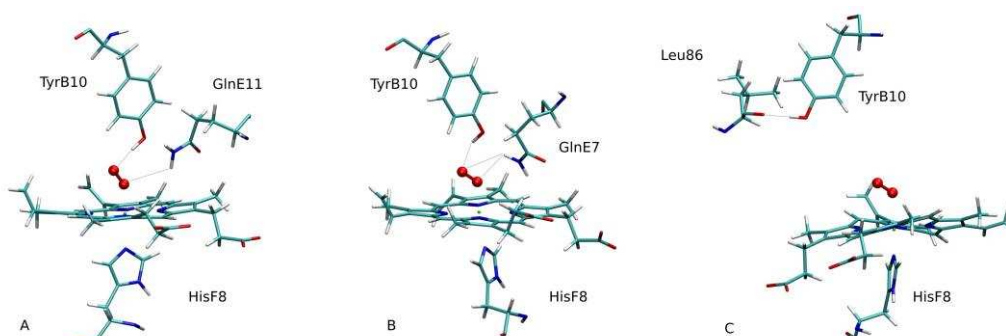


Figura 5.1-5 - Sitio activo y residuos donores de puente de hidrógeno en la cavidad distal de Hemoglobina de *Ascaris* (A), Hemoglobina truncada N de *Mycobacterium tuberculosis* (B) and Protoglobina de *Methanosarcina Acetivorans* (C).

Para evaluar la participación de este residuo en la alta afinidad reportada de MaPgb para O₂, se ha seguido la evolución temporal por Dinámica Molecular (véase la sección 3.2 para un análisis completo) de la distancia del grupo hidroxilo de TyrB10 con el ligando y con el carbonilo de LeuE4, tal como se muestra en la Figura 5.1-6 para el dímero oxigenado. El grupo hidroxilo de TyrB10 se acerca al ligando (3 Å) en contadas ocasiones a lo largo de la trayectoria, pero se encuentra la mayor parte del tiempo lejos del mismo (6 Å), manteniendo una interacción de puente de hidrógeno con Leu94. El mismo comportamiento se da en el monómero oxigenado, confirmando que no existe una relación entre la k_{off} reportada y la estabilización por puentes de hidrógeno por residuos de la cavidad distal.

Estos resultados nos llevan a pensar que las hemoproteínas podrían explotar el modo bre como un mecanismo adicional para regular la afinidad por ligandos dado que valores positivos o negativos para este modo aumentan o disminuyen, respectivamente, la afinidad. En particular, podría existir una relación entre la inusual afinidad por O₂ reportada para MaPgb y la elevada distorsión del hemo que se observa en la estructura de rayos X de alta resolución (1.3 Å) depositada en el PDB con código 2VEB.

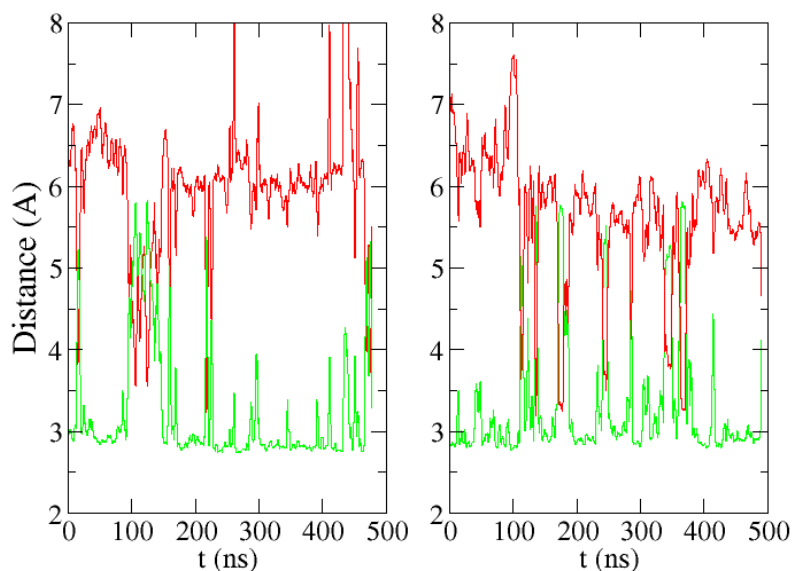


Figura 5.1-6 - Evolución a lo largo de 500 ns de simulación por dinámica molecular de la distancia entre el grupo hidroxilo de TyrB10 con O₂ (rojo) y con LeuE4 (verde) para el dímero oxigenado de MaPgb. Ambas subunidades del dímero están representadas a izquierda y derecha

Se han utilizado métodos híbridos QM/MM para estimar las energías de unión de O₂ en distintas globinas permitiendo identificar efectos dinámicos, distales y proximales que pueden afectar la k_{off} ^{4,8-13}. La Tabla 5.1-4 muestra algunos valores de energía de unión ($\Delta(\Delta E)$, respecto al hemo aislado) y la correspondiente constante de disociación (k_{off}) para varias globinas que presentan diferente número de residuos dadores de puente de hidrógeno en la cavidad distal (ver Figura 5.1-5). Aquéllas que poseen 2 puentes de hidrógeno (AscHb, MtTrHbN) exhiben un mayor aumento en la energía de unión, seguidas de Mb y de la mutante YB10A de trHbN, que sólo poseen 1. Finalmente, la mutante HE7V de Mb, carente de puentes de hidrógeno que puedan estabilizar al ligando, prácticamente recupera el valor de energía de unión del hemo libre.

La excepción a la tendencia observada para todos los casos descritos es la MaPgb que, como se mencionó anteriormente, no posee ningún residuo capaz de formar un puente de hidrógeno permanente que estabilice al ligando. Esto refuerza la hipótesis de que la alta afinidad por O₂ está relacionada, al menos en parte, con la distorsión del hemo.

Tabla 5.1-4 - Valores de k_{off} (s^{-1}) y $\Delta(\Delta E)$ (kcal/mol) de cálculos QM/MM (relativos al hemo aislado)

Proteína	$\Delta(\Delta E)$	Puentes de H con O_2	k_{off} (s^{-1})
Pgb	-7.6	0	$9.4 \times 10^{-3} - 9.2 \times 10^{-2}$
Mb	-5.0	1	12
Mb HE7V	-0.7	0	1600
trHbN	-15.2	2	0.2
trHbN YB10A	-7.8	1	45
AscHb	-12.3	2	0.4×10^{-2}

En la Tabla 5.1-5 se presenta el análisis de NSD para el hemo de MaPgb, incluyendo el grupo hemo de Mb como referencia. Las diferencias más notables las encontramos en el modo ruf y bre, donde MaPgb presenta 1.42 Å y 0.202 Å contra 0.02 Å y -0.023 Å de Mb. Estas diferencias ciertamente deben influir en la afinidad por O_2 para estas globinas. Es interesante resaltar que un valor tan positivo para bre implica un hemo comprimido, lo cual es un hecho estructural altamente inusual. En la Figura 5.1-7 y la Tabla 5.1-6 se muestra una comparación entre MaPgb y Mb de parámetros estructurales del hemo que dan cuenta de esta compresión.

Tabla 5.1-5 - Análisis por NSD de los hemos oxigenados de mioglobin (Mb, PDB 1A6M) y protoglobin (MaPgb, PDB 2VEB). Valores en Å

OOP	Mb	Pgb	IP	Mb	Pgb
sad	-0.185	-0.390	mst	-0.017	-0.042
ruf	-0.017	+1.419	nst	-0.021	+0.006
dom	+0.206	-0.013	trx	+0.000	+0.055
wax	-0.026	-0.132	try	-0.014	+0.087
way	+0.186	+0.235	bre	-0.023	+0.202
pro	-0.071	+0.019	rot	+0.012	+0.013

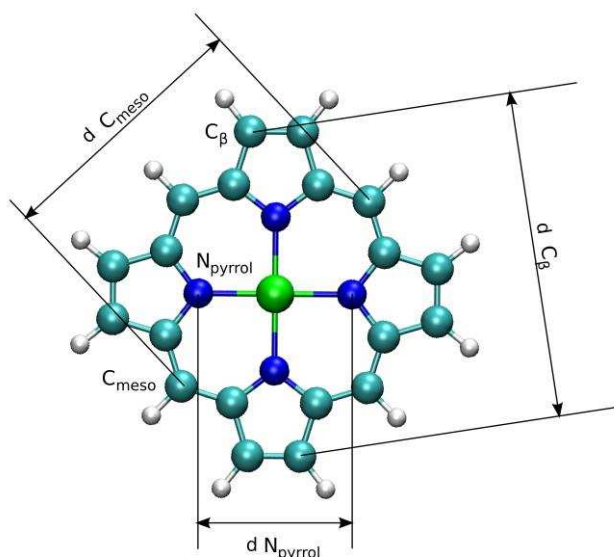


Figura 5.1-7 - Definición de parámetros estructurales utilizados en la Tabla 5.1-6 para evaluar la compresión-expansión del hemo

Tabla 5.1-6 - Parámetros estructurales del hemo (definidos en Figura 5.1-7) para Mb (1A6M), referencia de porfirina de Fe(II) y MaPgb (2VEB). Valores en Å

	Mb (1A6M)	PorFe(II) 4c (referencia)	Pgb (2VEB)
$d C_{\text{meso}}$	6.80 ± 0.01	6.84	6.65 ± 0.04
$d N_{\text{pirrol}}$	4.01 ± 0.01	3.98	3.963 ± 0.008
$d C_{\beta}$	8.54 ± 0.03	8.53	8.44 ± 0.04
$d \text{Fe-N}$	2.01 ± 0.01	1.99	1.99 ± 0.01

Los parámetros estructurales presentados en la Tabla 5.1-6 son consistentes con la descomposición por NSD para Mb y MaPgb. La compresión se ve claramente expresada en términos de las distancias entre C_{meso} , C_{β} y N_{pirrol} , puesto que en todos los casos MaPgb presenta valores inferiores a la referencia (en 0.2 \AA , 0.1 \AA y 0.02 \AA , respectivamente) y a Mb. La Mb, por el contrario, coincide prácticamente con la estructura de referencia.

Para evaluar el impacto de la distorsión del hemo en la energía de enlace del ligando, resulta útil examinar las contribuciones de los modos mencionados a la energía de unión respecto al hemo modelo. Esto se presenta en la Tabla 5.1-7, donde la evaluación se ha llevado a cabo a distintos niveles de teoría por métodos de estructura electrónica. Pese a las diferencias en cuanto a la magnitud del efecto, en todos los casos

la conclusión es la misma: el modo bre implica una estabilización en la energía de enlace que supera la desestabilización por ruf. Por ejemplo, si tomamos B3LYP/6-31G**, mientras que la distorsión fuera del plano desestabiliza 1.5 kcal/mol, la distorsión en el plano estabiliza 4.7 kcal/mol, resultando en una estabilización neta de 3.2 kcal/mol de MaPgb respecto al hemo modelo.

Tabla 5.1-7 - Efecto de la distorsión en el plano (IP) y fuera del plano (OOP) del hemo de MaPgb en un hemo modelo en la energía de unión a distintos niveles de teoría (Valores en kcal/mol)

Hemo modelo de MaPgb	B3LYP 6-31G**	PBE 6-31G**	OLYP 6-31G**	B3LYP TZVPP (CHNO) QZVPP (Fe)	PBE (SIESTA)
OOP+IP	-3.2	-2.1	-4.3	-4.3	-2.1
Only IP	-4.7	-4.0	-5.5	-4.9	-4.1
Only OOP	1.5	1.9	1.1	0.6	2.0

5.1.5 Efectos en la distribución de densidad electrónica

La distorsión del hemo produce una redistribución de la densidad electrónica que afecta a las distancias de enlace y permite comprender las tendencias observadas en la afinidad por O₂. En concreto, la influencia ejercida por el modo bre se presenta en la Tabla 5.1-8, donde se muestran valores para bre de -0.5, 0 y +0.5. El análisis se ha hecho utilizando la metodología NBO.

En el modelo sin distorsión la carga neta obtenida para el O₂ es -0.343e y se ve prácticamente inalterada para valores negativos de bre (expansión). Por el contrario, se observa un enriquecimiento de densidad electrónica sobre el O₂ (-0.357e) para valores positivos de bre (compresión). Respecto al ligando, mientras que para el hemo expandido ambos átomos poseen cargas comparables (O1/O2: -0.169e/-0.173e), el hemo sin distorsión y el comprimido muestran una asimetría en la densidad electrónica, que incrementa la carga negativa en el oxígeno no ligado al Fe (O1/O2: -0.156e/-0.187e y -0.161e/-0.196e, respectivamente). El átomo de Fe sufre un aumento de 0.045e de carga positiva con la expansión, mientras que la carga se ve reducida en 0.054e en la compresión. El imidazol no parece sufrir alteraciones en su densidad electrónica con el

bre. Los átomos de N del anillo porfirínico sufren una modesta ganancia en densidad electrónica (-0.041e) con la expansión, y la pierden (-0.064e) con la compresión. Esto es compatible con las variaciones de carga observadas para Fe-O₂: +0.046e para expansión y -0.068e para compresión.

El efecto neto en MaPgb es una transferencia de densidad electrónica del anillo porfirínico al O₂, dado que presenta bre positivo (compresión). Esto produce cambios en las distancias de enlace. Mientras que para la expansión se acortan los enlaces Fe-O1 y Fe-N_{im}, la compresión del anillo porfirínico conlleva un alargamiento de las mismas, tal como se muestra en la Tabla 5.1-8. La compresión del hemo que presenta MaPgb desplaza densidad electrónica de los orbitales ecuatoriales hacia los axiales, capaces así de interactuar con O₂ y el imidazol a distancias mayores. A su vez, se produce un alargamiento del enlace O-O por el aumento de densidad electrónica en el orbital antiligante.

Tabla 5.1-8 – Análisis de carga NBO y distancias de enlace para una porfirina expandida o comprimida. q_{O1} es la carga del oxígeno unido al Fe y q_{O2} la del restante. En paréntesis la carga de la molécula de O₂

bre (Å)	d(O-O)	d(Fe-O)	q_{O1}/q_{O2} (q_{O2})	q_{Fe}	q_{im}
-0.5	1.268	1.737	-0.169/ -0.173 (-0.342)	1.505	0.128
0.0	1.271	1.741	-0.156/-0.187 (-0.343)	1.460	0.125
0.5	1.274	1.745	-0.161/-0.196 (-0.357)	1.406	0.122

5.2 *Dinámica Molecular de MaPgb: relación entre la estructura cuaternaria y la unión de ligando*

En esta sección nos centraremos en el análisis estructural y dinámico de MaPgb, prestando atención a la influencia de la estructura cuaternaria y a la presencia/ausencia de ligando unido al hemo. Estudiaremos las características estructurales de la proteína a lo largo de simulaciones de dinámica molecular, así como la migración de ligandos diatómicos desde el solvente a la cavidad distal del grupo hemo.

Puesto que la proteína fue cristalizada como un dímero y ésta es también la estructura cuaternaria en experimentos por filtración en gel, se han llevado adelante simulaciones por dinámica molecular tanto de la estructura dimérica como monomérica. La última interesa como referencia para estudiar el efecto de la dimerización en las propiedades estructurales y dinámicas de la proteína. Es a su vez relevante puesto que una gran parte de las globinas se halla como monómero.

Para cada una de las simulaciones se han recogido trayectorias para las formas oxigenada y libre del grupo hemo partiendo de las estructuras de rayos X obtenidas por el grupo de Martino Bolognesi¹⁴ y depositadas en el PDB con códigos 2VEE y 2VEB. De este modo se han completado un total de 2 μ s de dinámica molecular divididos en 4 simulaciones de 500 ns de monómero oxigenado (MPgb-O₂), monómero deoxigenado (MPgb-deoxy), dímero oxigenado (DPgb-O₂) y dímero deoxigenado (DPgb-deoxy). Todas las simulaciones fueron llevadas a cabo a 1 atm y 298K en una caja de aguas explícitas luego de un adecuado equilibrado. Se utilizó el programa AMBER9¹⁵ y el campo de fuerzas parm99SB¹⁶. Detalles adicionales de las condiciones de simulación pueden ser encontrados en el artículo enviado a publicación: “Small Molecule Migration in Methanosarcina acetivorans Protoglobin: Effect of Ligand Binding and Quaternary Structure” (Forti F. et al.).

5.2.1 Evolución estructural de *MaPgb*

La evolución estructural a lo largo de la trayectoria se puede estimar en base a la desviación cuadrática media -root mean square deviation o rmsd- del backbone. En este estudio en concreto, es de particular relevancia para evaluar las diferencias estructurales observadas entre las simulaciones de las distintas estructuras cuaternarias (monómero versus dímero) y el rol de la presencia/ausencia de ligando unido al hemo.

En la Figura 5.2-1 se representa el rmsd a lo largo de las 4 simulaciones realizadas respecto a las estructuras iniciales. Éste fue calculado para el backbone descartando los residuos de los extremos N (1-22) y C (187-196) terminal, que carecen de una estructura secundaria bien definida, y que por ello introducen ruido en la comparación de la evolución estructural de cada trayectoria.

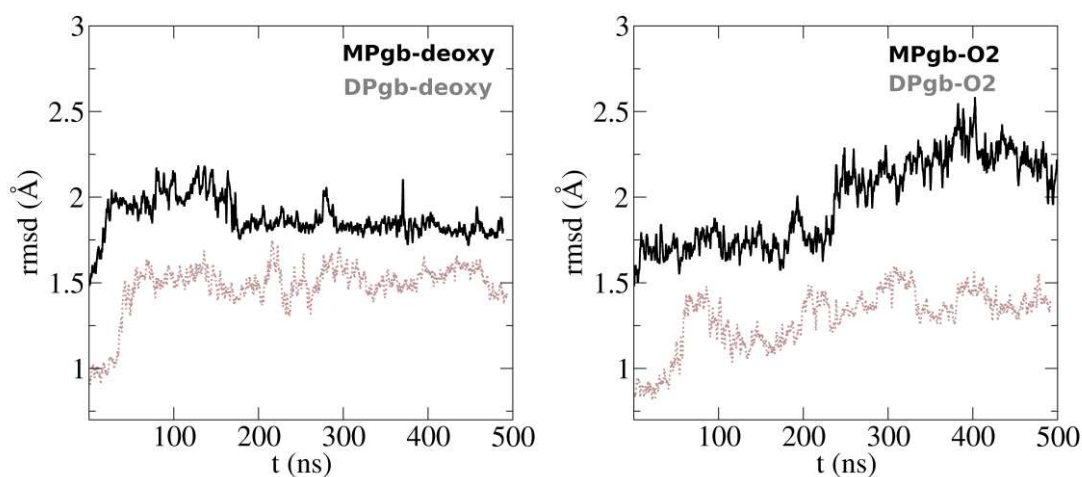


Figura 5.2-1. rmsd (Å) del backbone (residuos 23-189) a lo largo de 500 ns de DM para los monómeros (negro) y dímeros (gris) en las formas oxigenada (derecha) y deoxigenada (izquierda) respecto a las estructuras cristalográficas (códigos PDB 2VEB y 2VEE).

La primera conclusión que se extrae de la Figura 5.2-1 es que los monómeros muestran un mayor desplazamiento respecto a la estructura inicial que los dímeros (~ 2 Å versus ~ 1.5 Å). Este resultado es esperable si se tiene en cuenta que las estructuras iniciales utilizadas para las simulaciones corresponden a dímeros (códigos PDB 2VEB y 2VEE). Pese a que no se trata de un cambio importante en términos de la estructura general de la proteína, es evidente que la ausencia de la segunda subunidad provoca reajustes estructurales en el caso de los monómeros que no se observan cuando ambas subunidades se hallan formando el dímero, es decir, la estructura de la proteína nota una diferencia en la exposición a solvente del área de contacto entre subunidades, básicamente las hélices G y H: el par G/H-H/G que forma la interface entre ambas subunidades se pierde en el caso de los monómeros.

En este punto resulta útil realizar un análisis detallado del rmsd por hélice, como se presenta en la figura Figura 5.2-2. El desplazamiento más importante corresponde a la hélice G en el monómero deoxigenado, que alcanza casi los 3 Å, mientras que para el dímero sólo llega a 1.2 Å. Como se comentó, esta hélice forma la superficie de contacto con la otra subunidad en el dímero por lo que la exposición a solvente en el monómero impacta directamente sobre la estructura de la hélice.

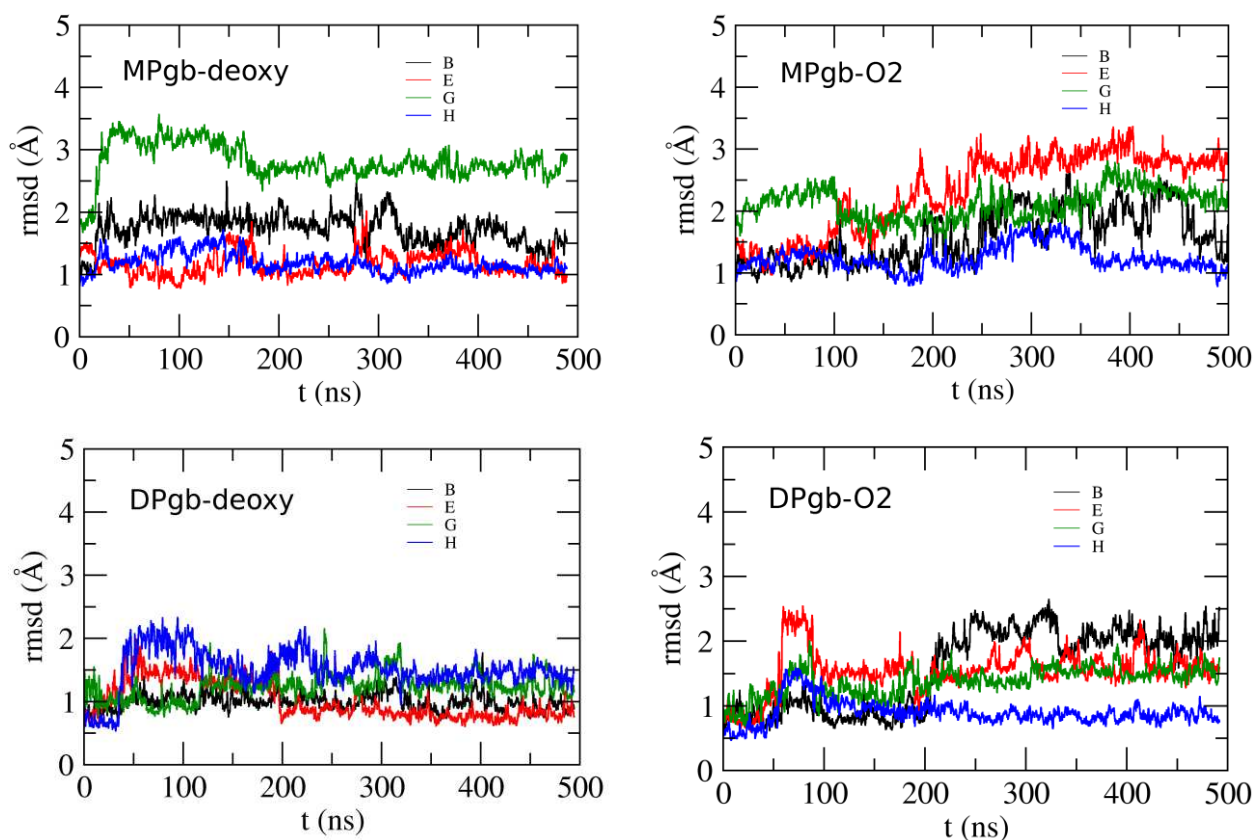


Figura 5.2-2. rmsd (Å) del backbone de las hélices A, E, G y H a lo largo de 500 ns de DM para los monómeros (superior) y dímeros (inferior) en las formas oxigenada (derecha) y deoxigenada (izquierda) respecto a las estructuras cristalográficas (códigos PDB 2VEB y 2VEE)

Aunque esto también sucede para el caso del monómero oxigenado respecto al dímero oxigenado, el cambio es menor (2.2 Å versus 1.6 Å) y el efecto parece propagarse a otras hélices: B y E. La hélice E para MPgb-O₂ alcanza un rmsd de casi 3 Å, mientras que está entre 1.5 Å y 2 Å para casi toda la trayectoria de DPgb-O₂. La hélice B en ambos casos oscila alrededor de los 2 Å.

Los resultados sugieren que la relajación de la hélice G está relacionada con la exposición a solvente que sufre la misma en los monómeros. Sin embargo, las características estructurales de las hélices B y E parecen estar relacionadas además con la presencia o ausencia del ligando unido al hemo. La hélice B muestra mayor variabilidad en las simulaciones realizadas para las formas oxigenadas, y la hélice E un mayor desviación estructural para estos casos que para las formas deoxigenadas. Dado que estas 3 hélices son las que delimitan los túneles para migración de ligandos en MaPgb, las fluctuaciones estructurales observadas comparando dímero/monómero y oxy/deoxy podrían ser relevantes en términos de la funcionalidad de la proteína.

Con el objeto de profundizar la caracterización de las diferencias estructurales y dinámicas y estructurales entre las 4 simulaciones, se llevó adelante un análisis de dinámica esencial (DE). Los autovalores disminuyen suavemente en todos los casos, lo que sugiere una alta riqueza dinámica de la proteína. Si nos concentramos en el primer modo esencial (que da cuenta de entre un 13% y 19% de la fluctuación total) observamos una clara distinción entre las formas oxigenada y deoxigenada de la proteína, tal como se aprecia en la Figura 5.2-3.

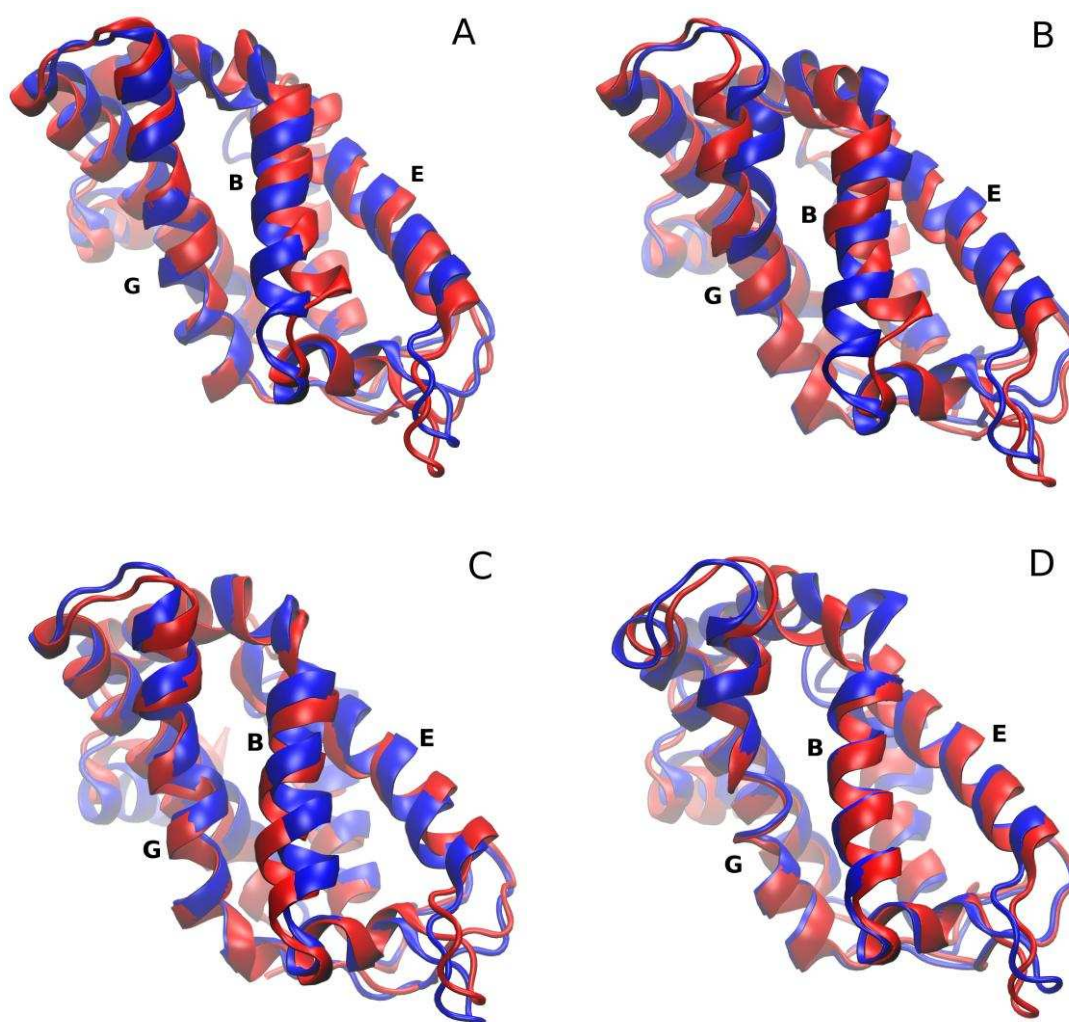


Figura 5.2-3. Representación de la deformación a lo largo del primer modo esencial para el backbone de (A) MPgb-O₂, (B) DPgb-O₂, (C) MPgb-deoxy, and (D) DPgb-deoxy

Mientras que las formas deoxigenadas concentran el modo esencial principal en la hélice A (y las partes cercanas de G y H), las formas oxigenadas muestran mayores fluctuaciones en las hélices B y E. Es particularmente relevante el movimiento de la

hélice B para las formas oxigenadas en contraste con las deoxigenadas, dado que determina la forma de ambos túneles. Esto sugiere una posible relación entre la migración del ligando al sitio distal y la unión del mismo al hemo.

Teniendo en cuenta los resultados expuestos resulta lógico pensar que las diferencias estructurales observadas entre dímeros y monómeros, así como proteína con y sin ligando unido tienen un impacto no despreciable en la migración de ligandos a través de los túneles 1 y 2. Esto es así dado que las hélices B, E y G que forman este nuevo sistema de túneles evolucionan de manera diferente a lo largo de la dinámica en cada uno de los 4 casos estudiados. Por ello, estos resultados sugieren la hipótesis de que la estructura cuaternaria y la unión del ligando determinan las propiedades asociadas a la migración de ligandos.

5.2.2 Detección del ligando unido y comunicación al resto de la proteína

En la mayoría de las globinas estudiadas hasta la fecha la presencia de un ligando unido al hemo es detectado mediante la formación de puentes de hidrógeno con residuos donores presentes en la cavidad distal. Generalmente esto es comunicado al resto de la proteína induciéndose cambios conformacionales locales o globales que afectan las propiedades estructurales y dinámicas. A su vez, la formación de estos puentes de hidrógeno estabiliza al ligando unido al hemo, hallándose generalmente una relación entre el número de puentes de hidrógeno y la afinidad por el ligando. Tal es el caso de MtHbN, Mb o AscHb, como se describió en la sección 5.1.4. Sin embargo, MaPgb representa una excepción entre las globinas dado que no forma ningún puente de hidrógeno con el ligando, aún exhibiendo alta afinidad por O₂. Teniendo en cuenta los resultados expuestos en la sección anterior, en particular la ΔE , resulta evidente que deben existir otros residuos capaces de sentir la presencia del ligando.

Una detallada inspección de las estructuras recogidas a lo largo de las simulaciones demuestra que hay dos residuos aromáticos cercanos al ligando: Phe(145)G8 y Phe(93)E11. Esto es compatible con los resultados que arrojan los análisis de las estructuras de rayos X: Phe(93)E11 se halla a 3 Å del ligando en la forma oxigenada y muestra desorden conformacional, mientras que adopta una conformación definida para la forma libre¹⁴.

Phe(145)G8 está a alrededor de 6 Å del hierro en la estructura de rayos X, distancia que se mantiene en promedio a lo largo de las simulaciones, con la excepción de MPgb-deoxy. La distancia media para esta última es de alrededor de 1 Å menor (Figura 5.2-4), cercana a los 4.9 Å. Esto puede ser atribuido a la recolocación de la hélice G observada para MPgb-deoxy (Figura 5.2-2). Es interesante notar que esto no sucede para MPgb-O₂, que mantiene los 6 Å de distancia media de Phe(145)G8 al hierro.

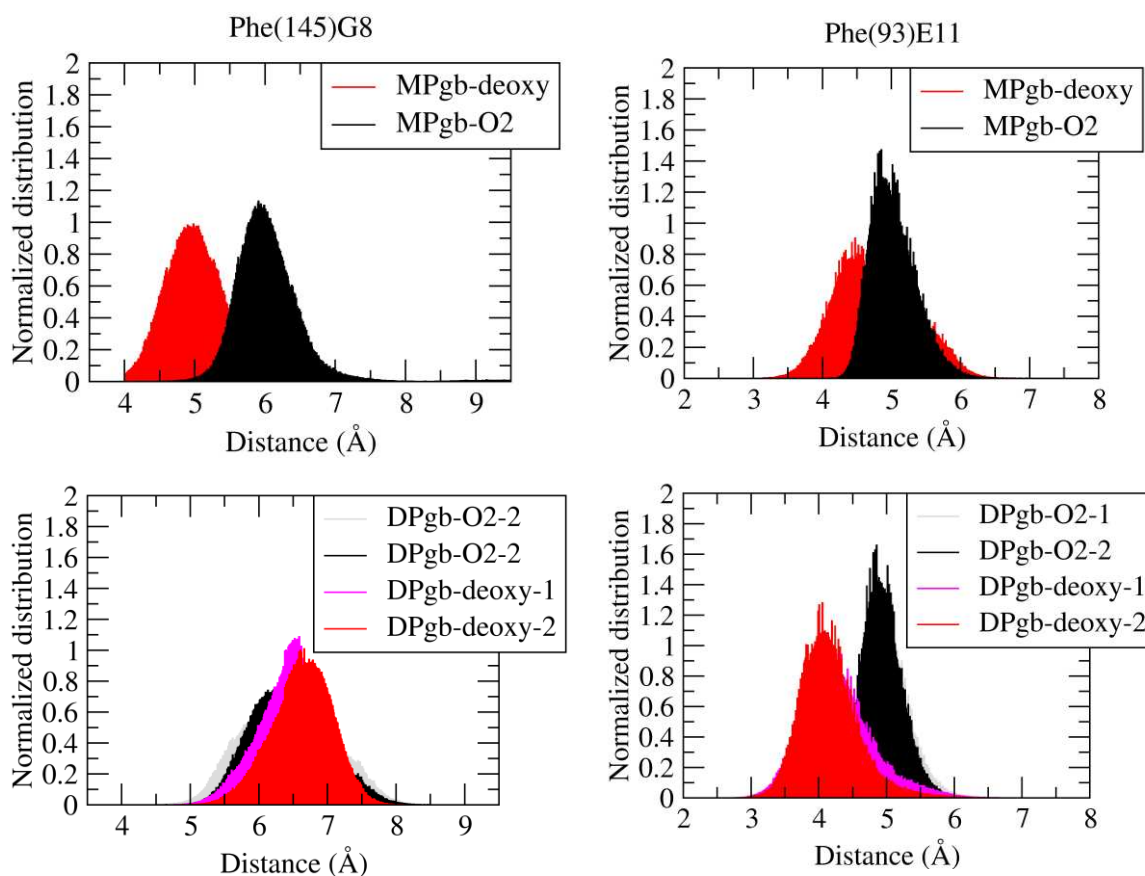


Figura 5.2-4. Distribución de distancias entre la cadena lateral de Phe(145)G8 (izquierda) y Phe(93)E11 (derecha) al hierro. Monómeros (superior) y dímeros (inferior) en las formas oxigenada y deoxigenada. Las subunidades de los dímeros están indicadas con "1" y "2"

DPgb-O₂ y DPgb-deoxy también mantienen una distancia compatible con el cristal (Figura 5.2-4). Resulta claro que la relajación de la hélice G que se observa para el monómero deoxigenado acerca la Phe(145)G8 hacia el hierro, dado que la hélice ya no se halla sostenida por la hélice H de la inexistente segunda subunidad. Esto no sucede para el monómero oxigenado pese a carecer de la segunda subunidad, puesto que la presencia del ligando impide el acercamiento del residuo al hierro.

Phe(93)E11 nota la presencia del ligando tanto en la forma monomérica como dimérica. Para las formas deoxigenadas la distancia de la cadena lateral al hierro está entre 4,0 Å y 4,5 Å. Cuando se analizan las formas oxigenadas se observa un aumento de 0,5 Å para el monómero y de 1,1 Å para el dímero. Esto significa que existe un alejamiento neto de este residuo en presencia del ligando, al margen de la estructura cuaternaria de la proteína. El impedimento estérico introducido por el ligando contribuye a los cambios estructurales y dinámicos observados en las hélices B y E, tal como se muestra en la Figura 5.2-3. Aunque no se trate de puentes de hidrógeno con el ligando como es el caso de MtHbN^{17,18}, estos resultados revelan que la presencia del ligando también puede ser comunicada al resto de la proteína mediante un mecanismo que podríamos definir como estérico.

5.2.3 Migración de ligandos

En esta sección analizaremos la influencia de las diferencias estructurales analizadas en la sección precedente en la migración de ligandos a través de los túneles 1 y 2.

En la estructura de rayos X el túnel 1, delimitado por las hélices G y B, se halla perfectamente definido. Sin embargo, cuando se analiza su evolución a lo largo de las simulaciones, se observa que Phe(145)G8 adopta también una conformación distinta a la hallada en los cristales, que bloquea el túnel. Este residuo se encuentra en 2 conformaciones a lo largo de las simulaciones que están definidas por la torsión N-C_α-C_β-C_γ: 270 y 210 grados (Figura 5.2-6). La primera coincide con la hallada en las estructuras cristalográficas (267.2 grados en 2VEB, de 264.7 a 270.2 grados en 8 subunidades independientes de 2VEE). A esta conformación la denominaremos abierta dado que la migración de ligandos no parece impedida. En la otra conformación el anillo aromático de la Phe bloquea el túnel, por lo que la llamaremos cerrada.

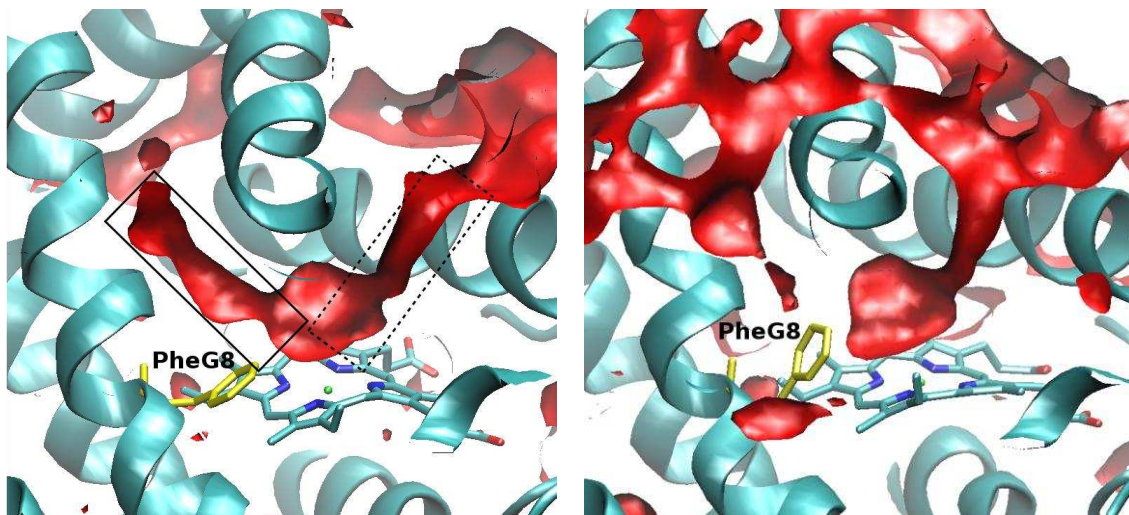


Figura 5.2-5. Isosuperficies de energía libre obtenidas por ILS para DPgb-deoxy en la forma cerrada (derecha) y abierta (izquierda) de Phe(145)G8. El túnel 1 en recuadro sólido y el túnel 2 en recuadro punteado. En amarillo la Phe(145)G8.

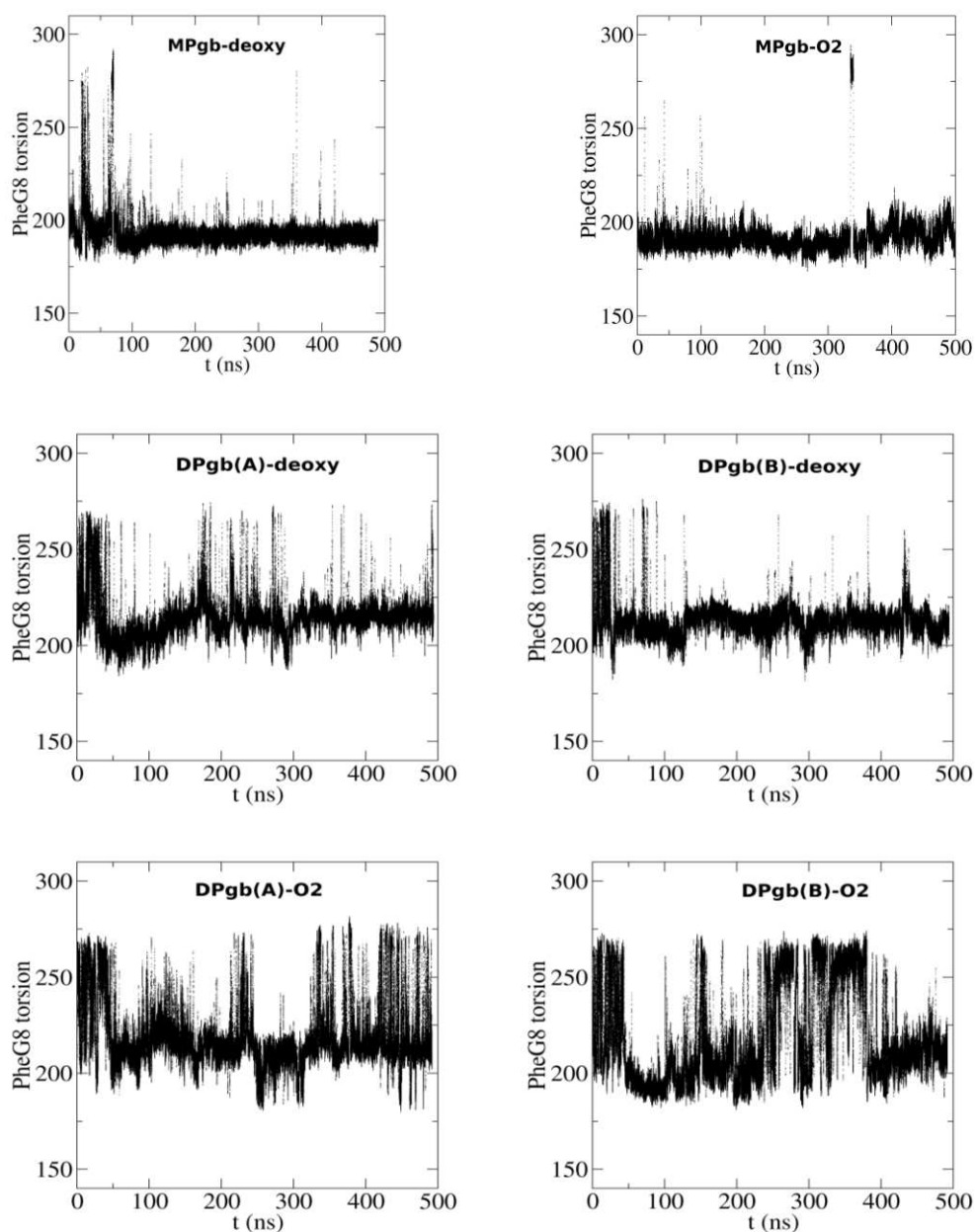


Figura 5.2-6. Evolución de la torsión N-C α -C β -C γ de Phe(145) para MPgb-deoxy (arriba, izquierda), MPgb-O2 (arriba, derecha), DPgb-deoxy, ambas subunidades (medio) y DPgb-O2 ambas subunidades (abajo)

El efecto de este residuo traducido en términos de migración de ligandos queda claramente definido si se observan las isosuperficies de energía libre obtenidas por mediante el método ILS (Figura 5.2-5). Allí se aprecia que el túnel 1 está bien definido en la forma abierta pero se halla cortado en la cerrada. Esto confirma el efecto de bloqueo a la migración de un ligando diatómico del solvente al sitio distal. En este

sentido Phe(145)G8 podría actuar como una puerta para la migración de ligandos diatómicos que recuerda al efecto de la PheE15 sobre el canal largo (LT) de MtHbtrN^{19,20}. Por el contrario, el túnel 2 está siempre abierto.

Tal como se observa en la Figura 5.2-6, que muestra la evolución del residuo G8, la proporción entre los estados conformacionales abierto/cerrado de Phe(145)G8 depende tanto de la estructura cuaternaria como de la presencia de ligando unido al hemo. Si se analizan los últimos 250 ns de cada simulación se observa que para los monómeros este residuo está en la forma cerrada el 99,9% y 98,0% de la trayectoria para la forma deoxigenada y oxigenada, respectivamente. En el caso de los dímeros se aprecia una mayor proporción de la forma abierta, alcanzando 5% y 8% en ambas subunidades de la forma deoxigenada. Para el dímero oxigenado esta tendencia se incrementa hasta alcanzar un 25% de la forma abierta. Estos resultados permiten formular la hipótesis de que la presencia de O₂ promueve una mayor apertura del túnel 1 pero que la dimerización es necesaria para consolidar este efecto. Contrariamente a lo que sucede con el túnel 1, el túnel 2 se halla siempre abierto a la migración de ligandos.

Estos resultados sugieren que MaPgb podría estar involucrada en procesos químicos bimoleculares. El primer ligando podría entrar a la cavidad distal a través del túnel 2. Si esto sucediera en la forma monomérica el túnel 1 estaría esencialmente cerrado. La dimerización facilitaría la apertura de este túnel, efecto que podría estar relacionado con la mayor flexibilidad de la hélice B promovida por la unión del primer ligando al hemo. En definitiva, la apertura del túnel 1 mediante estos mecanismos redundaría en una mayor facilidad para la captación de un segundo ligando en MaPgb.

Con el objeto de cuantificar el efecto de la estructura cuaternaria en la migración de ligandos se llevaron a cabo simulaciones de Multiple Steered Molecular Dynamics (MSMD). Éstas permiten estimar el perfil de energía libre asociado a la migración del ligando desde el solvente a la cavidad distal mediante la relación de Jarzinsky (ver Métodos). El perfil para DPgb-deoxy a lo largo del túnel 1 muestra que existe una pequeña barrera de 2.2 kcal/mol alrededor de los 11 Å del hierro seguida de una región plana hasta los 6 Å y una segunda barrera baja de 2 kcal/mol para llegar al hemo (Figura 5.2-7). Por el contrario, MPgb-deoxy muestra una barrera alta (6 kcal/mol) y ancha alrededor de los 12 Å del hierro para la migración de O₂.

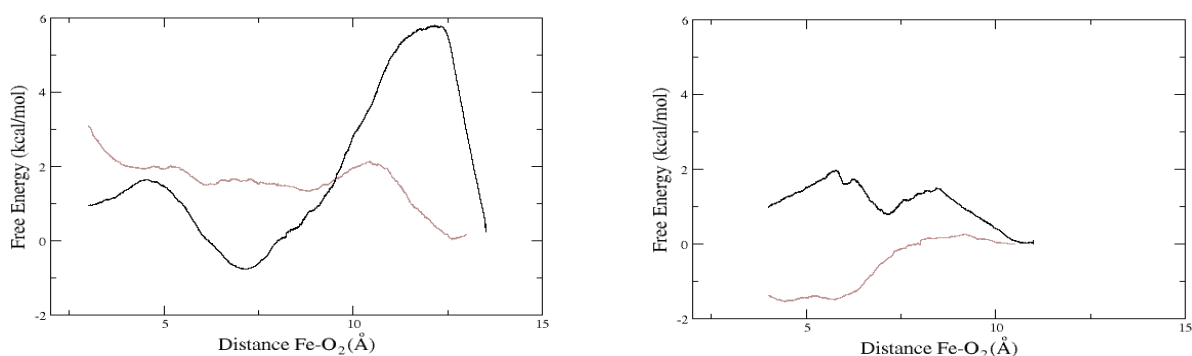


Figura 5.2-7. Perfiles de energía libre para la migración de O₂ a través del túnel 1 (izquierda) y túnel 2 (derecha) en las formas monomérica (negro) y dimérica (gris) de MaPgb

En el caso del túnel 2, tal como se observó en las isosuperficies ILS, la migración de una molécula diatómica no presenta dificultades a temperatura ambiente. Para MPgb-deoxy se aprecia una barrera de sólo 1 kcal/mol mientras que en el caso de DPgb-deoxy no se observa barrera para la migración a través del túnel 2.

Estos resultados están de acuerdo con los previamente presentados: la dimerización facilita la migración a través del túnel 1 lo que aumenta la accesibilidad global de ligandos diatómicos en el dímero en relación al monómero.

5.3 *Discusión acerca de las implicaciones funcionales*

La cristalización de MaPgb (códigos PDB 2VEB y 2VEE), los experimentos por filtración en gel, los estudios cinéticos y los resultados por simulación computacional presentados en esta memoria abren un nuevo capítulo en el ya diverso mundo de las globinas. Pese a que aún se desconoce la función de esta proteína, los resultados obtenidos permitan formular diversas conclusiones.

Los experimentos por filtración en gel, en concordancia con las estructuras de rayos X, muestran que la proteína eluye como dímero¹⁴. En la presente memoria el efecto de la dimerización parece estar relacionado con la apertura del túnel 1 para la migración de ligandos diatómicos. Esto se pone de manifiesto por la mayor proporción de la forma abierta de la Phe(145)G8 en la forma dimérica y está de acuerdo con los análisis por ILS y MSMD/Jarzinsky. Este efecto se ve potenciado con la presencia de ligando unido al hemo. Se trata entonces de un efecto combinado debido a varios factores. Por un

lado, la unión del ligando promueve una mayor flexibilidad en la hélice B, cuya presencia es captada por Phe(93)E11. Por otro lado, la dimerización sostiene la hélice G por su contacto con la H de la otra subunidad permitiendo la aparición del túnel 1, que no se está bien definido en el monómero. Estas hélices delimitan este túnel. Por último, la presencia del ligando unido al hemo promueve la apertura del túnel 1.

Estos resultados sugieren una participación de la proteína en procesos químicos bimoleculares en los que un segundo ligando vería facilitado el acceso al sitio distal una vez que el primero se ha unido al hemo. Este efecto sería aún mayor si el primer ligando fuera una molécula de mayor tamaño (triatómica) dado que el efecto estérico sobre Phe(145)G8 y Phe(93)E11 sería más pronunciado. Por su parte, la dimerización parece ser un mecanismo de regulación para la migración de ligandos.

Estos hallazgos concuerdan con experimentos cinéticos recientes de unión de CO llevados a cabo por el grupo de Cristiano Viapiani (manuscrito en preparación) donde se encuentra que la proteína se halla en 2 estados caracterizados por unión rápida (fast, R) y otra lenta (slow, T), tal como se muestra en el esquema cinético presentado en la Figura 5.3-1. En ese trabajo se describe que existe una mezcla de $\text{MaPgb}^{*\text{R}}\text{CO}$ y $\text{MaPgb}^{*\text{T}}\text{CO}$. Luego de la fotodisociación el CO puede migrar a un docking site secundario $(\text{MaPgb}^{*\text{R}}:\text{CO})_2/(\text{MaPgb}^{*\text{T}}:\text{CO})_2$ desde el primario $(\text{MaPgb}^{*\text{R}}:\text{CO})_1/(\text{MaPgb}^{*\text{T}}:\text{CO})_1$ o escapar a solvente $(\text{MaPgb}^{*\text{R}} + \text{CO})/(\text{MaPgb}^{*\text{T}} + \text{CO})$. El rebinding sucede a través de 2 caminos distintos que involucran $\text{MaPgb}^{*\text{T}}$ o $\text{MaPgb}^{*\text{R}}$. Diferentes constantes de equilibrio conectan $\text{MaPgb}^{*\text{T}}$ con $\text{MaPgb}^{*\text{R}}$ y $\text{MaPgb}^{*\text{R}}\text{CO}$ con $\text{MaPgb}^{*\text{T}}\text{CO}$

En ese trabajo los resultados indican que la unión de CO promueve la forma rápida (fast). Es decir, que la unión del ligando modifica la población relativa de ambos estados enriqueciendo la fast. En este sentido, el estudio computacional descrito en esta tesis coincide con lo hallado por el grupo de Cristiano Viapiani dado que nuestros resultados muestran que la unión del ligando al hemo promueve cambios conformacionales y en la dinámica de la proteína (en particular en las hélices B y E) que resultan en una mayor proporción de forma abierta para el túnel 1. En definitiva, según lo expuesto, la presencia del ligando abre la proteína a la entrada de un segundo ligando, desviando el equilibrio hacia la forma de combinación rápida. Se trataría de una regulación alostérica homotrópica dado que la primera molécula que se liga al hemo actuaría a la vez como sustrato y regulador de la entrada de una segunda molécula.

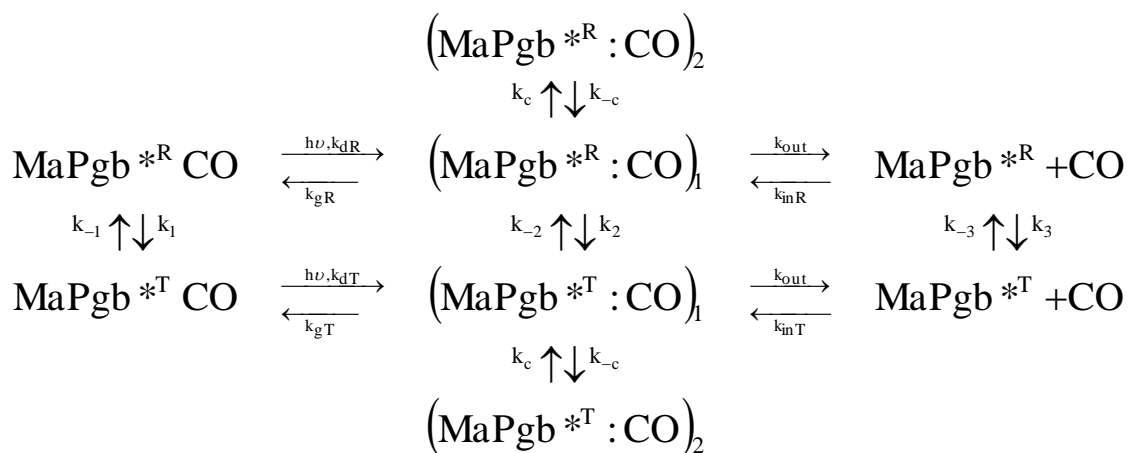


Figura 5.3-1. Esquema cinético para el CO elaborado por el grupo de Cristiano Viapiani

Por otro lado la baja k_{off} para O_2 reportada en los experimentos cinéticos¹⁴ no estaría relacionada con el cierre del túnel 1 dado que el túnel 2 permanece siempre abierto, en todas las formas estudiadas. Es decir, la salida del ligando de la proteína al solvente no estaría limitada por la migración a través de la matriz proteica.

Existen numerosos estudios acerca de los distintos mecanismos de regulación de la afinidad por ligandos que existen entre las globinas²¹. Tal como se comentó, la presencia de residuos donores de puente de hidrógeno en la cavidad distal puede favorecer la estabilización de O_2 ligado, disminuyendo así la k_{off} . En general, cuanto mayor es el número de residuos capaces de establecer un puente de hidrógeno con el ligando, mayor es la afinidad por O_2 y, por lo tanto, menor el valor de k_{off} ²¹. En el caso de MaPgb el único residuo capaz de formar puente de hidrógeno estabilizante con el ligando es Tyr(61)B10. Sin embargo, como se demostró en esta memoria, este residuo alterna un puente de hidrógeno con el ligando y con Leu(86)E4 por lo que no estaría implicado en una estabilización permanente del ligando por efecto distal.

La afinidad puede ser también regulada por una modificación en la reactividad del hemo. Ejemplos de este tipo lo representan leghemoglobina (LHb) y mutantes proximales de Mb. En estos casos cambia la carga de la histidina proximal²². El mecanismo propuesto en este estudio implica un cambio en la reactividad por distorsión de la estructura plana habitual del hemo. Teniendo en cuenta los resultados presentados se puede concluir que el bre positivo (+0,202) observado en el hemo de MaPgb (compresión) contribuye a la baja k_{off} para O_2 reportada por el grupo de Martino Bolognesi. Por otra parte, puesto que el hemo en la mayor parte de las globinas es muy

cercano a la planaridad y ésta es su forma estable aislado en solución, la presencia de un hemo altamente distorsionado como el de MaPgb no puede ser casual, sino que podría reflejar un mecanismo de regulación de la afinidad por O₂.

La existencia de un mecanismo basado en la distorsión del hemo plantea preguntas acerca de los condicionantes evolutivos que han contribuido a incrementar la tensión sobre el grupo hemo. Naturalmente, deben existir residuos en posiciones específicas que empujan al hemo selectivamente para lograr distorsiones fuera del plano y en el plano. Seguramente, ello refleja un efecto cooperativo, que no puede atribuirse a un único residuo concreto, sino al conjunto de residuos que definen la cavidad donde se hospeda el grupo hemo. Estudios de mutagénesis, tanto experimental como computacional, serán sin duda altamente informativos acerca de los residuos involucrados en el mecanismo de compresión del hemo.

5.4 Bibliografía

- (1) Jentzen, W. S., M. C.; Hobbs, J. D.; Song, X. Z.; Ema, T.; Nelson, N. Y.; Medforth, C. J.; Smith, K. M.; Veyrat, M.; Mazzanti, M.; Ramasseul, R.; Marchon, J. C.; Takeuchi, T.; Goddard, W. A. III; Shelnut, J. A. *J Am Chem Soc* **1995**, 117, 11085.
- (2) Shelnut, J. A.; Song, X. Z.; Ma, J. G., Jia, S. L.; Jentzen, W.; Medforth, C. J. . *Chem. Soc. Rev.* **1998**, 27, 31.
- (3) Jentzen, W. S., X.Z.; Shelnut, J. A. *J. Phys. Chem. B* **1997**, 101, 1684.
- (4) Capece, L. M. M. A. C., A.; Doctorovich, F. A.; Estrin, D. A. . *J Am Chem Soc* **2006**, 128, 12455.
- (5) Lu, C. E., T.; Wainwright, L. M.; Poole, R. K.; Yeh, S. R. . *J Biol Chem* **2007**, 282, 13627.
- (6) Peterson, E. S. H., S.; Wang, J.; Miller, L.M.; Vidugiris, G.; Kloek, A. P.; Goldberg, D. E.; Chance, M. R.; Wittenberg, J. B.; Friedman, J. M. *Biochemistry* **1997**, 36, 13110.
- (7) Mukai, M. S., P.Y.; Ouellet, H.; Guertin, M.; Yeh, S.R. *Biochemistry* **2002**, 41, 3897.
- (8) Guallar, V. J., A. A.; Friesner, R. A.; Spiro, T. G. *J. Chem. Inf. Model.* **2006**, 128, 5427.
- (9) Angelis, F. D. J., A. A.; Car, R.; Spiro, T. G. *J. Phys. Chem. B* **2005**, 109, 3065.
- (10) 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, 5611.
- (11) Martí, M. A. B., D.E.; Crespo, A.; Nardini, M.; Bolognesi, M.; Estrin, D. A. *Proteins* **2005**, 62, 641.
- (12) Martí, M. A. C., L.; Bikiel, D.E.; Falcone, B.; Estrin, D.A. *Proteins* **2007**, 68, 480.
- (13) Spiro, T. G. J., A. A. *Curr. Opin. Chem. Biol.* **2001**, 5, 715.
- (14) Nardini M, P. A., Thijs L, Saito JA, Dewilde S, Alam M, Ascenzi P, Coletta M, Ciaccio C, Moens L, Bolognesi M. *EMBO Rep.* **2008**, 9, 157.
- (15) Case, D., Darden TA, Cheatham TE, III, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Pearlman DA, Crowley M, Walker RC, Zhang W, Wang B, Hayik S, Roitberg A, Seabra G, Wong KF, Paesani F, Wu X, Brozell S, Tsui V, Gohlke H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Beroza P, Mathews DH, Schafmeister C, Ross WS, Kollman PA. *AMBER, version 9 University of California: San Francisco, CA, 2006.*
- (16) Hornak, V., Okur, A., Rizzo, R. C., Simmerling, C. *Proteins* **2006**, 65, 712.
- (17) Mukai, M. O., Y.; Ouellet, H.; Guertin, M.; Yeh, S. R. . *Biochemistry* **2004**, 43, 2764.
- (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, 4433.
- (19) Bidon-Chanal, A.; Marti, M. A.; Crespo, A.; Milani, M.; Orozco, M.; Bolognesi, M.; Luque, F. J.; Estrin, D. A. *Proteins* **2006**, 64, 457.

(20) Bidon-Chanal, A.; Marti, M. A.; Estrin, D. A.; Luque, F. J. *J Am Chem Soc* **2007**, 129, 6782.

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

(22) Kundu, S. T., J. T., III; Hargrove, M. S. *Trends Plant Sci.* **2003**, 8, 387.

Capítulo 6

Conclusiones

Metodología Multinivel

- El desarrollo del hamiltoniano semiempírico RM1 hace necesario derivar una nueva versión parametrizada del modelo MST, dado que la transferabilidad de los parámetros AM1 y PM3 puede llevar a errores notables para solventes polares. El modelo parametrizado provee estimaciones precisas de la energía de solvatación para un gran grupo de compuestos neutros no incluidos en el conjunto de entrenamiento.
- El tratamiento MST/RM1 de compuestos iónicos es más delicado, dado que existe una dependencia fuerte entre el factor de escalado utilizado para modular la frontera electrostática entre el solvente y el soluto y el grupo ionizable. El uso de un factor de escalado adaptado al entorno químico provee una estrategia computacional viable para obtener estimaciones precisas de la energía de solvatación para compuestos cargados.
- La metodología Multinivel persigue proporcionar un método eficiente para la exploración conformacional de moléculas tipo fármaco, permitiendo identificar estructuras estables y su población relativa. La estrategia hace uso de una exploración conformacional a nivel RM1 y posterior refinamiento a nivel más robusto de teoría: B3LYP y MP2. De este modo se consigue un buen balance costo/calidad para la estimación precisa de la población relativa de los conformeros relevantes.
- Los resultados obtenidos para la histamina neutra son alentadores considerando la complejidad conformacional y tautomérica que presenta y el delicado balance que existe entre sus diversas formas en solución. Los resultados reflejan este ajustado balance y, aunque existe aún controversia respecto a las preferencias conformacionales de la histamina neutra, coinciden con la mayoría de los estudios previos.

Protoglobina de *Methanosarcina Acetivorans*

- Además de los conocidos mecanismos que regulan la afinidad del hemo por ligandos en las distintas globinas (efecto distal, efecto proximal), hemos estudiado el impacto de las distorsiones de la planaridad de este grupo en la afinidad por O₂. Hemos hallado que éstas tienen un impacto no despreciable y pueden modular la afinidad de una globina por ligandos bimoleculares.
- Las distorsiones fuera del plano siempre resultan en una disminución de la afinidad por O₂. El modo breathing, que implica la compresión-expansión del grupo hemo en el plano, es el único modo a lo largo del cual la distorsión puede aumentar la afinidad por O₂. El breathing positivo del hemo de MaPgb podría contribuir a la alta afinidad por O₂ reportada experimentalmente dado que no presenta efecto distal, habitualmente responsable de la alta afinidad hallada en otras globinas.
- La accesibilidad del ligando a la proteína se logra a través de un sistema de túneles novedoso definido entre las hélices G y B (túnel 1) y B y E (túnel 2). El túnel 2 está siempre abierto, permitiendo la libre migración de ligandos. Esto no sucede en el túnel 1. Phe(145)G8 se halla en 2 conformaciones: abierta y cerrada, que permiten o impiden la migración de ligandos a través del mismo, respectivamente.
- La dimerización facilita la apertura del túnel 1, mientras que el monómero presenta prácticamente siempre cerrado el acceso a ligandos por éste. Sin embargo, la dimerización no es suficiente, siendo necesaria la presencia de ligando unido al hemo para favorecer a su vez la migración a través del túnel 1. La presencia del ligando unido al hemo es detectada por Phe(93)E11, y transmitida a Phe(145)G8 a través del cambio estructural y dinámico de las hélices B y E.
- Los resultados anteriores proporcionan una explicación al comportamiento experimental derivado de los estudios cinéticos realizados por Cristiano Viappiani. De hecho, cabe pensar que las conformaciones de ligación lenta y rápida reflejan la apertura del túnel 1 cuando se ha fijado un primer ligando en el grupo hemo. A su vez, ello sugiere que MaPgb podría participar en un proceso bimolecular, de manera que la entrada de un primer ligando facilitara el acceso de un segundo ligando.

Capítulo 7

Publicaciones

Extension of the MST Continuum Solvation Model to the RM1 Semiempirical Hamiltonian

FLAVIO FORTI,¹ XAVIER BARRIL,^{1,2} F. JAVIER LUQUE,¹ MODESTO OROZCO^{3,4,5}

¹Facultat de Farmàcia, Departament de Físicoquímica and Institut de Biomedicina, Universitat de Barcelona, Av. Diagonal 643, E-08028, Barcelona, Spain

²ICREA, Facultat de Farmàcia, Departament de Físicoquímica and Institut de Biomedicina, Universitat de Barcelona, Av. Diagonal 643, E-08028, Barcelona, Spain

³Facultat de Química, Departament de Bioquímica i Biologia Molecular, Universitat de Barcelona, cl. Martí i Franqués 1, 08028, Barcelona

⁴Institut Nacional de Bioinformàtica and Unitat de Modelització Molecular i Bioinformàtica, Institut de Recerca Biomèdica. Parc Científic de Barcelona, cl. Josep Samitier 1, 08028 Barcelona, Spain

⁵Computational Biology Program, Barcelona Supercomputer Center, Jordi Girona 31, Edifici Torre Girona, Barcelona 08028, Spain

Received 8 June 2007; Accepted 7 July 2007

DOI 10.1002/jcc.20814

Published online 17 August 2007 in Wiley InterScience (www.interscience.wiley.com).

Abstract: The need to simulate large-sized molecules or to deal with large series of compounds is a challenging topic in computational chemistry, which has stimulated the development of accurate semiempirical methods, such as the recently reported Recife Model 1 (RM1; *J Comput Chem* 2006, 27, 1101). Even though RM1 may prove to be of value simply due to the enhanced quantitative accuracy in gas phase, it is unclear how the new parameters optimized for RM1 affect the suitability of this semiempirical Hamiltonian to study chemical processes in condensed phases. To address this question, we report the parametrization of the MST/RM1 continuum model for neutral solutes in water, octanol, chloroform and carbon tetrachloride, and for ions in water. The results are used to discuss the transferability of the solvation parameters implemented in previous MST/AM1 and MST/PM3 models.

© 2007 Wiley Periodicals, Inc. *J Comput Chem* 29: 578–587, 2008

Key words: RM1 Hamiltonian; semiempirical method; solvation; MST model; continuum solvation

Introduction

Many biological processes, including those of pharmacological or toxicological importance, are related to the differential solubility of solutes in aqueous and nonaqueous solvents.^{1,2} Thus, prediction of the solvation-free energy in both polar and apolar environments (and related properties) is crucial for understanding the biological profile of chemical compounds. This challenging goal has stimulated the development of a plethora of methods that combine different formalisms to represent solute and solvent molecules.^{3–8} Among them, quantum mechanical self-consistent reaction field (QM-SCRF) models combine a QM representation of the solute with a classical polarizable continuum treatment of the solvent.^{4,6,8,9} These methods have found widespread acceptance due to (i) the accurate description of the solute at a suitable QM level, which permits to gain insight into the solvent-induced changes in the properties of the solute, (ii) the availability of well-defined formalisms to couple the solute charge distribution and the solvent reaction field, (iii) the

reduced computational cost, and (iv) the availability of QM-SCRF methods in a variety of easy-to-use computer programs.

From a computational point of view, the *ab initio* QM-SCRF treatment of large size solutes is still very demanding, which limits the routine application to studies involving large sets of compounds, such as the screening of libraries of druglike compounds. This justifies the development of a variety of QM-SCRF methods based on the NDDO (MNDO,¹⁰ AM1,¹¹ PM3)¹² framework.^{13–26} Specific parametrizations of semiempirical QM-SCRF

This article contains supplementary material available via the Internet at <http://www.interscience.wiley.com/jpages/0192-8651/suppmat>

Correspondence to: F. J. Luque; e-mail: fjlucque@ub.edu or M. Orozco; e-mail: modesto@mmb.pcb.ub.es

Contract/grant sponsor: Spanish Ministerio de Educación y Ciencia; contract/grant numbers: CTQ2005-08797-CO2-01/BQU and BIO2006-01602

Contract/grant sponsor: Centre de Supercomputació de Catalunya

Table 1. Parameters^a Defined for MST/RM1 Continuum Calculations in Water, Octanol, Chloroform, and Carbon Tetrachloride.

Scaling factor (κ)			
Water	Octanol	Chloroform	Carbon tetrachloride
1.25	1.50	1.60	1.80
Atomic radii (R)			
General	Specific for water ^b		
H ^c	1.20 (0.80)		
C	1.50		
N	1.50	N in –N–	1.40
		N in –NH ₂	1.55
O	1.40	O in –OH	1.45
F	1.35		
P	1.80		
S	1.75	S in –SH	1.85
Cl	1.80		
Br	1.95		

^a R : in Å.^bModified radii for specific chemical groups of neutral compounds in aqueous solution.^cThe radius for a hydrogen atom bound to heteroatom is given in parenthesis.

methods, including refined descriptions of the solute-solvent boundary and of the electrostatic and nonelectrostatic components of the solvation-free energy, have been recently reported, as in the case of the SM5 solvation model developed by Cramer coworkers,^{27–30} or the semiempirical versions of the MST model developed in our group.^{31–33}

Since the introduction of the most elaborate semiempirical methods, AM1 and PM3, several strategies have been undertaken to correct the limited accuracy in reproducing experimental data for certain classes of compounds in the gas phase.³⁴ They are mainly focused on obtaining improved molecular energies, such as the addition of torsional correction terms for amides,³⁵ the adjustment of specific parameters by fitting to high level *ab initio* data,^{36,37} the modification of the core repulsion functions,^{38,39} the inclusion of bond and group equivalents to estimate heats of formation,⁴⁰ or to restore orbital orthogonality.⁴¹ A complete reoptimization of the standard set of parameters has also been attempted, as in the case of the PM5 model.⁴² In this latter spirit, Rocha et al.⁴³ have recently reported the RM1 (Recife Model I), which retains the formalism and the parameters defined for the AM1 Hamiltonian, but relies on a refined parametrization using a reference set of 1736 compounds (containing C, H, N, O, P, S, F, Cl, Br, and I) relevant in organic and biochemical areas. Compared to the AM1, PM3, and PM5 models, the superiority of the RM1 Hamiltonian is supported by the lower errors in predicting properties such as the enthalpy of formation, dipole moments, ionization potentials, and geometrical parameters.

Even though RM1 may prove to be of value simply due to the enhanced quantitative accuracy in gas phase, it is unclear

how the new parameters optimized for RM1 can affect other molecular properties. In particular, our interest is to calibrate the suitability of the RM1 Hamiltonian to study chemical processes in condensed phases. To address this question, we report the parametrization of the MST/RM1 continuum model for neutral solutes in water, octanol, chloroform and carbon tetrachloride, and for ions in water. The results are used to discuss the transferability of the solvation parameters implemented in previous MST/AM1 and MST/PM3 models.

Methods

The MST Model

The free energy of solvation (ΔG_{sol}) is expressed as the sum of electrostatic (ΔG_{ele}), cavitation (ΔG_{cav}), and van der Waals (ΔG_{vw}) contributions [Eq. 1; for a detailed description of the MST model, see ref. 9].

$$\Delta G_{\text{sol}} = \Delta G_{\text{ele}} + \Delta G_{\text{cav}} + \Delta G_{\text{vw}} \quad (1)$$

The *electrostatic* term, which accounts for the work spent in charging up the solute's charge distribution, is determined following the polarizable continuum model (PCM)^{44,45} originally developed by Miertus, Scrocco, and Tomasi (MST). Accordingly, the *pseudo-Schrödinger* equation given by eq. (2) is solved, where H^0 represents the Hamiltonian of the solute, V_{R} stands for the perturbation operator that accounts for the solvent reaction field, and Ψ denotes the wavefunction of the solute.

$$(H^0 + V_{\text{R}})\Psi = E\Psi \quad (2)$$

The free energy of cavitation, that is, the work necessary to generate a cavity large enough to accommodate the solute in the bulk solvent, is computed following Pierotti's scaled particle theory⁴⁶ adapted to molecular-shaped cavities according to the procedure proposed by Claverie.⁴⁷ Accordingly, the free energy of cavitation of atom i ($\Delta G_{\text{C-P},i}$) is determined weighting the cavitation-free energy of the isolated atom in Pierotti's formalism ($\Delta G_{\text{P},i}$) by the ratio between its contribution (S_i) to the solvent-exposed surface and the total solvent-exposed surface (S_{T}) of the molecule [Eq. (3)]

$$\Delta G_{\text{cav}} = \sum_{i=1}^N \Delta G_{\text{C-P},i} = \sum_{i=1}^N \frac{S_i}{S_{\text{T}}} \Delta G_{\text{P},i} \quad (3)$$

Finally, the van der Waals component (arising from dispersion-repulsion interactions between solute and solvent) is determined using a linear relationship with the accessible surface [Eq. (4)]. The atomic surface tensions, ξ_i , are determined by fitting the experimental-free energies of solvation for a training set of molecules.

$$\Delta G_{\text{vw}} = \sum_{i=1}^N \Delta G_{\text{vw},i} = \sum_{i=1}^N \xi_i S_i \quad (4)$$

Table 2. Free Energy of Solvation (kcal/mol) Determined from MST/RMI Calculations for the Training Set of Compounds in Water, Octanol, Chloroform, and Carbon Tetrachloride (Calculated/Experimental).

Compound	Water	Octanol	Chloroform	Carbon tetrachloride
<i>C, H</i>				
Ethane	1.4/1.8	-0.9/-0.6		
Neopentane		-1.7/-1.7		
Cyclohexane	1.9/1.2	-3.3/-3.7		
<i>n</i> -Hexane		-3.1/-3.3		
<i>n</i> -Heptane		-3.6/-4.1		
<i>n</i> -Octane		-4.2/-4.6		
Ethylene	0.1/1.3	-1.8/-0.3		
2-Methylpropene	0.5/1.2	-2.6/-2.0		-2.7/-2.6
1-Butene	0.7/1.4	-2.6/-1.9		-2.9/-2.5
E-2-pentene	0.8/1.3			-3.5/-3.5
Benzene	-1.4/-0.9	-4.3/-3.7	-5.1/-4.6	-4.6/-4.5
Toluene	-1.3/-0.9	-4.8/-4.6	-5.5/-5.5	-5.0/-5.1
<i>m</i> -Xylene	-1.2/-0.8		-6.0/-5.9	-5.4/-5.7
<i>C, H, O</i>				
Water	-6.5/-6.3	-5.6/-4.4		
Methanol	-5.4/-5.1	-4.4/-3.9	-3.5/-3.4	-2.6/-2.2
Ethanol	-5.2/-5.0	-4.7/-4.4	-4.0/-3.9	-3.1/-3.1
<i>n</i> -Propanol	-4.9/-4.8	-5.2/-5.0	-4.6/-4.3	-3.7/-3.7
2-Propanol	-4.2/-4.8		-4.6/-4.3	-3.6/-3.2
<i>n</i> -Butanol	-4.6/-4.7	-5.8/-5.7	-5.3/-5.3	-4.3/-4.2
<i>n</i> -Pentanol	-4.5/-4.5	-6.4/-6.4	-6.0/-5.9	-5.0/-5.0
<i>n</i> -Hexanol	-4.4/-4.4	-6.9/-7.1	-6.6/-6.7	-5.6/-5.7
<i>n</i> -Heptanol	-4.2/-4.2	-7.6/-7.8	-7.3/-7.5	-6.2/-6.5
<i>n</i> -Octanol		-8.1/-8.1		
Phenol	-6.1/-6.6	-7.4/-8.7	-7.0/-7.1	-6.0/-6.1
<i>o</i> -Methylphenol	-5.2/-5.9		-7.2/-7.6	-6.3/-6.8
<i>p</i> -Methylphenol	-5.6/-6.1		-7.3/-7.6	-6.4/-6.1
Dimethylether	-1.3/-1.9	-2.1/-2.1		
Diethylether	-1.6/-1.6		-4.3/-4.3	
Tetrahydrofuran	-2.5/-3.7	-3.9/-3.9		
Anisole	-2.4/-1.0	-5.5/-5.5	-6.5/-6.2	-5.7/-5.5
Acetaldehyde	-3.6/-3.5		-4.0/-3.7	
Benzaldehyde	-4.1/-4.0	-6.5/-6.1	-7.3/-7.1	-6.4/-6.1
<i>p</i> -Hydroxybenzaldehyde	-9.1/-10.5		-9.2/-10.4	-7.8/-8.2
Acetone	-3.8/-3.9	-4.1/-3.2	-4.6/-4.3	-3.4/-3.5
Butanone	-3.4/-3.6		-5.1/-5.4	-4.0/-4.1
Acetophenone	-4.6/-4.6		-7.8/-8.4	-6.8/-6.4
Acetic acid	-7.5/-6.7	-6.1/-6.4	-5.4/-4.5	-4.2/-3.4
Propionic acid	-6.8/-6.5	-6.4/-6.9	-5.8/-5.2	-4.7/-4.3
Butiric acid	-6.5/-6.4	-6.9/-7.6	-6.5/-6.0	-5.3/-5.0
Methylethanoate	-3.1/-3.3	-3.9/-3.5	-4.9/-4.9	-3.8/-3.7
Ethylethanoate	-3.1/-3.1		-5.5/-5.5	-4.4/-4.4
Methylpropanoate	-2.8/-2.9		-5.4/-5.5	-4.3/-4.4
Methylbenzoate	-3.6/-4.3	-6.9/-7.3	-8.1/-7.8	-7.1/-7.2
<i>C, H, N</i>				
Ammonia	-3.5/-4.3			
Methylamine	-4.9/-4.6	-3.9/-3.8	-3.5/-3.2	-2.4/-2.5
Ethylamine	-4.6/-4.5	-4.4/-4.1	-4.1/-4.0	-3.0/-2.8
<i>n</i> -Propylamine	-4.6/-4.4	-5.0/-4.8	-4.8/-4.7	-3.6/-3.6
<i>n</i> -Butylamine	-4.4/-4.3	-5.5/-5.4	-5.4/-5.1	-4.2/-4.2
Dimethylamine	-3.5/-4.3		-3.5/-3.7	-2.5/-2.8
Diethylamine	-4.1/-4.1	-4.6/-4.8	-4.9/-5.2	-3.8/-4.1
Trimethylamine	-2.2/-3.2	-2.5/-3.6	-3.9/-4.0	-2.6/-3.1
Piperidine	-4.7/-5.1	-5.5/-5.3		
Aniline	-5.8/-5.0		-7.0/-6.8	-5.9/-5.7
Acetonitrile	-4.0/-3.9	-3.4/-3.2	-4.8/-4.4	

(continued)

Table 2. (Continued)

Compound	Water	Octanol	Chloroform	Carbon tetrachloride
Propanenitrile	-3.9/-3.9	-3.9/-3.7		
Butanenitrile	-3.7/-3.6	-4.4/-4.3		
Benzonitrile	-4.7/-4.1	-6.4/-6.1	-8.0/-7.2	-6.8/-6.3
Pyridine	-4.3/-4.7	-4.7/-5.3	-6.2/-6.7	-4.9/-5.0
2-Methylpyridine	-4.2/-4.6		-6.6/-7.0	
4-Methylpyridine	-4.4/-4.9		-6.7/-7.5	
<i>C, H, N, O</i>				
Acetamide	-10.1/-9.7	-7.4/-8.0	-6.5/-7.1	
Nitroethane	-4.3/-3.7	-4.4/-3.9		
Nitropropane	-3.6/-3.3			-4.9/-4.5
Nitrobenzene	-4.2/-4.1	-6.5/-6.6	-8.5/-7.8	-7.0/-6.9
<i>m</i> -Nitrophenol			-10.1/-10.5	-8.3/-8.8
<i>p</i> -Nitrophenol			-10.5/-11.0	-8.5/-9.4
<i>C, H, S</i>				
Hydrogen sulfide	-1.2/-0.7			
Metanethiol	-1.2/-1.2			
1-Propanethiol		-3.8/-3.5		
Thiophenol	-2.0/-2.6	-5.7/-6.0	-7.6/-7.6	
Dimethylthioether	-1.4/-1.5	-3.3/-4.2		
Diethylthioether	-2.0/-1.4	-4.6/-4.1	-5.5/-6.4	
Thioanisole	-2.2/-2.7	-6.5/-6.5	-7.5/-6.0	-5.7/-5.7
Thiophene		-4.3/-3.9	-5.2/-5.8	
<i>C, H, Halogens</i>				
Fluoromethane	-0.4/-0.2			
Tetrafluoromethane	4.3/3.2	1.7/1.5		
1,1-Difluoroethane	-0.4/-0.1	-1.6/-1.1		
2,2,2-Trifluoroethanol	-6.2/-4.3	-5.0/-4.8	-3.2/-3.0	
Fluorobenzene		-4.1/-3.9	-4.3/-4.3	-3.6/-3.6
Chloroform		-3.9/-3.8		
Chloroethane		-3.5/-2.6		
Chlorobenzene	-1.3/-1.1		-5.8/-5.5	-5.4/-5.2
<i>p</i> -Dichlorobenzene	-0.6/-1.0	-6.0/-5.7	-6.4/-6.3	-6.2/-6.3
Fluorotrichloromethane		-2.2/-2.6	-2.7/-2.6	
Difluorodichloromethane	3.8/1.7	-0.9/-1.3	-1.2/-1.6	
Dichloromethane	-2.0/-1.4			
Bromomethane	-1.5/-0.8	-2.9/-2.4		
Bromoethane	-1.8/-0.7	-3.6/-2.9		
Bromobenzene	-1.7/-1.5	-5.7/-5.5		-6.3/-6.3
<i>p</i> -Dibromobenzene	-1.3/-2.4	-6.8/-7.5		
1-Bromo-2-chloroethane	-1.6/-2.0			
<i>p</i> -Bromophenol	-6.9/-7.1		-8.6/-8.6	-7.9/-7.9
<i>C, H, P</i>				
Phosphine	0.6/0.6			

Computational Details

Following previous studies,⁴⁸ a dual cavity strategy is used. Accordingly, nonelectrostatic components of the solvation-free energy are determined using a van der Waals surface built up from suitable atomic van der Waals radii, while the electrostatic term is computed using the solvent-exposed surface defined upon an appropriate scaling of those atomic radii. This scaling factor (κ) was generally adjusted from the comparison of the electrostatic response determined from QM SCRF MST computations and the electrostatic component of the solvation-free

energy obtained from classical discrete simulations (Molecular Dynamics or Monte Carlo Free Energy Calculations). The solvent-dependent scaling factor and the atomic radii adopted in preceding MST parametrizations³³ have been retained here (see Table 1). The GEPOL procedure is used to determine the solute/solvent interface for both electrostatic and nonelectrostatic components.⁴⁹ A relevant aspect to be mentioned is the procedure used to compute the solvent's reaction field, which, in turn, depends on the electrostatic potential. In contrast to *ab initio* methods, where the electrostatic potential is uniquely defined, several procedures have been proposed in the context of semi-

Table 3. Statistical Parameters of the Fitting Between Experimental and Theoretical Solvation-Free Energies Derived from MST/RM1 Computations.

Parameter ^a	Water	Octanol	Chloroform	Carbon tetrachloride
Training set				
<i>n</i>	81	66	57	50
msd	0.0	0.1	0.0	0.0
rmsd	0.6	0.5	0.5	0.3
<i>c</i>	0.99	1.00	1.00	1.00
<i>r</i>	0.99	0.99	1.00	1.00
<i>F</i>	3,194	5,853	10,247	15,442
Validation set				
<i>n</i>	32	24	16	12
msd	0.1	0.0	0.0	-0.1
rmsd	0.7	0.7	0.8	0.6
<i>c</i>	0.97	1.01	0.99	1.02
<i>r</i>	0.99	0.99	0.99	0.99
<i>F</i>	2068	1572	798	851

^aNumber of compounds (*n*); mean-signed deviation (msd; kcal/mol); root-mean square deviation (rmsd; kcal/mol); scaling coefficient (*c*), Pearson's correlation coefficient (*r*), and *F* value.

empirical methods.^{50–54} Here, the three-centered integrals needed to compute the electrostatic potential are solved using the NDDO-based approach.^{51,53} Moreover, those integrals are evaluated by canceling out the empirical parameters ρ_{A,I_1} and ρ_{B,I_2} that modify the electrostatic multipole interaction in the framework of the Dewar-Sabelli-Klopmann approximation, and the electrostatic interaction between the atomic core and the charges that simulate the solvent reaction field is determined using a Coulombic expression (see refs. 20 and 23 for more details).

A total of 81, 66, 57, and 50 compounds containing prototypical (bio)organic functional groups have been considered to carry out the parametrization of the MST/RM1 model for water, octanol, chloroform, and carbon tetrachloride, respectively. The statistical significance of the model has been further assessed by comparison of calculated and experimental solvation-free energies for a test set of compounds. For ions, we have chosen 97 ionic compounds taken from the data recently compiled by Kelly et al.⁵⁵ and corrected by a factor of 1.9 kcal/mol, which accounts for the difference between the theoretical value determined by Zhan and Dixon⁵⁶ for the hydration-free energy of the proton (used in ref. 55 to obtain the experimental hydration-free energies of ions) and the most accurate experimental value obtained by Tissandier et al.⁵⁷ RM1 calculations were performed using a locally modified version of MOPAC 6.0.

Results and Discussion

Neutral Compounds

The parametrization of the MST model coupled to the RM1 Hamiltonian was carried out trying to reduce to the minimum the number of adjustable parameters to obtain more transferable models and to avoid “overtraining” artefacts. Thus, we decided to limit ourselves to fit only atomic surface tensions [ξ ; Eq. (5)]

by comparison between computed and experimental-free energies of solvation.

$$\Delta G_{vW} = \Delta G_{\text{sol}}^{\text{exp}t} - \Delta G_{\text{ele}} - \Delta G_{\text{cav}} \quad (5)$$

Table 2 shows the solvation-free energies in the four solvents for the training set of compounds, and Table 3 reports the statistical parameters for the comparison of the experimental and theoretical solvation-free energies (extended data are provided as Supporting information). In all cases, the mean signed deviations are close to zero, and the scaling coefficient of the regression equation between experimental and theoretical values is close to unity, thus indicating the lack of systematic errors for a large, diverse series of neutral compounds. The root-mean square deviation (rmsd) between theoretical and experimental solvation-free energies amount to 0.6, 0.5, 0.5, and 0.3 kcal/mol for water, octanol, chloroform, and carbon tetrachloride, which are very close to the uncertainties obtained for the MST/AM1 and MST/PM3 models and for the *ab initio* versions of the method.

Inspection of the different components of the solvation-free energy (see Supporting information) reveals that the electrostatic component, on average, accounts for around 132%, 39%, 24%, and 11% of the total solvation-free energy in water, octanol, chloroform, and carbon tetrachloride. Similarly, the change in the dipole moment induced upon solvation amounts to around 40%, 17%, 13%, and 6%, which is also in agreement with the solvent-induced changes in the dipole moment predicted by previous version of the MST algorithm and by discrete QM/MM calculations.^{58,59}

To test the parametrization of the MST/RM1 method, we have computed the solvation-free energies for an additional validation set of 84 solvation-free energies (water: 32; octanol: 24; chloroform: 16; carbon tetrachloride: 12; see Table 4). None of these molecules were considered in the parametrization set. The reliability of the MST/RM1 model out of the training set can be verified by the excellent agreement found between experimental and theoretical estimates (see Table 3 and Fig. 1).

The atomic surface tensions optimized for the MST/RM1 model in water, octanol, chloroform, and carbon tetrachloride (see Table 5) are generally close to the values optimized for MST/AM1 and MST/PM3 models, which were not unexpected considering the similarity between the three methods. The largest differences are found for C and N atom types, which can be attributed at least in part to the fact that both MST/AM1 and MST/PM3 parametrizations required the inclusion of a correction term for nitrile groups, while this correction term was unnecessary in the MST/RM1 model. On the other hand, taken together the data for the four solvents, the rmsd values between the ΔG_{ele} values determined for RM1, AM1, and PM3 are close to 0.5 kcal/mol (see Table 6). Nevertheless, the rmsd varies from 0.1 kcal/mol in carbon tetrachloride to around 0.9 kcal/mol in water, as expected from the reduction in the scaling factor κ that modulates the solute-solvent electrostatic boundary, which, in turn, leads to larger differences in the ΔG_{ele} term between the semiempirical formalisms.

Overall, the free energies of solvation (data not shown) obtained by assuming the transferability of the MST/AM1 or

Table 4. Free Energy of Solvation (kcal/mol) Determined from MST/RM1 Calculations for the Validation Set of Compounds in Water, Octanol, Chloroform, and Carbon Tetrachloride (Calculated/Experimental).

Compound	Water	Octanol	Chloroform	Carbon tetrachloride
<i>C, H</i>				
Propane	1.9/2.0	-1.4/-1.3		
Hexane	2.7/2.5			
Cyclopentene	-0.3/0.6			
Ethylbenzene	-1.1/-0.8	-5.3/-5.1	-6.1/-5.8	-5.5/-5.7
<i>p</i> -Xylene	-1.2/-0.8	-5.3/-5.2		
Naphtalene	-2.5/-2.4	-7.5/-7.0	-7.7/-7.9	-7.0/-7.6
Cyclohexane			-4.0/-4.5	
<i>C, H, O</i>				
2-Methyl-2-propanol	-4.0/-4.5	-5.0/-4.8	-4.9/-4.5	-3.8/-3.4
Dioxane	-2.7/-1.8	-4.1/-4.9	-4.9/-6.2	-4.0/-5.0
Methylpropylether	-1.3/-1.7	-3.3/-3.6		
Propanal	-3.7/-3.4	-4.1/-4.1		
Cyclopentanone	-4.2/-4.7	-5.4/-5.0		-4.6/-5.3
Methylmethanoate	-3.0/-2.8	-3.3/-2.8		
2-Propen-1-ol	-5.9/-6.8	-6.0/-5.3	-5.0/-4.3	
<i>p</i> -Cresol		-7.8/-8.8	-7.3/-7.6	
Butiric acid		-6.9/-7.6	-6.5/-6.0	-5.3/-4.8
Acetaldehyde			-4.0/-3.7	
Propionic acid			-5.8/-5.4	
Benzaldehyde				-6.4/-6.1
Butanone				-4.0/-4.1
Methylethanoate				-3.8/-3.8
<i>C, H, N</i>				
Pirrolidine	-5.1/-5.5			
Piperazine	-7.7/-7.4	-7.1/-5.8		
<i>p</i> -Methylaniline	-6.0/-5.6			
Methylpirazine	-4.8/-5.6			
Methyladenine	-12.9/-13.6			
Aminoaniline	-10.9/-9.9			
Aniline		-7.2/-6.7		
2-Methylpyridine		-5.0/-6.1		
4-Methylpyridine		-5.2/-6.6		
Ammonia			-3.1/-2.4	-2.2/-1.1
Piperidine			-5.3/-6.4	
Trimethylamine				-2.6/-3.1
<i>C, H, N, O</i>				
Benzamide	-11.8/-10.9			
Urea	-15.5/-13.8			
Nitromethane	-4.5/-4.0			
Morpholine	-5.5/-7.2		-5.6/-6.7	
Methylthymine	-9.8/-10.4			
Nitropropane		-4.7/-4.4		
<i>C, H, S</i>				
Ethanethiol	-1.5/-1.3			
Dimethylsulfide	-1.4/-1.8			
Diethylthioether		-4.6/-4.1	-5.5/-6.4	
Thioanisol			-7.5/-7.5	-5.7/-5.7
<i>C, H, Halogens</i>				
Fluorobenzene	-0.8/-0.8			
Dichloromethane	-1.9/-1.4			
<i>o</i> -Chlorotoluene	-0.8/-1.2			
Dibromomethane	-1.7/-0.7			
Chlorofluoromethane	-2.3/-0.8			
Tetrafluorobenzene		1.7/1.5		
2,2,2-Trifluoroethanol				-5.0/-4.8
Cholorobenzene		-5.3/-5.0		
<i>p</i> -Dicholorobenzene		-6.0/-5.7		
<i>p</i> -Bromophenol		-9.2/-10.6		
Bromobenzene			-6.7/-6.1	
<i>C, H, P</i>				
Triethylphosphate	-7.9/-7.8			

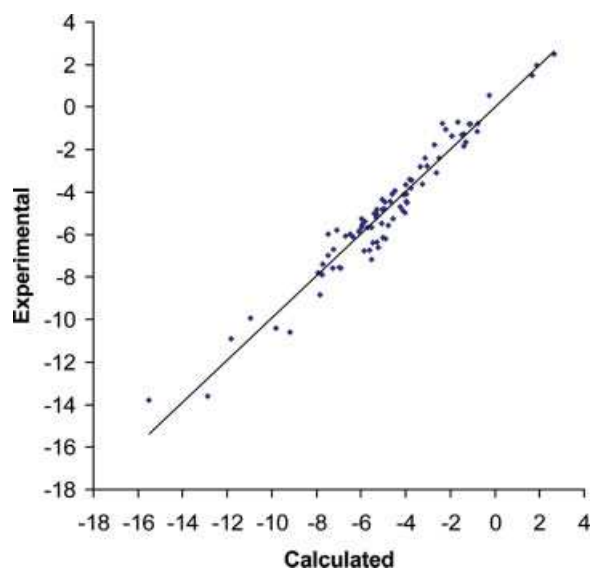


Figure 1. Representation of the experimental and theoretical solvation-free energies determined for the validation set of compounds in water, octanol, chloroform and carbon tetrachloride.

MST/PM3 $\Delta G_{\text{v,w}}$ values yields errors of 0.5 and 0.3 kcal/mol in chloroform and carbon tetrachloride, which compare with the values obtained for the optimized MST/RM1 model (see Table 2), but to uncertainties which are ~ 0.5 kcal/mol larger than those obtained for the optimized MST/RM1 model for the most polar solvents. Even though this is not a dramatic difference, it represents a non-negligible worsening in the chemical accuracy, confirming then the need to perform a specific parametrization for the RM1 Hamiltonian.

Table 5. Optimized Atomic Surface Tensions (ξ) for the MST/RM1 Model in Water, Octanol, Chloroform, and Carbon Tetrachloride.^a

ξ	Water	Octanol	Chloroform	Carbon tetrachloride
C	-0.169 (-0.217/-0.288)	-0.248 (-0.280/-0.280)	-0.161 (-0.239/-0.242)	-0.180 (-0.220/-0.220)
CH	-0.133 (-0.124/-0.125)	-0.197 (-0.192/-0.193)	-0.158 (-0.151/-0.154)	-0.153 (-0.150/-0.151)
CH ₂	-0.130 (-0.130/-0.127)	-0.194 (-0.194/-0.192)	-0.148 (-0.150/-0.150)	-0.143 (-0.144/-0.144)
CH ₃	-0.120 (-0.114/-0.120)	-0.182 (-0.178/-0.178)	-0.143 (-0.140/-0.143)	-0.135 (-0.134/-0.135)
N	-0.073 (-0.128/-0.101)	-0.110 (-0.245/-0.204)	-0.182 (-0.256/-0.206)	-0.151 (-0.206/-0.190)
NH	-0.130 (-0.115/-0.145)	-0.227 (-0.325/-0.311)	-0.179 (-0.194/-0.185)	-0.174 (-0.174/-0.175)
O	-0.060 (-0.013/-0.031)	-0.150 (-0.138/-0.105)	-0.154 (-0.134/-0.116)	-0.152 (-0.139/-0.133)
OH	-0.174 (-0.152/-0.159)	-0.258 (-0.273/-0.265)	-0.186 (-0.179/-0.173)	-0.189 (-0.183/-0.182)
F	-0.048 (-0.047/-0.049)	-0.124 (-0.124/-0.118)	-0.074 (-0.077/-0.076)	-0.051 (-0.052/-0.048)
P	-0.074	-	-	-
PH	-0.074	-	-	-
S	-0.104 (-0.079/-0.058)	-0.191 (-0.205/-0.173)	-0.169 (-0.180/-0.156)	-0.120 (-0.119/-0.109)
SH	-0.093 (-0.064/-0.053)	-0.176 (-0.176/-0.157)	-0.188 (-0.187/-0.172)	-
Cl	-0.094 (-0.118/-0.110)	-0.163 (-0.167/-0.175)	-0.129 (-0.128/-0.128)	-0.126 (-0.125/-0.125)
Br	-0.102 (-0.115/-0.108)	-0.163 (-0.175/-0.173)	-0.141 (-0.135/-0.135)	-0.139 (-0.134/-0.135)

^aThe atomic surface tensions optimized for MST/AM1 and MST/PM3 methods are given in parentheses (AM1/PM3).

Ionic Compounds

According to our previous *ab initio* studies, the solute cavity should be modified by reducing the scaling factor κ by around 10% (i.e., $\kappa = 1.12$) for those atoms (the heavy atom and the hydrogen atoms attached to it) that bear the formal charge.^{60,61} Unfortunately, this strategy was found to be inadequate at the semiempirical level. As an example, whereas the hydration-free energy of CH_3OH_2^+ was predicted to be -92.8 kcal/mol, which compares with the experimental value of -92.9 kcal/mol, the calculated and experimental hydration-free energies of protonated CH_3NH_2 (-81.2 and -75.9 kcal/mol), $\text{O}(\text{CH}_3)_2$ (-71.4 and -79.9 kcal/mol), and CH_3COCH_3 (-66.6 and -76.9 kcal/mol) differ by 5.3, 8.5, and 10.3 kcal/mol. Similarly, the hydration-free energies of CH_3O^- and CH_3S^- , which are predicted to be -90.6 and -92.3 kcal/mol, differ from the experimental values (-95.1 and -74.1 kcal/mol, respectively) by -4.5 and 18.2 kcal/mol, thus indicating a moderate underestimation for CH_3O^- but a much larger overestimation for CH_3S^- .

These findings strongly argue against the use of a universal reduction in the scaling factor that modulates the electrostatic cavity associated with the atoms bearing the formal charge. Rather, the magnitude of the scaling factor appears to be very sensitive to the nature of the heavy atom and to the chemical environment. This behavior is not completely unexpected, as previous studies have pointed out that AM1 and PM3 predict intrinsic proton affinities and basicities with a nonsystematic scatter from the experimental values,⁶²⁻⁶⁵ it being suggested that the predictive capability of those semiempirical methods might be improved using a chemical environment-based scaling.⁶⁵

On the basis of the preceding considerations, we performed a systematic exploration to determine the optimum scaling factor for a series of monovalent cations and anions (see Methods). Table 7 reports the calculated and experimental hydration-free energies obtained for a series of 47 cations, which can be grouped into

Table 6. Statistical Parameters of the Comparison Between the Electrostatic Component of the Solvation-Free Energy in Water, Octanol, Chloroform, and Carbon Tetrachloride Derived from MST/RM1, MST/AM1, and MST/PM3 Computations.

Parameter ^a	msd	rmsd	c	r	F
All values					
RM1 vs. AM1	0.1	0.5	0.92	0.99	8736
RM1 vs. PM3	0.2	0.6	0.93	0.98	5062
AM1 vs. PM3	0.1	0.8	0.91	0.97	5319
Water					
RM1 vs. AM1	0.5	0.8	0.91	0.97	1215
RM1 vs. PM3	0.3	0.9	0.94	0.94	547
AM1 vs. PM3	-0.2	0.8	1.03	0.95	645
Octanol					
RM1 vs. AM1	-0.1	0.3	1.06	0.97	919
RM1 vs. PM3	0.2	0.5	0.89	0.94	489
AM1 vs. PM3	0.3	0.5	0.84	0.97	830
Chloroform					
RM1 vs. AM1	0.0	0.2	1.02	0.98	1107
RM1 vs. PM3	0.1	0.3	0.90	0.94	455
AM1 vs. PM3	0.1	0.3	0.88	0.96	571
Carbon tetrachloride					
RM1 vs. AM1	0.0	0.1	0.96	0.98	1240
RM1 vs. PM3	0.0	0.1	0.89	0.95	452
AM1 vs. PM3	0.0	0.1	0.92	0.96	621

^aMean-signed deviation (msd; kcal/mol); root-mean square deviation (rmsd; kcal/mol); scaling coefficient (*c*); Pearson's correlation coefficient (*r*); and F value.

three categories (sets C1-C3), depending on the value of the optimized scaling factor. The data in Table 7 indicate that there is an inverse relation between the magnitude of the scaling factor and the number of hydrogen atoms attached to the protonable atom. Thus, for ethers and thioethers, carbonyl and nitrile compounds, secondary aromatic and tertiary aliphatic amines or heterocycles such as pyridine (set C1), the optimum agreement between experimental and calculated hydration-free energies is obtained using a scaling factor of 0.99. For alcohols, secondary aliphatic amines or compounds such as pyrrole (set C2), the scaling factor increases to 1.15, and finally for primary aliphatic amines or aromatic unsubstituted amines (set C3) it amounts to 1.24. With regard to the anions (see Table 8), the optimized scaling factors also show a large dependence on the nature of the atom type, ranging from 1.09 for alcoxides and phenoxides to 1.24 for carboxylate anions, and even to an enlargement of the cavity ($\kappa = 1.34$) for methyl-sulfate and thiolate anions. As a simple rule-of-thumb, it can be stated that the magnitude of the scaling factor must be enlarged as the positive charge induced upon protonation can be partitioned into a larger number of hydrogen atoms, or as the negative charge is delocalized onto a larger number of atoms.

The use of a chemical environment-based scaling treatment of the solute-solvent boundary of the ionizable group appears to be promising to estimate the hydration-free energies of charged species at the semiempirical level. The implementation of this computational strategy is, nevertheless, challenged by the intrinsic uncertainty associated with the experimental hydration-free energies (around 3 kcal/mol⁵⁵), and by the limited availability of

Table 7. Calculated and Experimental Hydration-Free Energies (kcal/mol) of Monovalent Cations Obtained by Adopting a Chemical Environment Type-Dependent Scaling Factor to Define the Electrostatic Boundary Between Solute and Solvent.

	Calculated	Experimental
Set C1 ($\kappa = 0.99$)		
Dimethylether	-78	-80
Diethylether	-64	-72
Acetone	-72	-77
Acetophenone	-59	-65
Ethyletanoate	-68	-69
Acetamide	-80	-74
Formamide	-88	-83
Trimethylamine	-61	-61
Triethylamine	-49	-55
Acetonitrile	-75	-75
<i>N,N</i> -diethylaniline	-51	-54
<i>N,N</i> -dimethylaniline	-61	-57
Pyridine	-63	-61
<i>p</i> -Methyl- <i>N,N</i> -dimethylaniline	-59	-56
Quinoline	-59	-56
Dimethylthioether	-62	-65
Set C2 ($\kappa = 1.15$)		
Methanol	-91	-93
Ethanol	-83	-88
Azacycloheptane	-58	-63
Azetidine	-66	-68
Aziridine	-68	-71
Dimethylamine	-68	-69
Diethylamine	-59	-64
Morpholine	-75	-70
<i>N</i> -ethylaniline	-62	-62
<i>N</i> -methylaniline	-66	-63
Piperazine	-68	-66
Piperidine	-66	-64
Pyrrol	-68	-62
Pyrrolidine	-64	-66
Set C3 ($\kappa = 1.24$)		
1-Aminonaphthalene	-67	-68
<i>t</i> -Butylamine	-61	-67
Methylamine	-75	-76
Propylamine	-67	-72
Butylamine	-74	-72
Alylamine	-75	-72
Cyclohexanamine	-61	-69
Isopropylamine	-65	-70
<i>m</i> -Aminoaniline	-72	-66
<i>m</i> -Chloroaniline	-77	-75
<i>m</i> -Methylaniline	-70	-70
<i>o</i> -Methylaniline	-68	-70
<i>p</i> -Chloroaniline	-78	-74
<i>p</i> -Methylaniline	-68	-70
<i>p</i> -Methoxyaniline	-75	-71
<i>p</i> -Nitroaniline	-88	-76
Aniline	-71	-72
msd	-0.2	
rmsd	4.3	
<i>c</i>	1.00	
<i>r</i>	1.00	
<i>F</i>	11876	

Table 8. Calculated and Experimental Hydration-Free Energies (kcal/mol) of Monovalent Anions Obtained by Adopting a Chemical-Environment Type-Dependent Scaling Factor to Define the Electrostatic Boundary Between Solute and Solvent.

	Calculated	Experimental
Set A1 ($\kappa = 1.09$)		
Methanol	-94	-95
Ethanol	-90	-91
1,2-Ethanediol	-91	-85
2,2,2-Trifluoroethanol	-64	-78
Methoxyethanol	-90	-89
Propanol	-90	-88
2-Propanol	-89	-86
2-Propen-1-ol	-87	-86
<i>t</i> -Butanol	-87	-82
Hexafluoropropan-2-ol	-53	-65
2-Butanol	-86	-84
Benzilalcohol	-83	-85
Phenol	-72	-72
<i>m</i> -Hydroxyphenol	-74	-74
<i>m</i> -Methylphenol	-73	-71
<i>m</i> -Nitrophenol	-60	-62
<i>o</i> -Chlorophenol	-68	-66
<i>o</i> -Methylphenol	-70	-70
<i>o</i> -Nitrophenol	-62	-60
<i>p</i> -Chlorophenol	-65	-66
<i>p</i> -Hydroxyphenol	-75	-78
<i>p</i> -Methylphenol	-72	-72
<i>p</i> -Nitrophenol	-59	-58
Diphenylamine	-66	-54
Set A2 ($\kappa = 1.18$)		
Aniline	-75	-63
<i>p</i> -Nitroaniline	-58	-57
Acetamide	-81	-80
Propanenitrile	-68	-66
3-Pentanone	-69	-74
Acetaldehyde	-74	-76
Dimethylsulfoxide	-66	-68
Acetone	-74	-76
Nitromethane	-70	-76
Formaldehyde	-74	-76
Acetonitrile	-71	-72
Ethylethanoate	-70	-66
Set A3 ($\kappa = 1.24$)		
Formic acid	-77	-76
Acetic acid	-75	-78
Acrylic acid	-75	-74
Benzoic acid	-75	-71
Chloroacetic acid	-69	-70
Dichloroacetic acid	-64	-62
Hexanoic acid	-76	-74
Piruvic acid	-69	-68
Propionic acid	-74	-76
Trifluoroacetic acid	-55	-59
Set A4 ($\kappa = 1.34$)		
Methanesulfonic acid	-66	-72
Ethaneethiol	-73	-72
Propaneethiol	-73	-70
Thiophenol	-66	-63
Methaneethiol	-73	-74

(continued)

Table 8. (Continued)

	Calculated	Experimental
msd	0.0	
rmsd	4.4	
<i>c</i>	1.00	
<i>r</i>	1.00	
<i>F</i>	14296	

reliable values for charged compounds comprising a variety of prototypical functional groups. Under these circumstances, the calibration of the chemical environment-based scaling factors against hydration-free energy determined from high level *ab initio* computations might be a valuable alternative.

Conclusions

The development of a new set of parameters within the framework of the RM1 semiempirical Hamiltonian makes it necessary to derive a new parametrized version of the MST method, since transferability of AM1 or PM3 parameters can lead to non-negligible errors for polar solvents. The parametrized model provides accurate estimates of the solvation-free energy for a large series of neutral compounds not included in the training set. The MST/RM1 treatment of ionic compounds is, nevertheless, more delicate, as there is a large dependence between the scaling factor used to modulate the electrostatic boundary between solute and solvent and the nature of the ionizable group. The use of a chemical environment-adapted scaling factor provides a computationally feasible strategy to obtain accurate estimates of the hydration-free energy for charged compounds. The availability of semiempirical methods able to predict accurately the molecular energies in the gas phase and the solvation-free energies from QM-SCRF MST computations should contribute to enhance their widespread usage in a broad range of areas, such as the screening of large sets of bioorganic compounds in drug-design studies.

Supporting Information

Components of the solvation-free energy in water, octanol, chloroform, and carbon tetrachloride for the training and validation sets of neutral compounds, and for monovalent ionic compounds.

Acknowledgments

FF is fellowship from the Spanish MEC.

References

- 3D QSAR. In Drug Design: Theory, Methods and Applications; Kubinyi, H., Ed.; Kluwer/ESCOM: The Netherlands, 2000.

2. Gundertofte, E. K.; Jorgensen, F. S. *Pharmacokinetics Molecular Modeling and Prediction of Bioavailability*; Kluwer Academic: New York, 2000.
3. Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*; Wiley: New York, 1991.
4. Rivail, J. L.; Rinaldi, D. In *Computational Chemistry, Reviews of Current Trends*; Leszczynski, J., Ed.; World Scientific: Singapore, 1995; p. 139.
5. van Gunsteren, W. F.; Weiner, P. K.; Wilkinson, A. J., Eds. *Computer Simulation in Biomolecular Systems. Theoretical and Experimental Applications*, Vol. 3; Kluwer: Dordrecht, 1997.
6. Cramer, C. J.; Truhlar, D. G. *Chem Rev* 1999, 99, 2161.
7. Orozco, M.; Luque, F. J. *Chem Rev* 2000, 100, 4187.
8. Tomasi, J.; Mennucci, B.; Cammi, R. *Chem Rev* 2005, 105, 2999.
9. Luque, F. J.; Curutchet, C.; Muñoz-Muriedas, J.; Bidon-Chanal, A.; Soteras, I.; Morreale, A.; Gelpí, J. L.; Orozco, M. *Phys Chem Chem Phys* 2003, 5, 3827.
10. Dewar, M. J. S.; Thiel, W. *J Am Chem Soc* 1977, 99, 4899.
11. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J Am Chem Soc* 1985, 107, 3902.
12. Stewart, J. J. P. *J Comput Chem* 1989, 10, 209.
13. Kozaki, T.; Morihashi, K.; Kikuchi, O. *J Mol Struct (Theochem)* 1988, 168, 265.
14. Karelson, M. M.; Katritzky, A. R.; Szafran, M.; Zerner, M. C. *J Org Chem* 1989, 54, 6030.
15. Szafran, M.; Karelson, M. M.; Katritzky, A. R.; Koput, J.; Zerner, M. C. *J Comput Chem* 1993, 14, 371.
16. Ford, G. P.; Wang, B. *J Comput Chem* 1992, 13, 229.
17. Wang, B.; Ford, G. P. *J Chem Phys* 1992, 97, 4162.
18. Cramer, C. J.; Truhlar, D. G. *J Am Chem Soc* 1991, 113, 8305.
19. Cramer, C. J.; Truhlar, D. G. *Science* 1992, 256, 213.
20. Chudinov, G. E.; Napolov, D. V.; Basilevsky, M. U. *Chem Phys* 1992, 160, 41.
21. Rauhut, G.; Clark, T.; Steinke, T. *J Am Chem Soc* 1993, 115, 9174.
22. Negre, M.; Orozco, M.; Luque, F. J. *Chem Phys Lett* 1992, 196, 27.
23. Luque, F. J.; Negre, M. J.; Orozco, M. *J Phys Chem* 1993, 97, 4386.
24. Klamt, A.; Schüürmann, G. *J Chem Soc Perkin Trans 2* 1993, 799.
25. Furuki, T.; Umeda, A.; Sakurai, M.; Inoue, Y.; Chujo, R. *J Comput Chem* 1994, 15, 90.
26. Furuki, T.; Sakurai, M.; Inoue, Y. *J Comput Chem* 1995, 16, 378.
27. Giesen, D. J.; Gu, M. Z.; Cramer, C. J.; Truhlar, D. G. *J Org Chem* 1996, 61, 8720.
28. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J Phys Chem* 1996, 100, 19824.
29. Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J Phys Chem* 1996, 100, 16385.
30. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J Phys Chem B* 1998, 102, 3257.
31. Luque, F. J.; Bachs, M.; Orozco, M. *J Comput Chem* 1994, 15, 847.
32. Orozco, M.; Bachs, M.; Luque, F. J. *J Comput Chem* 1995, 16, 563.
33. Curutchet, C.; Salichs, A.; Barril, X.; Orozco, M.; Luque, F. J. *J Comput Chem* 2003, 24, 32.
34. Ahn, N. T.; Frison, G.; Solladie-Cavallo, A.; Patrick, M. *Tetrahedron* 1998, 54, 12841.
35. Stewart, J. J. P. *J Comput-Aided Mol Design* 1990, 4, 1.
36. Liu, Y.; Lu, D.; Gonzalez-Lafont, A.; Truhlar, D. G.; Garret, B. C. *J Am Chem Soc* 1993, 115, 7806.
37. Alhambra, C.; Wu, L.; Zhang, Z.-Y.; Gao, J. *J Am Chem Soc* 1998, 120, 3858.
38. Bernal-Uruchurtu, M. I.; Martins-Costa, M. T. C.; Millot, C.; Ruiz-López, M. F. *J Comput Chem* 2000, 21, 572.
39. Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. *J Comput Chem* 2002, 23, 1601.
40. Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. *J Comput Chem* 2002, 23, 498.
41. Weber, W.; Thiel, W. *J Theor Chem Acc* 2000, 1003, 485.
42. Stewart, J. J. P. *MOPAC 2002*; Fujitsu Ltd: Tokyo, Japan, 1999.
43. Rocha, G. B.; Freire, R. O.; Simas, A. M.; Stewart, J. J. P. *J Comput Chem* 2006, 27, 1101.
44. Miertus, S.; Scrocco, E.; Tomasi, J. *Chem Phys* 1981, 55, 117.
45. Miertus, S.; Tomasi, J. *J Chem Phys* 1982, 65, 239.
46. Pierotti, R. A. *Chem Rev* 1976, 76, 717.
47. Claverie, P. In *Intermolecular Interactions: From Diatomics to Biomolecules*; Pullman, B., Ed.; Wiley: Chichester, 1978.
48. Colominas, C.; Luque, F. J.; Teixidó, J.; Orozco, M. *Chem Phys* 1999, 240, 253.
49. Pascual-Ahuir, J. L.; Silla, E.; Tomasi, J.; Bonaccorsi, R. *J Comput Chem* 1987, 8, 778.
50. Luque, F. J.; Illas, F.; Orozco, M. *J Comput Chem* 1990, 11, 416.
51. Ferenczy, G. G.; Reynolds, C. A.; Richards, W. G. *J Comput Chem* 1990, 11, 159.
52. Cummins, P. L.; Gready, J. E. *Chem Phys Lett* 1990, 174, 355.
53. Alhambra, C.; Luque, F. J.; Orozco, M. *J Comput Chem* 1994, 15, 12.
54. Luque, F. J.; Reuter, N.; Cartier, A.; Ruiz-López, M. F. *J Phys Chem A* 2000, 104, 10923.
55. Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J Chem Theor Comput* 2005, 1, 1133.
56. Zhan, C.; Dixon, D. A. *J Phys Chem A* 2001, 105, 11534.
57. Tissandier, M. D.; Cowen, K. A.; Feng, W. Y.; Gundlach, E.; Cohen, M. H.; Earhart, A. D.; Coe, J. V.; Tuttle, T. R., Jr. *J Phys Chem A* 1998, 102, 7787.
58. Orozco, M.; Luque, F. J.; Habibollah-Zadeh, D.; Gao, J. *J Chem Phys* 1995, 102, 6145.
59. Luque, F. J.; Zhang, Y.; Alemán, C.; Bachs, M.; Gao, J.; Orozco, M. *J Phys Chem* 1996, 100, 4269.
60. Orozco, M.; Luque, F. J. *Chem Phys* 1994, 182, 237.
61. Curutchet, C.; Bidon-Chanal, A.; Soteras, I.; Orozco, M.; Luque, F. J. *J Phys Chem B* 2005, 109, 3565.
62. Dewar, M. J. S.; Dieter, K. M. *J Am Chem Soc* 1986, 108, 8075.
63. Voets, R.; François, J.-P.; Martin, J. M. L.; Mullens, J.; Yperman, J.; Van Pouche, L. C. *J Comput Chem* 1989, 10, 449.
64. Szafran, M.; Koput, J. *J Comput Chem* 1991, 12, 675.
65. Ozment, J. L.; Schmiedekamp, A. M. *Int J Quantum Chem* 1992, 43, 783.

Role of Heme Distortion on Oxygen Affinity in Heme Proteins: The Protoglobin Case

Damián E. Bikiel,[†] Flavio Forti,[‡] Leonardo Boechi,[†] Marco Nardini,[§] F. Javier Luque,[‡] Marcelo A. Martí,^{†,||,*} and Darío A. Estrin^{†,*}

Departamento de Química Inorgánica, Analítica y Química Física/INQUIMAE-CONICET, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pabellón 2, Buenos Aires, C1428EHA, Argentina, Departament de Fisicoquímica and Institut de Biomedicina (IBUB), Facultat de Farmàcia, Universitat de Barcelona, Av. Diagonal 643, 08028, Barcelona, Spain, Department of Biomolecular Sciences and Biotechnology, CNR-INFM, University of Milano, Milano, Italy, and Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pabellón 2, Buenos Aires, C1428EHA, Argentina

Received: March 9, 2010

The chemical properties of heme proteins largely reflect the electronic properties of their heme group. Often, the porphyrin ring of the heme exhibits significant distortions from its isolated structure, but the impact of these distortions on the chemical properties of the heme is yet uncertain. A systematic study focused on the effects of the distortion of the macrocycle on the binding affinity for oxygen is presented. The results show that out-of-plane distortions decrease the binding affinity, while in-plane distortions can increase or decrease it. Among in-plane distortions, only the *breathing* mode, which involves the symmetric compression–expansion of the porphyrin ring, strongly modulates the binding affinity. These findings shed light into the peculiar binding affinity of *Methanosarcina acetivorans* protoglobin, a protein that contains a highly distorted heme. Overall, the results highlight that in-plane distortions might be exploited by certain classes of heme proteins to modulate the ligand affinity.

Introduction

Heme proteins can play a wide variety of functional roles including storage and transport of small ligands, redox processes, or sensing of diatomic compounds. Valuable qualitative indications about the role of heme proteins come from the binding properties of small diatomic ligands, such as the kinetic rate constants of ligand association (k_{on}) and dissociation (k_{off}) processes.^{1,2} Typical k_{on} values range from 10^5 to 10^8 $\text{M}^{-1} \text{s}^{-1}$, while k_{off} values can vary by up to 7 orders of magnitude (from 10^{-3} to 10^4 s^{-1}).^{3,4} O_2 storage or transport proteins (such as myoglobin, Mb, or hemoglobin, Hb) display moderate affinities ($K_{\text{O}_2} \approx 1 \mu\text{M}^{-1}$) with large association rates and moderate to high dissociation rate constants (1 – 10 s^{-1}). In contrast, proteins acting as NO dioxygenases, such as *Ascaris* Hb (AscHb)⁵ or *M. tuberculosis* truncated Hb N (MfTrHbN), display high O_2 affinities due to low or very low dissociation rate constants (0.001 – 0.1 s^{-1}).⁶ Clearly, understanding the factors that modulate ligand affinity is crucial for a comprehensive knowledge of the physiological role of heme proteins and has been a constant effort of many different research groups over the past few decades.⁷

Association processes mainly depend on the accessibility of ligands to the heme distal site, which can involve either a gated mechanism, as in Mb,⁸ or even migration through cavities/tunnels, as in truncated Hbs.^{9,10} Dissociation rate constants are determined by various factors such as the interaction between the ligand and the heme (i.e., the strength of the Fe–ligand bond), the interactions between the ligand and the surrounding protein matrix (i.e., hydrogen-bond (HB) or other forms of noncovalent

interactions with distal residues that can either be stabilizing or destabilizing) and the probability of escape from the ligand pocket. For instance, the very low O_2 dissociation from oxygenated MfTrHbN can be attributed to multiple HB interactions with TyrB10 and GlnE11 residues.^{6,11} In other cases, like leghemoglobin, the strength of the Fe–O bond can be further modulated by proximal interactions.¹² Ligand escape has also been found to be a key issue in determining kinetic constants, specifically k_{off} .¹³

Collman¹⁴ and Ibers¹⁵ studied synthetic models of porphyrins to shed light into the steric effect of the environment on the binding of CO and O_2 in heme proteins, principally by strains over the ligand. However, a much less studied factor in controlling the intrinsic heme reactivity concerns the distortions of the porphyrin ring due to the protein environment. The reason for such a lack of investigation on porphyrin distortion and heme reactivity may be related to the limited experimental resolution achieved in many crystallographic investigations that did not allow mapping with the required precision the contained deviations from planarity of the porphyrin system. A systematic classification of heme distortions, referred to as normal-coordinate structural decomposition (NSD), has been proposed by Jentzen et al.,¹⁶ who identified the most relevant out-of-plane (saddling, ruffling, doming, X-waving, Y-waving, and propeller) and in-plane (meso-stretching, N-pyrrole stretching, pyrrole translation (X,Y), breathing, and pyrrole rotation) normal modes that relate the structure of a distorted heme compared to a reference D_{4h} structure (see Figure 1).^{17,18} Noteworthy, it has been shown that different heme protein families retain similar heme distortions.¹⁶

A highly distorted heme has recently been found in *Methanosarcina acetivorans* protoglobin (Pgb),¹⁹ which is a single domain archeal heme protein related to the globin domain of

* To whom correspondence should be sent. E-mail: D.A.E., dario@qi.fcen.uba.ar; M.A.M., marcelo@qi.fcen.uba.ar.

[†] Departamento de Química Inorgánica, Universidad de Buenos Aires.

[‡] Universitat de Barcelona.

[§] University of Milano.

^{||} Departamento de Química Biológica, Universidad de Buenos Aires.

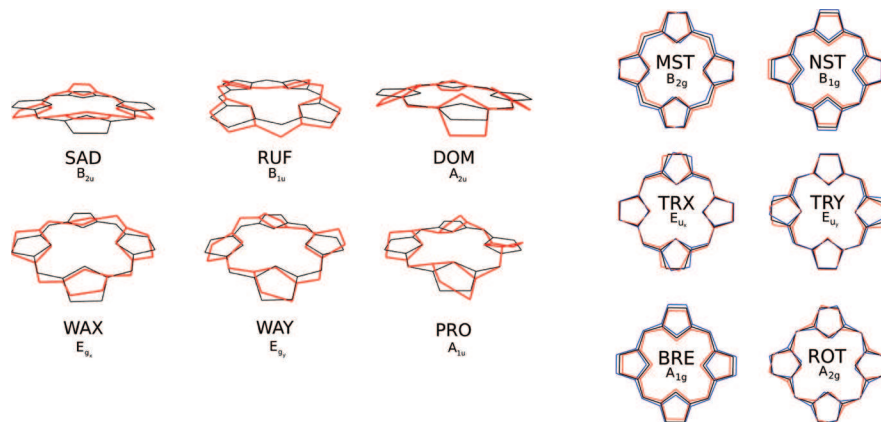


Figure 1. Heme distortions. Left: out-of-plane distortions (red) are doming (*dom*), ruffling (*ruf*), saddling (*sad*), waving (*wax*, *way*) and propelling (*pro*). Right: positive (red) and negative (blue) in-plane distortions involving meso-stretching (*mst*), *N*-pyrrole stretching (*nst*), pyrrole translation (*trx*, *try*), breathing (*bre*), and pyrrole rotation (*rot*). The reference D_{4h} structure is shown in black.

globin-coupled sensor proteins.^{20–22} Pgb can bind O₂ reversibly in vitro, but the very low O₂ dissociation rate (0.092–0.0094 s⁻¹) is completely unexpected as there is no distal HB stabilization of the heme-bound O₂.¹⁹ In addition, inspection both of X-ray structure and of the molecular dynamics trajectory show an accessible heme moiety, suggesting that no ligand escape effect is present. This structural feature raises challenging questions about the precise function of Pgb, and particularly on the effect of the unusual heme distortion on Pgb reactivity.^{21,22}

Heme distortions have been suggested to modulate heme reactivity in proteins. Thus, the effect of the out-of-plane distortions on properties such as the UV–vis spectrum, redox potential, and ligand affinity has been examined in recent studies.^{23–30} Moreover, the significant ruffling observed in nitrophorins has been related to the relative stability of the ferrous/ferric states.²³ Finally, heme distortions in nitric oxide synthase have been related to a subtle tuning of its electronic structure.³¹ Nevertheless, to the best of our knowledge, there is no systematic study of the effect of heme distortions on ligand binding affinity. To fill this gap, we examine here the relationship between heme distortion and O₂ affinity. On the basis of the results of this analysis, we suggest an explanation for the strong structural distortion of the heme in Pgb and its unusual O₂ affinity.

Computational Methods

Quantum Mechanical Calculations. In the NSD framework the distortion of a porphyrin ring (D_{obs} ; eq 1) is quantified from the displacements observed for C and N atoms in the structure of the macrocycle (S_{obs}) relative to a reference D_{4h} ring (S_{ref} ; see eq 1). In turn, D_{obs} can be written as a linear combination of the normal mode displacements (\hat{D}_i ; eq 2).

$$D_{\text{obs}} = S_{\text{obs}} - S_{\text{ref}} \quad (1)$$

$$D_{\text{obs}} = \sum_i d_i \hat{D}_i \quad (2)$$

where d_i is the coefficient associated with the normal mode displacement vector \hat{D}_i .

The most relevant (six out-of-plane and six in-plane; see Figure 1) normal modes were derived from the normal-mode analysis carried out for the geometry of a Fe(II)–porphyrin (tetracoordinated, 4c) optimized at the B3LYP/6-31G** level.^{32–34}

To explore the effect of the heme distortion on the O₂ binding, ad hoc distorted 4c structures (six for each mode) were built up by projecting each relevant normal mode on the porphyrin structure in both positive and negative directions until a maximum value of positional root-mean square (determined only for C and N atoms relative to the reference structure) of approximately 0.5 or 1.2 Å (for in-plane and out-of-plane distortions, respectively) which are typical values observed in the crystals. An imidazole was coordinated to the iron to simulate the environment at the active site of the proteins. The coordinated imidazole defines the proximal site in the model, leaving vacant the other side (distal) for diatomic ligand binding. Pentacoordinated (5c, distorted porphyrin and imidazole) and hexacoordinated (6c, distorted porphyrin, imidazole, and oxygen) structures were calculated as noncharged quintuplets and singlets, respectively, according to the experimental spin states.^{35–37} The O₂ molecule was calculated as a triplet. Each distorted structure (X) was partially optimized, constraining all the C atom positions, but allowing geometrical relaxation of Fe, H, and N atoms of the porphyrin, the Fe-bound O₂ and the coordinated imidazole using the *modredundant* key in Gaussian'03 package. Despite each mode comprises all the C and N atoms, almost a negligible contamination was observed due to the relaxation of the N. This protocol allowed us to obtain optimized structures with an heme C-backbone distorted. 5c and 6c structures were partially optimized independently for each structure. Finally, the binding affinity was estimated from the difference between the energy of bonded and nonbonded states (eq 3).

$$\Delta E_{\text{binding}}^{\text{X}} = E_{\text{porO}_2}^{\text{X}} - E_{\text{por}}^{\text{X}} - E_{\text{O}_2} \quad (3)$$

where $E_{\text{porO}_2}^{\text{X}}$, $E_{\text{por}}^{\text{X}}$, and E_{O_2} denote the energy of the Fe-bound O₂, the separate porphyrin, and O₂, respectively, for a certain distorted structure X.

From a practical viewpoint, it is convenient to compare $\Delta E_{\text{Binding}}^{\text{X}}$ with the binding energy of the reference D_{4h} structure ($\Delta E_{\text{Binding}}^{\text{ref}}$), which in turn allows us to express the change in binding energy due to heme distortion ($\Delta(\Delta E)$) in terms of the energy cost due to distortion in both oxy ($\Delta E_{\text{porO}_2}^{\text{X}}$) and deoxy ($\Delta E_{\text{por}}^{\text{X}}$) states of the heme (eqs 4 and 5).

$$\Delta(\Delta E) = \Delta E_{\text{Binding}}^{\text{X}} - \Delta E_{\text{Binding}}^{\text{ref}} \quad (4)$$

$$\Delta(\Delta E) = E_{\text{PorO}_2}^{\text{X}} - E_{\text{PorO}_2}^{\text{ref}} - E_{\text{Por}}^{\text{X}} + E_{\text{Por}}^{\text{ref}} = \Delta E_{\text{PorO}_2}^{\text{X}} - \Delta E_{\text{Por}}^{\text{X}} \quad (5)$$

Hybrid Quantum-Classical (QM-MM) Calculations. The initial structure for QM-MM calculations was built up starting from the crystal structure of monomeric Pgb (PDB code 2veb).¹⁹ Molecular dynamics simulations using the AMBER force field showed that the standard parameters included for heme were not accurate enough to describe properly heme distortions. To maintain the observed distortion in the heme, we decided to alter the standard protocol used in QM-MM calculation. A minimization of the crystal structure was performed by freezing all the C's atom in heme, relaxing the system without compromising the distorted structure of the heme. Then, hybrid QM-MM geometry optimizations were performed using a conjugate gradient algorithm at the DFT level with the SIESTA code³⁸ using our own QM-MM implementation,³⁹ fixing all the C atoms in the heme. The SIESTA code showed an excellent performance for medium and large systems and also proved to be appropriate for biomolecules, and specifically for heme models.^{12,40–44} Only residues located less than 10 Å apart from the heme reactive center were allowed to move freely in QM-MM runs. Calculations were performed using the generalized gradient approximation functional proposed by Perdew, Burke, and Ernzerhof.^{45,46} For all atoms, basis sets of double plus polarization quality were employed as previous works showed that both QM-MM optimized structures and binding energies are properly described at this level of theory. The QM subsystem included the Fe–porphyrin ring (excluding side chains) and the relevant ligands (imidazole and O₂), while the rest of the protein and water molecules were treated classically. The scaled position link atom method^{47,48} adapted to our SIESTA code was used for the interface between QM and MM regions. Further technical details about the QM-MM implementation can be found elsewhere.⁴⁹ Finally, O₂ binding energies ($\Delta E_{\text{binding}}$) were calculated as

$$\Delta E_{\text{Binding}} = E_{\text{protO}_2} - E_{\text{prot}} - E_{\text{O}_2} \quad (6)$$

where E_{protO_2} , E_{prot} , and E_{O_2} are the energy of the O₂-bound protein, the free protein, and isolated O₂.

Results

Normal Decomposition of the Penta- and Hexacoordinated Structures in the Porphyrin Ring. The NSD analysis of the fully optimized structures (nonconstrained structures) of the penta- and hexacoordinated porphyrin ring is shown in Table 1. In the pentacoordinated case (which leaves the sixth Fe coordination site immediately available for ligand binding), the principal distortions are the out-of plane negative *dom*, which

TABLE 1: Normal Structure Decomposition Analysis (OOP = Out-of-Plane; IP = In-Plane) of the Full-Optimized Structures of Penta- and Hexacoordinated Porphyrin Ring (in Å)

OOP	penta	hexa	IP	penta	hexa
<i>sad</i>	−0.030	−0.010	mst	0.000	0.036
<i>ruf</i>	0.000	0.234	nst	0.041	0.000
<i>dom</i>	−0.231	−0.043	trx	0.001	0.001
<i>wax</i>	0.000	−0.004	try	0.000	−0.001
<i>way</i>	0.001	−0.008	bre	−0.227	−0.051
<i>pro</i>	0.000	−0.000	rot	0.000	0.000

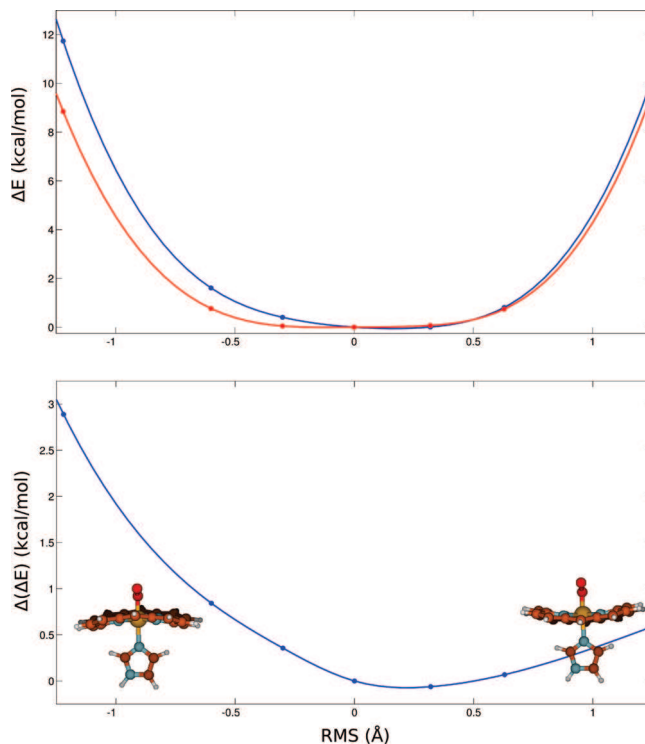


Figure 2. Top: energy cost profiles (kcal/mol) for the distortion of the heme following the *ruf* mode in oxy (blue) and deoxy (red) states. The reference value is the respective nondistorted structure (see eqs 4 and 5). Bottom: difference between the two distortion energy profiles. In the X-axis, the distortion value is in rms (Å).

TABLE 2: $\Delta(\Delta E)$ (kcal/mol) Values for Selected Distortions (in Å) along Out-of-Plane Modes

	rmsd (Å)						
	−1.2	−0.6	−0.3	0.0	0.3	0.6	
<i>sad</i>	1.9	0.5	0.1	0.0	0.1	0.5	2.4
<i>ruf</i>	2.9	0.8	0.4	0.0	−0.1	0.1	0.6
<i>dom</i>	7.0	2.6	1.0	0.0	−0.3	0.0	2.1
<i>wax/way</i>	3.6	0.7	0.2	0.0	0.3	1.2	3.5
<i>pro</i>	2.2	0.5	0.1	0.0	0.1	0.5	2.2

indicates that the imidazole is pulling the Fe (out-of-plane), causing a doming toward the imidazole, and the in-plane negative *bre*, which indicates an expanded ring compared to the reference structure. In contrast, for the hexacoordinated structure the distortion is dominated by the *ruf* mode, while the in-plane negative *bre* distortion almost vanishes.

Effect of Out-of-Plane (OOP) Distortions on the Oxygen Affinity. The energy cost for the *ruf* distortion (see eqs 4 and 5) of the heme in both oxy and deoxy states (Figure 2) reflects an almost harmonic behavior. After the zero of both energy profiles is set at their minimum value (i.e., minimum energy structure), subtraction of both curves allows us to estimate the effect of ruffling on the O₂ binding affinity compared to that on the ideal planar ring ($\Delta(\Delta E)$; see Figure 2). A positive *ruf* distortion leads to a slight destabilization of the binding energy (about 0.5 kcal/mol for a ruffling displacement of 1.2 Å), while negative distortions give rise to a larger destabilization (about 3 kcal/mol for a ruffling displacement of −1.2 Å).

Table 2 summarizes the corresponding $\Delta(\Delta E)$ values for the relevant out-of-plane distortions (the corresponding energy plots are reported in Supporting Information; see Figures S1–S4). In general, all the out-of-plane distortions tend to decrease the binding affinity. Only very minor positive doming and ruffling

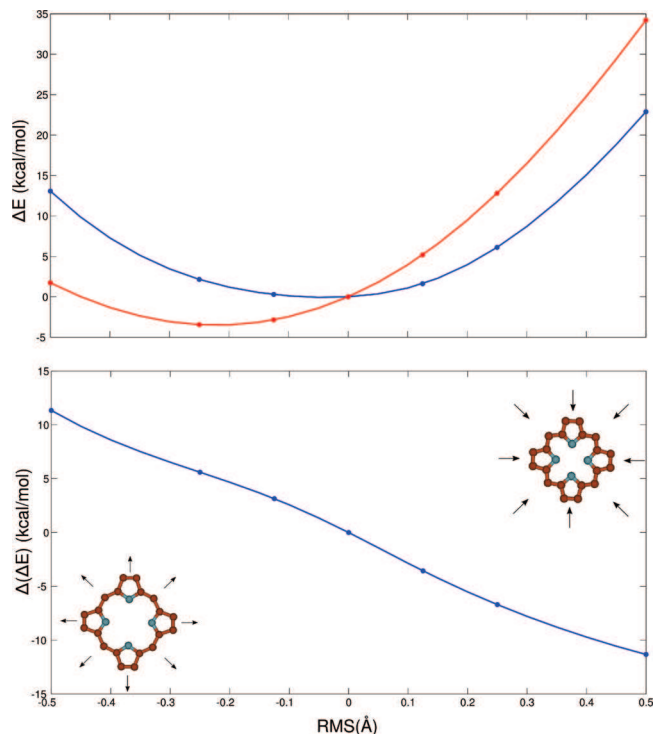


Figure 3. Top: energy cost profiles (kcal/mol) for the distortion of the heme following the *bre* mode in oxy (blue) and deoxy (red) states. The reference value is the respective nondistorted structure (see eqs 4 and 5). Bottom: difference between the two distortion energy profiles. In the X-axis, the distortion value is in rms (Å).

distortions lead to a very small increase in the binding affinity (by -0.3 kcal/mol at a rmsd of 0.3 Å), this effect being likely due to differences in the heme equilibrium geometry in the oxy/deoxy states. Moreover, a negative modulation of O_2 binding affinity by out-of-plane distortions is difficult to envisage, since a reduction in binding energy larger than 4 kcal/mol demands an extreme deformation of the porphyrin plane, especially for doming. On the other hand, the asymmetry observed for some of the modes can be mainly ascribed to two features: the position of the energy minimum in both penta- and hexacoordinated structures, and the curvatures of the energy profiles around the minimum energy structure. If both minima are located at nearly the same distortion value, then an almost symmetric behavior is observed.

In summary, out-of-plane distortions tend to decrease the oxygen binding energy, though the magnitude of these effects is generally small for structural distortions around the minimum energy structure, suggesting a minor contribution to modulation of the oxygen affinity in heme proteins.

Effect of the In-Plane (IP) Distortions on the Oxygen Affinity. Figure 3 shows the distortion energy determined for the *bre* mode in oxy and deoxy states. As noted above, the energy profiles exhibit a harmonic behavior. The differences in both curvatures and, principally, in the location of the minimum energy structure give rise to an interesting trend in the $\Delta(\Delta E)$ curve. First, as expected, the breathing distortion is much stiffer than out-of-plane modes, as noted in the fact that an rmsd of 0.5 Å yields $\Delta(\Delta E)$ values close to 12 kcal/mol. Second, negative displacements (i.e., an expansion of the porphyrin ring) are associated with a destabilization of the O_2 binding energy, whereas the opposite trend is found for positive displacements (i.e., a compression of the porphyrin ring).

The overall impact of the different in-plane distortions on the binding affinity can be gained from inspection of Table 3

TABLE 3: $\Delta(\Delta E)$ (kcal/mol) for Selected Positive and Negative Distortions (in Å) along In-Plane (IP) Modes

	rms						
	-0.5	-0.25	-0.125	0.0	0.125	0.25	0.5
<i>mst</i>	2.1	0.8	0.2	0.0	-0.3	-0.2	0.1
<i>nst</i>	2.0	1.1	0.3	0.0	0.3	1.1	2.0
<i>trx/try</i>	0.3	0.1	0.0	0.0	0.0	0.0	0.2
<i>bre</i>	11.3	5.6	3.1	0.0	-3.6	-6.7	-11.3
<i>rot</i>	2.0	0.6	0.2	0.0	0.0	0.2	1.3

(energy profiles available in the Supporting Information; see Figures S5–S8). Negative meso-stretching (*mst*) distortion reduces the binding energy, but positive displacements along this normal mode have very little impact on the O_2 affinity. In contrast, both negative and positive distortions along the N-pyrrole stretching (*nst*) decrease the O_2 binding energy. However, the most striking finding concerns breathing (*bre*), as this mode leads to an unexpected increase in the binding affinity of the heme group for O_2 .

Overall, both out-of-plane and in-plane distortions tend to reduce the O_2 affinity of the porphyrin ring, even though subtle differential trends regarding the magnitude and the symmetrical behavior of such a destabilization for positive or negative displacements are found for the different normal modes. The distortion energy determined for the deoxy heme model is generally softer than that obtained for the oxy one. Accordingly, if the minima of both energy profiles are similar, an increase in the distortion is then associated with a stabilization of the pentacoordinated heme model relative to the hexacoordinated one, leading to a net reduction in the O_2 affinity of the distorted heme compared to that of the reference structure. The only exception to this general behavior involves breathing, as the substantial difference found between the minimum energy structures of penta- and hexacoordinated states makes that positive *bre* displacements increase the O_2 affinity.

Comparison of the Effect on CO Affinity. Similar calculations were carried out using CO as ligand. In Figure 4, selected modes (*bre* and *ruf*) for O_2 and CO are plotted together. As can be seen, the trends are similar for both ligands. Moreover, the fact that both ligands respond in the same way to *bre*, suggest that the effect is not an artifact. Further analysis should be done to understand the discrimination between CO and O_2 . In view of these results, a reassessment of the experimental information should be done.

Electronic Effects of Breathing Mode. The influence of the expansion–compression due to the breathing distortion on the Fe– O_2 bond can be gained from Table 4. For the nondistorted model the net charge (derived from NBO analysis) on the bound oxygen is $-0.343e$. This value is little affected upon expansion of the porphyrin ring (negative *bre* distortion). Nevertheless, while in the nondistorted heme there is an asymmetry in the charge distribution between the two oxygen atoms ($O1$, $-0.156e$; $O2$, $-0.187e$), the atomic charge is largely similar in the expanded structure ($O1$, $-0.169e$; $O2$, $-0.173e$). In contrast, porphyrin compression (positive *bre* distortion) results in increased net charge on the oxygen molecule and also in an enhanced asymmetry in the charge distribution between the two oxygens ($O1$: $-0.161e$; $O2$: $-0.196e$). With regard to the iron atom, expansion of the porphyrin ring increases the positive charge by $0.045e$, whereas compression reduces the positive charge by $0.054e$. Finally, the net charge of the imidazole ring remains nearly unaffected by the breathing distortion (see Table 4).

The net effect reflects an electronic density transfer from the porphyrin to the Fe-bound O_2 . Thus, the overall change in the

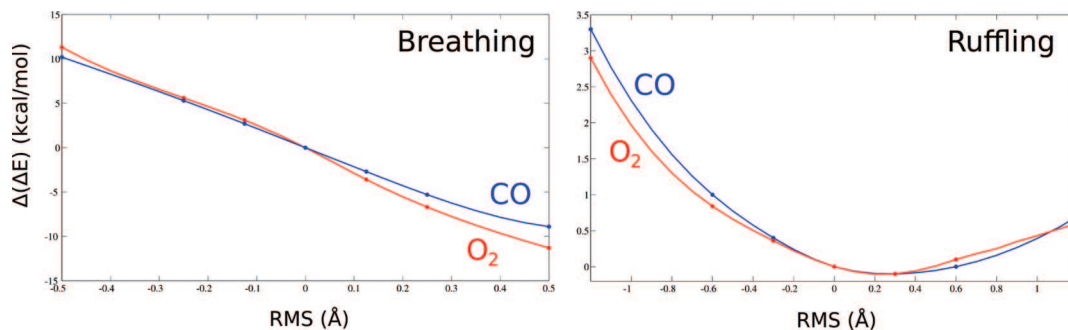


Figure 4. $\Delta(\Delta E)$ (kcal/mol) versus distortion (\AA) for CO (blue) and O_2 (red). Comparison of the *bre* (left panel) and *ruf* (right panel) modes.

TABLE 4: Comparative Analysis of the NBO Charge (in Units of e) and Bond Lengths (in \AA) in the Expanded and Compressed Porphyrins^a

<i>bre</i> (\AA)	$d(\text{O}-\text{O})$	$d(\text{Fe}-\text{O})$	$q_{\text{O}_1}/q_{\text{O}_2}$ (q_{O_2})	q_{Fe}	$q_{\text{Imidazole}}$
-0.5	1.268	1.737	-0.169/-0.173 (-0.342)	1.505	0.128
0.0	1.271	1.741	-0.156/-0.187 (-0.343)	1.460	0.125
0.5	1.274	1.745	-0.161/-0.196 (-0.357)	1.406	0.122

^a q_{O_1} is the charge on the oxygen closer to iron, q_{O_2} is the charge on the remaining oxygen. In parentheses is the net charge on O_2 .

$\text{Fe}-\text{O}_2$ charge upon expansion (+0.046 e) or compression (-0.068 e) of the porphyrin ring nearly matches the change in the charge of the porphyrin N atoms (-0.041 e and +0.064 e for expansion and compression, respectively). These changes are associated with shortening of the $\text{Fe}-\text{O}_1$ and $\text{Fe}-\text{N}_{\text{im}}$ bonds upon expansion of the ring, and with enlargement of those bonds upon porphyrin compression (see Table 4; see also the Supporting Information, Figure S9). As the compression proceeds, the density is “pushed” from the equatorial orbitals to the axial ones, which are then capable of interacting with both O_2 and imidazole at longer distances. Moreover, compression is accompanied by an enlargement of the $\text{O}-\text{O}$ distance, which results from an increased population of a partially occupied antibonding orbital (see Figure S10 in the Supporting Information). It could be inconsistent that higher oxygen binding energies are not correlated with the $\text{Fe}-\text{O}$ bond distance. However, as the binding energy arises from the comparison of two independent structures (5c and 6c) both subject to the distortion constraint, while the $\text{Fe}-\text{O}$ distance is only observed on the 6c species, if the distortion effect is more pronounced in the 5c state, a change in the binding energy may not necessarily be correlated with $\text{Fe}-\text{O}$ distance. Parts a and b of Figure 5a,b show selected NBO frontier orbitals located on Fe. These orbitals show clearly the effect of the *bre* mode. In (a) it can be observed that the orbitals from the pyrrolic nitrogens in the expanded structure (left) interact with the d_{z^2} of the iron. However, upon compression, this interaction is reduced, as can be observed on the enlargement of the axial component of the iron orbital. The result is that, upon compression, there exists more density on the axial component. In (b) the interaction between the pyrrolic nitrogens and the iron orbitals that participate on the $\text{Fe}-\text{O}_2$ bonding orbitals can be observed (see Figure S12, Supporting Information). As the compression proceeds, the orbitals become more compact. This increase in density can be transferred to the antibonding orbital of the O_2 via the iron d orbitals with proper symmetry.

Decomposition of the Pgb Oxygen Affinity. The preceding findings made us speculate that heme proteins might exploit the breathing mode to regulate the binding of ligands, since a positive/negative *bre* distortion would increase/decrease the heme–ligand affinity. In fact, the large distortion of the heme

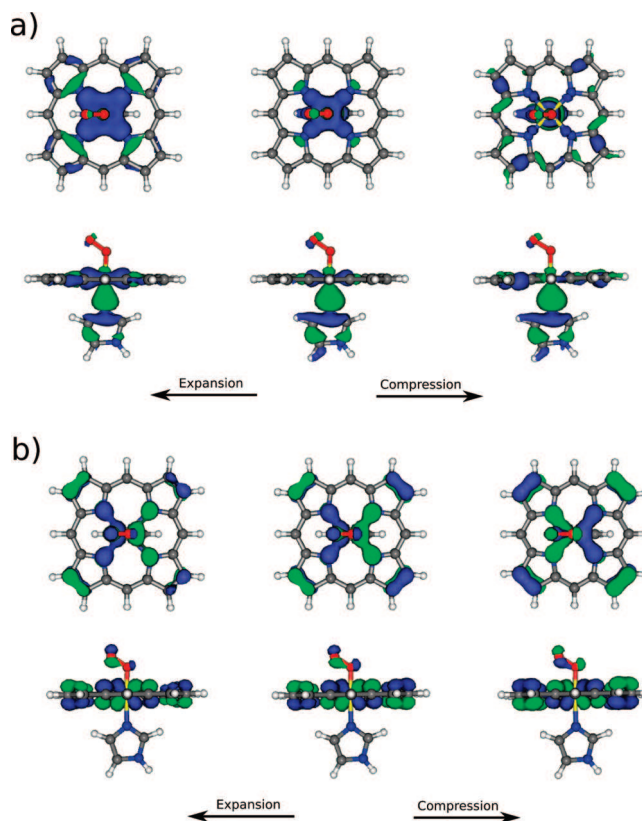


Figure 5. Side and upper views of selected NBO frontier orbitals.

found in the 1.3 \AA X-ray crystal structure of Pgb could explain the unusual low O_2 dissociation rate. At this point, it is worth noting that such a low rate cannot be justified a priori by additional interactions with the heme-bound O_2 due to the absence of distal residues able to form a hydrogen-bond with the ligand. In fact, the only distal residue that might stabilize the heme-bound O_2 is TyrB10, which has been found to provide hydrogen bonding to the heme ligand in several globins^{11,12,50–52} (see Figure 6). Nevertheless, the X-ray structure of Pgb reveals that TyrB10 not only forms a hydrogen-bond interaction with the backbone of LeuE4 but is also too distant from the ligand to establish a stable hydrogen-bond interaction. This trend was confirmed from the analysis of the trajectory collected in a 500 ns molecular dynamics simulation of oxygenated Pgb. Even after reorienting TyrB10 in the starting simulation structure to favor a hydrogen-bond to the heme ligand, no significant hydrogen-bonded structure was observed (see Figure S11 in the Supporting Information).

Previous works have shown that QM/MM estimates of the O_2 binding energies are valuable to explain the O_2 dissociation rate constant in heme proteins, allowing us to detect distal,

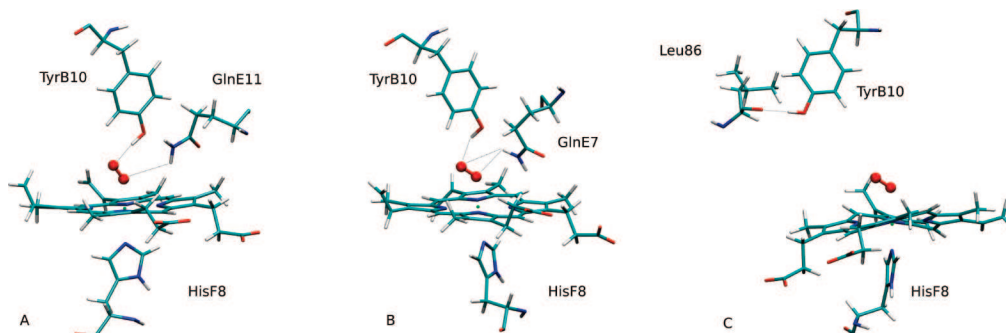


Figure 6. Active site and distal hydrogen-bond donors of *Ascaris hemoglobin* (A), *Mycobacterium tuberculosis truncated hemoglobin N* (B), and *Protoglobin* (C).

TABLE 5: Comparison between Experimental k_{off} and QM/MM $\Delta(\Delta E)$ (kcal/mol) Relative to the Isolated Heme

protein	$\Delta(\Delta E)$	no. of distal HB	k_{off} (s^{-1})	ref
Pgb	-7.6	0	9.4×10^{-3} to 9.2×10^{-2}	6
Mb	-5.0	1	12	58
Mb HE7 V	-0.7	0	1600	59
trHbN	-15.2	2	0.2	58
trHbN YB10A	-7.8	1	45	60
AscHb	-12.3	2	0.4×10^{-2}	5

TABLE 6: NSD Analysis of the Heme in the Oxygenated Forms of Myoglobin (Mb, PDB code 1A6M) and Protoglobin (Pgb, PDB code 2VEB) (Distortions in Å)

OOP	Mb	Pgb	IP	Mb	Pgb
<i>sad</i>	-0.185	-0.390	mst	-0.017	-0.042
<i>ruf</i>	-0.017	+1.419	nst	-0.021	+0.006
<i>dom</i>	+0.206	-0.013	trx	+0.000	+0.055
<i>wax</i>	-0.026	-0.132	try	-0.014	+0.087
<i>way</i>	+0.186	+0.235	bre	-0.023	+0.202
<i>pro</i>	-0.071	+0.019	rot	+0.012	+0.013

proximal, and also dynamical effects.^{3,12,53–57} Table 5 provides the $\Delta(\Delta E)$ values relative to the isolated heme obtained for O₂ binding to wild type (wt) forms of Mb, AscHb, MtTrHbN, and Pgb. As expected, AscHb and MtTrHbN possess the largest increase in binding affinity (12.3 and 15.2 kcal/mol, respectively) followed by Pgb (7.6 kcal/mol) and Mb (5.0 kcal/mol), which can be explained by the presence of two (MtTrHbN, AscHb; see Figure 6) and one (Mb) hydrogen-bond between heme-bound O₂ and distal residues in the heme cavity (TyrB10 and GlnE7 in AscHb, TyrB10 and GlnE11 in MtTrHbN, and HE7 in Mb). The critical role of these hydrogen bonds on the O₂ affinity is demonstrated by the results obtained for the mutants HisE7→Val in Mb and TyrB10→Ala in MtTrHbN, which reveal a drastic decrease in $\Delta(\Delta E)$ (see Table 5). In this context, the O₂ affinity predicted for Pgb is surprisingly large, keeping in mind the absence of stabilizing hydrogen-bonds with distal residues. In turn, it can be speculated that the origin of the intrinsic O₂ affinity found for Pgb could be related to the large distortion of the heme.

The NSD analysis of the porphyrin ring in Pgb reveals that the main out-of-plane contribution to the heme distortion is ruffling, which gives rise to a displacement of 1.42 Å, whereas the *ruf* distortion in the heme of Mb only amounts to 0.02 Å (Table 6). Regarding the in-plane distortions, the main difference between the heme moieties in Mb and Pgb affect the breathing mode, which leads to a displacement around 10-fold larger in Pgb compared to that in Mb. The presence of such a high positive *bre* distortion, and hence a high heme compression,

TABLE 7: Heme Structural Parameters for Myoglobin (1A6M), Fe(II) Porphyrin Reference, and Protoglobin (2VEB)^a

	Mb (1A6M)	Fe(II) Por 4c (reference)	Pgb (2VEB)
$d_{\text{C}_{\text{meso}}}$	6.80 ± 0.01	6.84	6.65 ± 0.04
$d_{\text{N}_{\text{pyrrol}}}$	4.01 ± 0.01	3.98	3.963 ± 0.008
$d_{\text{C}_{\beta}}$	8.54 ± 0.03	8.53	8.44 ± 0.04
$d_{\text{Fe-N}}$	2.01 ± 0.01	1.99	1.99 ± 0.01

^a The values are averaged (and standard deviations) in Å. See Figure 5 for the definitions.

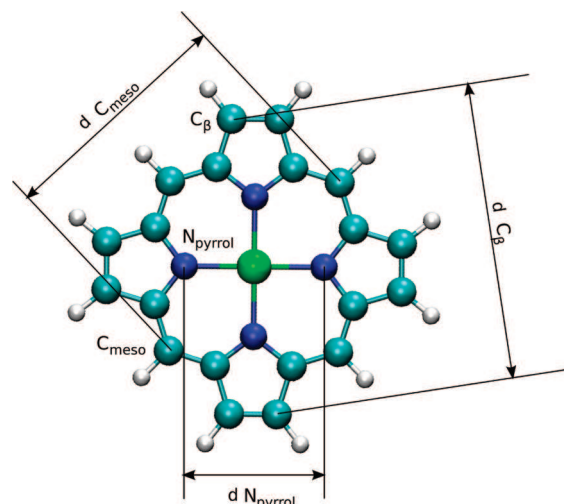


Figure 7. Definition of the structural parameters used in Table 7 to evaluate the porphyrin expansion-compression.

was never noticed before. Accordingly, both ruffling and breathing should be the main normal modes that might affect the binding properties of the heme in Pgb.

A classic view of the compression in the Pgb heme ring can be observed by measuring some structural parameters.⁶¹ In Table 7 we present the comparison of the structural parameters exemplified in Figure 7 for the myoglobin (1A6M), the protoglobin (2VEB) and the planar porphyrin 4c reference used in the NSD analysis.

The most direct measure of the ring core is the distance between faced nitrogens ($d_{\text{N}_{\text{pyrrol}}}$) and the Fe-N distance. While the porphyrin reference presents a value of 3.98 Å, Pgb presents a decreased value of 3.96 Å (0.02 Å from the reference) and the Mb an increased value of 4.01 Å. These values are in agreement with the *bre* distortion observed by NSD. The variation in Fe-N distance is small. For Pgb almost no modification from the reference is observed, while for Mb only a small increase is found. The distances between C_β shows a clear compression for Pgb (0.1 Å from the reference), while a

TABLE 8: Effect of the Out-of-Plane (OOP) and In-Plane (IP) Distortions Detected in Pgb on the O₂ Binding Affinity for a Heme Model System at Different Levels of Density Functional Theory (Energy in kcal/mol)

Pgb heme model	B3LYP 6-31G**	PBE 6-31G**	OLYP 6-31G**	B3LYP TZVPP (CHNO) QZVPP (Fe)	PBE (SIESTA)
OOP+IP	-3.2	-2.1	-4.3	-4.3	-2.1
Only IP	-4.7	-4.0	-5.5	-4.9	-4.1
Only OOP	1.5	1.9	1.1	0.6	2.0

negligible expansion to Mb. Finally, measuring the faced meso carbons, it can be observed that Pgb presents a high compression (0.2 Å from the reference) and Mb a minor compression. Pgb presents the three parameters in accord to a compression due to the *bre* mode, while Mb almost present the values of the reference, in agreement with the NSD.

To further explore the contribution of ruffling and breathing to the O₂ affinity, we have examined the effect of the separate distortions seen for Pgb on the binding energy in the heme model system (see Table 8). The out-of-plane distortion reduces the binding affinity, leading to a destabilization of 1.5 kcal/mol. In contrast, breathing leads to a sizable stabilization of the ligand binding energy, which clearly surpasses the destabilization due to ruffling. These changes, which agree with the trends noted above for the impact of ruffling and breathing on the binding affinity (see Figures 2 and 3 and Tables 2 and 3), indicate a net stabilization of the heme-bound O₂ compared to the ideal planar reference heme model. Notably, these trends are fully confirmed by the values determined at the PBE/6-31G**,^{45,46} OLYP/6-31G**,^{34,62,63} and B3LYP/TZVPP/QZVPP levels^{32–34} (see Table 8).

Functional Implications. Several studies have related the presence and number of hydrogen-bond interactions in the distal cavity with the O₂ binding energy and the corresponding dissociation rate.² Less information is available as to how much the oxygen affinity can be regulated by subtle changes in the intrinsic reactivity of the heme. The most representative examples of modulation of the dissociation rate by heme reactivity are leghemoglobin (LHb) and proximal site mutants of Mb, which affect the iron reactivity by modulating the charge density on the proximal histidine.⁶⁴ Previous QM/MM calculations from our group were able to reproduce the trend for wt and mutant Mb and LHb¹² showing that a $\Delta(\Delta E)$ of -1.6 kcal/mol results in a 3-fold increase in LHb k_{off} , while a $\Delta(\Delta E)$ of 1.4 kcal/mol yielded a 3-fold decrease in Mb dissociation rate.

Using these data as a reference, an increase of 3–5-fold of the dissociation rate can be predicted for the out-of-plane distortions generally observed in heme proteins. For the in-plane distortions, negative breathing will also result in an increase in k_{off} values. The only possibility of significantly decreasing k_{off} is the in-plane positive breathing, which may lead to a more than 5–10-fold decrease. Overall, the magnitude of $\Delta(\Delta E)$ due to moderate distortions is similar to those described previously for the proximal effects, and much smaller than those observed due to the presence of multiple hydrogen-bonds.²

As mentioned in the Introduction, Pgb is unusual due to both the relevant distortion of the heme and the low oxygen dissociation rate, which cannot be attributed to stabilizing hydrogen-bond interactions in the distal site. Our results point out that the observed in-plane distortion of the heme contributes to an increase in the oxygen binding affinity, which would thus agree with the available experimental data. This finding, nevertheless, raises the challenging question on how the protein

moiety can induce structural distortions in the heme. Out-of-plane distortions can be promoted by the presence of bulky residues (usually leucines or phenylalanines) perpendicularly oriented to the heme plane, which may thus selectively push the pyrrole rings out of the porphyrin plane, as suggested by site-directed mutagenesis studies in the case of nitrophorins.²⁴ Regarding the in-plane modes, it is difficult to envisage an effective way to expand the porphyrin ring through protein interactions, as required in negative breathing. However, positive breathing can be induced by forcing the porphyrin ring to fit a tighter cavity. In this scenario, evolutionary events could subtly regulate oxygen affinity by shaping the heme cavity to favor out-of-plane distortions to decrease ligand affinity, or to compress the porphyrin ring to promote the reverse effect.

Conclusions

Our results show that heme distortions are generally expected to decrease oxygen affinity. A notable exception corresponds to positive in-plane distortions for the breathing mode, which would increase oxygen affinity. Moreover, the magnitude of the stabilization due to heme distortion can be compared to the distal and proximal effects is reported for selected heme proteins. This innovative observation suggests that the deformations of the heme moiety can be understood as an inherent modulation mechanism for ligand affinity in globins. The application to Pgb of this analysis indicates the in-plane distortion of the heme, and particularly its positive breathing mode, as a key role player in determining the high oxygen affinity observed for this protein.

Acknowledgment. This work was partially supported by the University of Buenos Aires, ANPCyT, CONICET, the Spanish Ministerio de Innovación y Ciencia (SAF2008-05595 and PCI2006-A7-0688), the Generalitat de Catalunya (2009-SGR00298), and the EU FP7 program (project NOSTRESS).

Supporting Information Available: Energy profiles for remaining out-of-plane and in-plane mode distortions, *bre* distortion effect on Fe–O and O–O bonds, molecular dynamics details, and frontier orbitals. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References and Notes

- (1) *The Smallest Biomolecules: Diatomics and their Interactions with Heme Proteins*; Ghosh, A., Ed.; Elsevier Science: Amsterdam, 2008.
- (2) Martí, M. A.; Crespo, A.; Capece, L.; Boechi, L.; Bikiel, D. E.; Scherlis, D. A.; Estrin, D. A. *J. Inorg. Biochem.* **2006**, *100* (4), 761.
- (3) Scott, E. E.; Gibson, Q. H.; Olson, J. S. *J. Biol. Chem.* **2001**, *276* (7), 5177.
- (4) Bikiel, D. E.; Boechi, L.; Capece, L.; Crespo, A.; De Biase, P.; Di Lella, S.; González Lebrero, M. C.; Martí, M. A.; Nadra, A.; Perissinotti, L. L.; Scherlis, D. A.; Estrin, D. A. *Phys. Chem. Chem. Phys.* **2006**, *8* (48), 5611.
- (5) Goldberg, D. *Chem. Rev.* **1999**, *99*, 3371.
- (6) 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*, 97.
- (7) Bolognesi, M.; Bordo, D.; Rizzi, M.; Tarricone, C.; Ascenzi, P. *Prog. Biophys. Mol. Biol.* **1997**, *68*, 29.
- (8) Olson, J. S.; Phillips, G. N. *J. Biol. Chem.* **1996**, *271* (30), 17593.
- (9) Bidon-Chanal, A.; Martí, M. A.; Estrin, D. A.; Luque, F. J. *J. Am. Chem. Soc.* **2007**, *129* (21), 6782.
- (10) Boechi, L.; Martí, M. A.; Milani, M.; Bolognesi, M.; Luque, F. J.; Estrin, D. A. *Proteins* **2008**, *73* (2), 372.
- (11) 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.
- (12) Capece, L.; Martí, M. A.; Crespo, A.; Doctorovich, F. A.; Estrin, D. A. *J. Am. Chem. Soc.* **2006**, *128* (38), 12455.
- (13) Martí, M. A.; González Lebrero, M. C.; Roitberg, A. E.; Estrin, D. A. *J. Am. Chem. Soc.* **2008**, *130* (5), 1611.

- (14) Collman, J. P.; Fu, L. *Acc. Chem. Res.* **1999**, *32* (6), 455.
- (15) Slebodnick, C.; Ibers, J. A. *J. Biol. Inorg. Chem.* **1997**, *2* (4), 521.
- (16) Jentzen, W.; Simpson, M. C.; Hobbs, J. D.; Song, X. Z.; Ema, T.; Nelson, N. Y.; Medforth, C. J.; Smith, K. M.; Veyrat, M.; Mazzanti, M.; Ramasseul, R.; Marchon, J. C.; Takeuchi, T.; Goddard, W. A., III; Shelnutz, J. A. *J. Am. Chem. Soc.* **1995**, *117*, 11085.
- (17) Shelnutz, J. A.; Song, X. Z.; Ma, J. G.; Jia, S. L.; Jentzen, W.; Medforth, C. J. *J. Chem. Soc. Rev.* **1998**, *27*, 31.
- (18) Jentzen, W.; Song, X. Z.; Shelnutz, J. A. *J. Phys. Chem. B* **1997**, *101*, 1684.
- (19) Nardini, M.; Pesce, A.; Thijs, L.; Saito, J. A.; Dewilde, S.; Alam, M.; Ascenzi, P.; Coletta, M.; Ciacci, C.; Moens, L.; Bolognesi, M. *EMBO Rep.* **2008**, *9* (2), 157.
- (20) Hou, S.; Freitas, T.; Larsen, R. W.; Piatibratov, M.; Sivozhelezov, V.; Yamamoto, A.; Meleshkevitch, E. A.; Zimmer, M.; Ordal, G. W.; Alam, M. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98* (16), 9353.
- (21) 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.
- (22) Freitas, T. A. K.; Hou, S.; Dioum, E. M.; Saito, J. A.; Newhouse, J.; Gonzalez, G.; Gilles-Gonzalez, M. A.; Alam, M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (17), 6675.
- (23) Roberts, S. A.; Weichsel, A.; Qiu, Y.; Shelnutz, J. A.; Walker, F. A.; Montfort, W. R. *Biochemistry* **2001**, *40* (38), 11327.
- (24) Shokhireva, T. K.; Berry, R. E.; Uno, E.; Balfour, C. A.; Zhang, H.; Walker, F. A. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (7), 3778.
- (25) Maes, E. M.; Roberts, S. A.; Weichsel, A.; Montfort, W. R. *Biochemistry* **2005**, *44* (38), 12690. 2005.
- (26) Barkigia, K. M.; Chantranupong, L.; Smith, K. M.; Fajer, J. *J. Am. Chem. Soc.* **1998**, *110*, 7566.
- (27) Ravikanth, M.; Chandrashekar, T. K. *Coord. Chem.* **1995**, *82*, 105.
- (28) Barkigia, K. M.; Palacio, M.; Sun, Y.; Nogues, M.; Renner, M. W.; Varret, F.; Battioni, P.; Mansuy, D.; Fajer, F. *Inorg. Chem.* **2002**, *41* (22), 5647.
- (29) Jarzecki, A. A.; Spiro, T. G. *J. Phys. Chem. A* **2005**, *109* (3), 421.
- (30) Venkatesh Rao, S.; Yin, J.; Jarzecki, A. A.; Schultz, P. G.; Spiro, T. G. *J. Am. Chem. Soc.* **2004**, *126* (50), 16361.
- (31) Fernández, M. L.; Martí, M. A.; Crespo, A.; Estrin, D. A. *J. Biol. Inorg. Chem.* **2005**, *10* (6), 595.
- (32) Becke, A. J. *J. Chem. Phys.* **1993**, *98*, 1372.
- (33) Becke, A. J. *J. Chem. Phys.* **1993**, *98*, 5648.
- (34) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
- (35) Rovira, C.; Kunc, K.; Hutter, J.; Ballone, P.; Parrinello, M. *J. Phys. Chem. A* **1997**, *101*, 8914.
- (36) Scheidt, W. R.; Reed, C. A. *Chem. Rev.* **1981**, *81*, 543.
- (37) Momenteau, M.; Reed, C. A. *Chem. Rev.* **1994**, *94*, 659.
- (38) Soler, J.; Artacho, E.; Gale, J.; García, A.; Junquera, J.; Ordejón, P.; Sánchez-Portal, D. *J. Phys. Condens. Mater.* **2002**, *14*, 2745.
- (39) Crespo, A.; Scherlis, D. A.; Martí, M. A.; Ordejón, P.; Roitberg, A. E.; Estrin, D. A. *J. Phys. Chem. B* **2003**, *107*, 13728.
- (40) Scherlis, D. A.; Estrin, D. A. *Int. J. Quantum Chem.* **2001**, *87*, 158.
- (41) Martí, M. A.; Crespo, A.; Bari, S. E.; Doctorovich, F. A.; Estrin, D. A. *J. Phys. Chem. B* **2004**, *108*, 18073.
- (42) Martí, M. A.; Capece, L.; Crespo, A.; Doctorovich, F.; Estrin, D. A. *J. Am. Chem. Soc.* **2005**, *127* (21), 7721.
- (43) Martí, M. A.; Scherlis, D. A.; Doctorovich, F. A.; Ordejón, P.; Estrin, D. A. *J. Biol. Inorg. Chem.* **2003**, *8* (6), 595.
- (44) Bikiel, D. E.; Bari, S. E.; Doctorovich, F. A.; Estrin, D. A. *J. Inorg. Biochem.* **2008**, *102* (1), 70.
- (45) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *778*, 3865.
- (46) Perdew, J. P.; Burke, K.; Wang, Y. *Phys. Rev. B* **1994**, *54*, 16533.
- (47) Eichinger, M.; Tavan, P.; Hutter, J.; Ballone, P.; Parrinello, M. *J. Chem. Phys.* **1999**, *110*, 10452.
- (48) Rovira, C.; Schultze, B.; Eichinger, M.; Evanseck, J. D.; Parrinello, M. *Biophys. J.* **2001**, *81*, 435.
- (49) Crespo, A.; Martí, M. A.; Estrin, D. A.; Roitberg, A. E. *J. Am. Chem. Soc.* **2005**, *127* (19), 6940.
- (50) Lu, C.; Egawa, T.; Wainwright, L. M.; Poole, R. K.; Yeh, S. R. *J. Biol. Chem.* **2007**, *282*, 13627.
- (51) Peterson, E. S.; Huang, S.; Wang, J.; Miller, L. M.; Vidugiris, G.; Kloek, A. P.; Goldberg, D. E.; Chance, M. R.; Wittenberg, J. B.; Friedman, J. M. *Biochemistry* **1997**, *36* (42), 13110.
- (52) Mukai, M.; Savard, P. Y.; Ouellet, H.; Guertin, M.; Yeh, S. R. *Biochemistry* **2002**, *41* (12), 3897.
- (53) Angelis, F. D.; Jarzecki, A. A.; Car, R.; Spiro, T. G. *J. Phys. Chem. B* **2005**, *109* (7), 3065.
- (54) Guallar, V.; Jarzecki, A. A.; Friesner, R. A.; Spiro, T. G. *J. Am. Chem. Soc.* **2006**, *128* (16), 5427.
- (55) Spiro, T. G.; Jarzecki, A. A. *Curr. Opin. Chem. Biol.* **2001**, *5* (6), 715.
- (56) Martí, M. A.; Bikiel, D. E.; Crespo, A.; Nardini, M.; Bolognesi, M.; Estrin, D. A. *Proteins* **2005**, *62* (3), 641.
- (57) Martí, M. A.; Capece, L.; Bikiel, D. E.; Falcone, B.; Estrin, D. A. *Proteins* **2007**, *68* (2), 480.
- (58) 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*, 97.
- (59) Kundu, S.; Snyder, B.; Das, K.; Chowdhury, P.; Park, J.; Petrich, J. W.; Hargrove, M. S. *Proteins* **2002**, *46*, 268.
- (60) Couture, M.; Yeh, S.; Wittenberg, B. A.; Wittenberg, J. B.; Ouellet, Y.; Rousseau, D. L.; Guertin, M. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11223.
- (61) Kozłowski, P. M.; Bingham, J. R.; Jarzecki, A. A. *J. Phys. Chem. A* **2008**, *112*, 12781.
- (62) Handy, N. C.; Cohen, A. J. *Mol. Phys.* **2001**, *99*, 403.
- (63) Hoe, W.-M.; Cohen, A. J.; Handy, N. C. *Chem. Phys. Lett.* **2001**, *341*, 319.
- (64) Kundu, S.; Trent, J. T., III; Hargrove, M. S. *Trends Plant Sci.* **2003**, *8*, 387.

Small Molecule Migration in Methanosarcina acetivorans Protoglobin: Effect of Ligand Binding and Quaternary Structure

Flavio Forti,¹ Leonardo Boechi,² Damian Bikiel,² Marcelo A. Martí,^{2,3} Marco Nardini,⁴
Martino Bolognesi,⁴ Darío Estrin^{2*} and F. Javier Luque^{1*}

¹ Departament de Fisicoquímica and Institut de Biomedicina, Facultat de Farmàcia, Universitat de Barcelona, Av. Diagonal 643, E-08028, Barcelona, Spain

² Departamento de Química Inorgánica, Analítica, y Química Física, INQUIMAE-CONICET, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Buenos Aires, Argentina

³ Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Buenos Aires, Argentina

⁴ Dipartimento di Scienze Biomolecolari e Biotecnologie Università degli Studi di Milano, I-20131 Milano, Italy

* Send correspondence to fjluque@ub.edu or dario@qi.fcen.uba.ar

Abstract

Protoglobin is the first globin found in Archaea. Its biological role is still unknown, though this protein can bind O₂, CO and NO reversibly in vitro. A number of peculiar structural features have been identified in the X-ray structure of *Methanosarcina acetivorans* protoglobin (MaPgb). In particular, i) the heme is completely buried within the protein matrix, and ii) access of ligands to the heme distal site is granted by two orthogonal apolar tunnels, which are mainly delineated by helices G and B (tunnel 1), and helices B and E (tunnel 2). While tunnel 2 is found to be accessible to diatomic ligands in both deoxygenated and oxygenated forms of the protein, the accessibility of tunnel 1 is affected by the synergistic effect due to both the quaternary structure and the presence of the heme-bound ligand. Thus, dimerization affects the spatial arrangement of helix G (and H), which influences the shape of tunnel 1. The accessibility of ligands through this tunnel is regulated by Phe(145)G8, which can adopt two different (open/closed) conformations. Noteworthy, the ratio between open and closed states is modulated by protein dimerization and more strikingly by ligand binding. In particular, sensing of the ligand is mediated not only by Phe(145)G8, but especially Phe(93)E11. The steric hindrance between Phe(93)E11 and the heme-bound ligand alters the structural and dynamical behavior of helices B and E, and the net effect is to facilitate the opening of tunnel 1. This functional mechanism suggests that MaPgb might be involved in a ligand-controlled bimolecular chemical process.

Introduction

Globins are a family of heme-containing proteins found in all kingdoms of life. They have evolved to play a variety of biological roles, such as O₂ transport and storage, nitric oxide detoxification or sensing of diatomic gases.¹⁻³ Some of them are present as monomers under physiological conditions, but in other cases the biologically relevant form involves multimeric species, as illustrated by the prototypical cases of mammalian myoglobin (Mb) and hemoglobin (Hb), respectively.³⁻⁶ From a structural point of view, globins generally adopt a common fold characterized by a 3-over-3 helical sandwich, which contains the hydrophobic pocket that accommodates the heme group. The heme iron is coordinated to a conserved proximal HisF8,⁷ leaving the sixth coordination position in the distal side usually free for binding of the exogenous ligand.

The diversity of the globin superfamily has been disclosed by the progresses made in the last two decades, which have worked out novel hemoglobins with distinct structural and functional features. This can be exemplified by the subgroup of 2/2Hbs, which are typically smaller (around 20-30 residues shorter) than Mb and exhibit a characteristic 2-over-2 sandwich fold.^{8,9} 2/2Hbs have been proposed to act as small gas molecule sensors, oxygen carriers and pseudoenzymes. Another globin subgroup includes protoglobin (Pgb) and globin-coupled sensor (GCS) proteins, whose globin domain (~190 amino acids) is bigger than Mb.¹⁰⁻¹³ Protoglobins are the first single-domain GCS-related globins found in Archaea. They can bind O₂, CO and NO reversibly in vitro, but their function is still unknown.¹¹⁻¹³

The X-ray structure of *Methanosarcina acetivorans* protoglobin (MaPgb) has been recently reported.¹⁴ Contrary to most of the known globins, it contains 9 helices, as a pre-A helix (named Z) can be clearly distinguished. A peculiar structural feature is the large distortion of the heme moiety, which is completely buried in the protein matrix due to the occlusion of the propionate groups by the extended CE and FG loops and a 20-residue long N-terminal loop. Remarkably, access of diatomic ligands to the heme is, however, granted by an apolar V-shaped tunnel (see Figure 1), which is topologically different from the tunnels found in other globins as it consists of two branches mainly delineated by helices B/G (tunnel 1) and B/E (tunnel 2). As many other GCS proteins, MaPgb behaves as a dimer in solution and displays a 2,086 Å² association interface contributed mostly by residues belonging to the G- and H-helices (that build an intermolecular four-helix bundle), partly to the Z-helix, and to the BC and FG hinges.¹⁴

Figure 1

These unique structural properties, together with an unusually low O₂ dissociation rate (0.092-0.0094 s⁻¹), make MaPgb a relevant protein for studies aimed at gaining insight into structure-function relationships. To this end, the study of ligand association and dissociation is fundamental to explore the migration properties through the protein matrix. The association rate is mainly determined by the ligand migration, since chemical bonding to the iron is very fast.¹⁵ In turn, this process depends on the existence of transient barriers between dynamically fluctuating cavities or of gating mechanisms along tunnels.^{16,17} The dissociation rate is mainly determined by the interactions that stabilize the heme-bound ligand.¹⁵ In general, low k_{off} values can be ascribed to the presence of distal site residues that form hydrogen bonds with the ligand, thus stabilizing the bound state. After thermal breaking of the heme-bound ligand, barriers for migration through the protein matrix could also affect the dissociation rate.

In this study we explore the structural and dynamical behaviour of MaPgb in order to examine the distinctive features of ligand migration through the V-shaped tunnel delineated in the interior of the protein. In particular, attention is paid to the effect played by both dimerization and ligand binding on the migration properties. To this end, molecular dynamics simulations have been performed for monomeric and dimeric species of the protein in both bound and unbound states. The analysis of the ligand migration properties will be used to discuss the functional implications of MaPgb.

Methods

System set-up. The initial structure for molecular dynamics (MD) simulations of the monomeric and dimeric species of MaPgb in the oxygenated state was built up from the X-ray crystallographic structure of the O₂-bound protein (PDB code 2VEB).¹⁴ In turn, the ligand free crystal structure (PDB code 2VEE)¹⁴ was used to simulate the unbound form. The two X-ray structures of MaPgb are found as dimers and exhibit a similar quaternary fold.

The starting structures were immersed in a pre-equilibrated octahedral box of TIP3P water molecules.¹⁸ All systems were simulated at constant pressure (1 atm) and temperature (298K) using periodic boundary conditions and PME¹⁹ method for treating long range electrostatic interactions. The SHAKE algorithm²⁰ was used to keep bonds involving hydrogen atoms at their equilibrium length. A 1 fs time step was used for the

integration of Newton's equations. The parm99SB force field²¹ was used for the protein residues. For the heme, we adopted the parameters developed by our group and thoroughly tested in previous works,^{22,23} though slight adjustments mainly involving angles and torsions were introduced to account for the unusual high distortion of the heme found in the X-ray structures (see Table S1 in Supporting Information).

All simulations were performed with the PMEMD module of the AMBER9 program.²⁴ Equilibration protocols consisted in performing an initial optimization of the initial structures, followed by a slow heating up to the desired temperature. Energy minimization was performed in three subsequent steps where hydrogen atoms, water molecules and finally the whole system were minimized. Heating was performed in successive 50 ps runs where the temperature was gradually increased from 100 K to 298 K in four steps at constant volume, followed by an additional step run at constant pressure for 100 ps. Then, a series of 500 ns MD simulations (at 298 K and 1 atm) were run for the monomeric and dimeric protein in both deoxygenated and oxygenated states, covering a total simulation time of 2 μ s. The use of extended MD trajectories is necessary to allow the conformational rearrangements associated with the formation/breaking of the dimer into separate monomers in aqueous solution (see below). The analysis of the trajectories was performed by using frames collected every 1 ps during the production runs.

Ligand migration profiles. In order to examine the migration of ligands through the protein matrix, we have identified the preferred docking sites for the oxygen molecule and determined the migration free energy profiles along the protein tunnels that connect the solvent with the distal site. To this end, we first performed Implicit Ligand Sampling (ILS) calculations,²⁵ where the free energy of placing a probe ligand at a given position is estimated using Eq. 1.

$$G_{\text{ILS}}(r) = -k_{\text{B}}T \ln \frac{\sum_{n=1}^N \sum_{k=1}^C e^{-\Delta E_{n,k}(r)/k_{\text{B}}T}}{N C} \quad (1)$$

where N is the number of protein states, C stands for the number of equally probable ligand orientations and $\Delta E_{n,k}$ is the protein-ligand interaction energy.

ILS computations were performed using a 3D grid that encompasses the whole

protein with a 0.5 Å resolution. Moreover, 50 orientations of the probe molecule per grid point and a total set of 5000 snapshots taken every 2 ps were considered to determine the free energy at each grid point. This was repeated for different 10 ns windows along the simulations. Inspection of the free energy isosurfaces depicted on the 3D grid points allowed us to identify potential docking sites and migration paths through the protein matrix.

Since ILS estimates the free energy of binding accounting for the intrinsic thermal motion of the free protein, it implicitly assumes that the ligand interacts weakly with the protein. Moreover, it omits the effect of the dynamical interplay between thermal fluctuations of the probe and the internal cavities on the motion of the probe. To further refine the free energy profiles, constant velocity multiple Steered Molecular Dynamics (SMD) simulations were run to examine the migration of O₂ through the tunnels using Jarzynski's inequality (Eq. 2),²⁶ which relates the free energy with the irreversible work performed upon steering the probe along the reaction coordinate.

$$e^{-\Delta A(\xi)/k_B T} = \langle e^{-W(\xi)/k_B T} \rangle \quad (2)$$

where $W(\xi)$ is the external work performed on the system as it evolves from the initial to the final state computed by integrating the force acting on the steering potential along the reaction coordinate (ξ).

The steering potential forces the motion of the probe with constant velocity (v) along the reaction coordinate (Eq. 3). In the present study the reaction coordinate ξ was chosen as the iron-ligand distance, the force constant was 200 kcal mol⁻¹ Å⁻¹, and the pulling velocity was 0.025 Å ps⁻¹.

$$E(r) = k[r - (\zeta_0 + v\Delta t)]^2 \quad (3)$$

To compute the free energy profile of ligand migration along a selected tunnel, we have adopted the computational scheme used in previous studies for truncated hemoglobins by our group.^{22,23,27} Briefly, at least twenty SMD simulations were run pushing the ligand through the tunnel from the solvent towards the iron. Due to the expensiveness of SMD computations, the free energy profile was determined for only

one of the subunits. At this point, it is worth noting that similar results are expected for each of the two subunits, as noted from the high similarity between the ILS free energy isocontours determined for the two subunits of the dimeric protein (see below). The starting snapshot for each SMD was taken from the final structure of an equilibrated MD simulation run with the ligand placed at a fixed distance from the iron. Typically those distances were chosen from the preferred docking sites found during the motion of the free diatomic ligand in short MD simulations.

Results and Discussion

Overall structural analysis of MaPgb. Extended MD simulations were run to explore the structural integrity and flexibility of MaPgb. This is necessary to capture not only the fluctuations that modulate the shape and size of cavities and tunnels, but more importantly to examine the differences found for monomeric and dimeric forms of MaPgb, as well as the effect played by the presence of the ligand (O_2) bound to the heme. Accordingly, a series of 500 ns MD simulations were run for deoxygenated ligand-free (Pgb-deoxy) and oxygenated (Pgb- O_2) forms of the protein treated either as a monomer (MPgb) or a dimer (DPgb).

For all the simulations the time evolution of the rmsd (determined relative to the X-ray structure; see Figure S1 in Supporting Information) was computed for the backbone atoms of the protein core (residues 23-189), thus neglecting the contributions due to N- and C-terminal tails, which showed higher mobility compared to the core of the protein. The rmsd profile for DPgb-deoxy is stable at around 1.5 Å after the first 50 ns. A slightly lower value (rmsd of 1.3 Å) is obtained for DPgb- O_2 , though this value is achieved after a progressive increase in the rmsd profile during the first 220 ns. The rmsd determined for the monomers is larger than the value calculated for the dimeric forms, as it amounts to 1.8 and 2.2 Å for MPgb-deoxy and MPgb- O_2 at the end of the trajectories. This finding is not unexpected, as the two X-ray crystallographic structures (PDB codes 2VEB and 2VEE) used to build up the simulated systems correspond to dimeric forms of the protein.

The preceding results suggest that dissociation of the dimer into separate monomers does not promote a substantial alteration in the overall fold of the protein. However, the larger rmsd found for the monomers reveals the occurrence of local rearrangements in certain structural elements upon dissociation of the dimer. At this point, a detailed

structural analysis of the distinct helical segments showed that there is a relevant displacement of helix G (and at less extent helix H) in MPgb-deoxy compared to the X-ray crystallographic structure (PDB entry 2VEE; Figure 2). Thus, whereas the rmsd of helix G amounts to ~ 1.2 Å in the dimer (a value comparable with the rmsd of other helices), the rmsd is enhanced to near 2.9 Å in the monomer (Figure 2). Since helix G is directly involved in defining the interface between the interacting units in the dimeric form of MaPgb, dissociation of the dimer makes helix G to be exposed to the aqueous solvent, thus facilitating the conformational relaxation in the monomer compared to the dimer. Noteworthy, helix G, in conjunction with helix B, delineates the shape of tunnel 1 (see Figure 1). Therefore, the structural change observed for helix G might have a direct effect on the accessibility of ligands through this tunnel (see below).

Figure 2

Strikingly, when the same analysis is performed for the oxygenated forms of the protein in both monomeric and dimeric forms, the largest rmsd (determined relative to the X-ray crystallographic structure; PDB entry 2VEB) is mainly found for helices B, E and G (Figure 3). In DPgb-O₂ the structural rearrangement primarily affects helix B, whose rmsd (around 2.2 Å) is sensibly larger than that determined for helices E and G (around 1.6 Å). Nevertheless, in MPgb-O₂ the structural rearrangement primarily affects helices E (rmsd close to 2.8 Å) and G (around 2.2 Å). According to the analysis made for the deoxygenated monomer and dimer (see above and Figure 2), the conformational relaxation of helix G is not unexpected and can be explained by the dissociation of the dimer. However, present results suggest that the heme-bound ligand exerts additional effects, which primarily influence the structural properties of helices B and E (see Figure 3 bottom). It must be stressed that these two helices define the walls of tunnel 2. Therefore, these findings raise questions about the potential role played by the heme-bound ligand on the accessibility of the tunnels.

Figure 3

It is worth noting that the structural fluctuation of helices B and E in the oxygenated protein is associated with a drastic change in the dynamical behavior of the protein triggered upon ligand binding, as noted in the essential motions of the protein skeleton determined by diagonalizing the positional covariance matrix for the backbone atoms of the protein core.^{28,29} The calculated eigenvalues decrease in magnitude smoothly, thus indicating that a variety of motions determine the structural flexibility of the protein backbone. However, whereas the first ten principal components account for around 50%

of the structural variance, the first eigenvalue alone describes between 13% and 19% of the structural fluctuations. The nature of this single essential mode is drastically different in both deoxygenated and oxygenated forms of the protein (Figure 4). In the former case, it mainly involves displacements of helix A and those segments of helices G and H close to helix A. In contrast, the first essential mode for the heme-bound O₂ species mainly involves the displacement of helices B and E. Finally, since helix B contributes to define the shape of tunnels 1 and 2, the increased flexibility of this helix might be relevant to facilitate the migration of ligands to/from the heme cavity.

Figure 4

Based on these findings, we can conclude that both dissociation of the dimer and the presence of the heme-bound ligand promote distinctive structural and dynamical alterations in helices B, E and G of MaPgb. Because those helical segments delineate the V-shaped tunnel, it can be expected that the migration properties are likely affected by both the quaternary structure of the protein and the binding of ligands to the heme cavity. In turn, two main questions arise: i) how the presence of the ligand can be sensed by the protein and affect the structural and dynamical behavior of helices B and E, and ii) how the helical alterations triggered by both quaternary structure and ligand binding influence the accessibility of ligands through the tunnel.

Structural changes in the distal cavity. To answer these questions, we have first analyzed the structural changes induced in the distal cavity upon dissociation of the dimer and upon binding of O₂ to the heme.

The binding of ligands to the heme is generally sensed via formation of hydrogen bonds in the heme cavity. Hydrogen-bonding with distal residues not only contributes to stabilize the heme-bound ligand, but can also promote structural and dynamical changes in the protein as, for instance, reported for the 2/2HbN in *M. tuberculosis* (MtHbN), where the alteration of the hydrogen-bonding network formed by polar residues upon O₂ binding to the heme changed the main essential motions of the protein skeleton.^{30,31} Inspection of the X-ray structure of MaPgb, however, shows that there is no hydrogen bonding stabilization of the heme-bound O₂ in the distal cavity. The most feasible residue that could stabilize the ligand is Tyr(61)B10, but its hydroxyl group is too far from the heme-bound O₂ in PDB entry 2VEB (distance OH(TyrB10)---O(O₂) of 5.26 Å), and no water molecules were found to mediate the interaction between Tyr(61)B10 and O₂. Nevertheless, thermal fluctuations could facilitate the local reorientation of the

phenol group to enable hydrogen bonding with the ligand.

The potential involvement of distal hydrogen bonds in assisting the O₂ binding was examined from the time evolution of the distance between the hydroxyl oxygen of TyrB10 and O₂ in the oxygenated forms of both monomeric and dimeric MaPgb (Figure 5). Though there are few attempts to form a hydrogen bond with the heme-bound O₂, the distance OH(TyrB10)---O₂(O₂) is generally close to 6 Å. In fact, most of the time Tyr(61)B10 is hydrogen bonded to the carbonyl oxygen of Leu(86)E4. Similarly, the analysis of the snapshots ruled out a significant stabilization due to water-mediated bridges with Tyr(61)B10. Overall, present results exclude a significant contribution of Tyr(61)B10 in stabilizing the ligand. In turn, this finding supports the idea that the very low O₂ dissociation rate found in MaPgb has to be ascribed mostly to the strong distortion of the Pgb heme, as suggested by recent *in silico* analysis.³²

Figure 5

The absence of hydrogen bonding with the heme-bound O₂ prompted us to search for contacts between the ligand and vicinal residues in the distal cavity. To this end, we have analyzed the distribution of distances from the iron in order to adopt a common frame for both deoxygenated and oxygenated forms of the protein. On the basis of this analysis, two aromatic residues, Phe(145)G8 and Phe(93)E11 (see Figure S2 in Supporting Information), emerged as putative candidates in sensing the presence of the ligand in MaPgb. At this point, it is worth noting that in the X-ray structure of oxygenated MaPgb Phe(93)E11 was found to be the closest residue to O₂ (at 3 Å), with its side chain affected by conformational disorder. On the contrary, Phe(E11)93 adopts a defined conformation in the absence of the heme ligand.¹⁴

Regarding Phe(145)G8, the average distance determined from the benzene ring to the iron is 4.9 Å for MPgb-deoxy. This value is notably smaller than that found in the X-ray structures, as the closest distance from the side chain of Phe(145)G8 to the iron is close to 6 Å. Therefore, the smaller distance found for MPgb-deoxy can be attributed to the structural rearrangement of helix G after dissociation of the dimer. More importantly, the monomeric species is sensitive to the heme-bound O₂, as noted in the enlargement of the average distances determined from the benzene ring to the iron (from 4.9 and 6.0 Å for MPgb-deoxy and MPgb-O₂ forms; see Figure 6). In contrast, such sensitivity is slightly reduced in the dimeric species, as noted in the similar range of average distances (from 6.4 to 6.7 Å) determined for DPgb-deoxy and DPgb-O₂ (Figure 6). Clearly, the interaction between monomers in the dimer and the associated

rearrangement of helices G and H dominates the location of Phe(145)G8 in the distal cavity, and the friction those helices at the interface between subunits reduces the sensitivity of Phe(145)G8 relative to the isolated monomer.

Figure 6

Contrary to Phe(145)G8, the sensitivity of Phe(93)E11 to the heme-bound ligand is consistently observed in both monomeric and dimeric forms of MaPgb (Figure 6). In both MPgb-deoxy and DPgb-deoxy the distance of the Phe(93)E11 side chain to the iron, which ranges on average from 4.0 to 4.5 Å, increases upon binding of O₂. Such an increase amounts to around 0.5 Å when the monomeric forms are considered, but it is enlarged up to 1.1 Å in the case of the dimeric species. Thus, irrespective of the quaternary structure of the protein, the presence of the heme-bound ligand introduces steric hindrance onto the side chain of Phe(93)E11.

Overall, these findings suggest that the ligand sterically collides with the benzene ring of Phe(93)E11 and contributes not only to the structural change observed in helices B and E (Figure 3) observed upon ligand binding, but also to the alteration in the dynamical motion of these helical segments (Figure 4). Notably, these trends resemble the large-scale conformational change detected both theoretically³⁰ and experimentally³¹ in MtHbN. However, whereas in this latter protein the binding is sensed through alteration in the network of hydrogen bonds formed by TyrB10 and GlnE11 in the distal cavity, steric hindrance appears to mediate the ligand sensing properties in MaPgb.

Ligand migration. In order to explore the impact of the structural and dynamical rearrangements on the ligand migration properties of MaPgb, we have analyzed the structural and energetic features of tunnels 1 and 2.

The accessibility of ligands through tunnel 1 seems to be regulated by the side chain of Phe(145)G8, which is found in two main conformations along the MD trajectories (see below). Those conformations can be characterized by the dihedral angle N-C_α-C_β-C_γ, which adopts values centered in 270 and 210 degrees. The former value reflects the conformation found in the X-ray crystallographic structures (267.2 degrees in 2VEB, and varying from 264.7 to 270.2 degrees in the 8 independent protein molecules present in the asymmetric unit of 2VEE). In this conformational state migration of a diatomic ligand through tunnel 1 seems feasible. Nevertheless, in the other conformation the benzene ring protrudes into the tunnel, and it would therefore impede the migration of ligands. The impact of the conformational state adopted by Phe(145)G8 on the

accessibility of diatomic ligands through tunnel 1 is clearly reflected in the free energy isosurfaces derived from ILS computations (Figure 7; see also Figure S3 in Supporting Information). In the following, these two conformations will be denoted “open” and “closed”, respectively.

Figure 7

Noteworthy, the relative population of the two conformations of Phe(145)G8 depend on both the quaternary structure and the presence of the ligand in the distal cavity (Figure 8; see also Figure S4 in Supporting Information). Thus, the Phe(145)G8 side chain remains in the closed state during the last 250 ns of the trajectories collected for both the deoxygenated (99.9%) and oxygenated (98%) states of the monomeric protein. Dimerization of the protein promotes a slight increase in the population of the open state, which is found to populate between 5% and 8% in the two subunits of DPgb-deoxy. Remarkably, this trend is significantly enhanced upon binding of O₂ to the distal cavity. Thus, around 25% of the structures sampled in the last 250 ns of the trajectory are found in the open state. On the basis of the preceding results, it can be concluded that (i) binding of O₂ does not suffice to enhance the accessibility through tunnel 1 in MPgb, and (ii) that dimerization is necessary to facilitate the opening of tunnel 1. Finally, it is also worth noting that the enhanced conformational fluctuations triggered on helix B by the heme-bound O₂ (see above) should also facilitate ligand migration through tunnel 1 (see above).

Figure 8

In contrast to the preceding results, tunnel 2 is found to be accessible irrespective of the presence of heme-bound ligand and the quaternary structure of the protein, as noted in the ILS profiles shown in Figure 7 (see also Figure S3 in Supporting Information). This finding suggests that MaPgb might be involved in a ligand-controlled bimolecular chemical process. Accordingly, a small ligand could access the heme cavity through tunnel 2. However, if this would happen in MPgb, tunnel 1 would primarily remain in a closed state (Figure 8). In contrast, dimerization of the protein facilitates the transition of tunnel 1 to the open state, an effect that could be associated with the increased flexibility of helix B triggered by the presence of the heme-bound ligand (see Figures 3 and 4). According to these mechanistic events, it can be hypothesized that the ligand-assisted dynamical modulation of the accessibility through tunnel 1 should be reflected in a higher sensitivity of MaPgb to another incoming ligands.

To further check the role of the quaternary structure, SMD computations were

performed to determine the Potential of Mean Force (PMF) for the migration of O₂ through tunnels 1 and 2 using Jarzinsky's inequality for MPgb-deoxy and DPgb-deoxy (Figure 9). The PMF profile determined for MPgb-deoxy shows the existence of a high and wide barrier (around 6 kcal/mol) at around 12 Å from the iron atom. In contrast, the PMF profile determined for DPgb-deoxy reveals a drastic alteration in the PMF profile along tunnel 1, as the ligand must surpass a barrier of only 2.2 kcal/mol at around 11 Å from the iron. Then, there is a rather flat region between 9 and 6 Å, this latter distance corresponding to the entrance to the distal cavity. Finally, access to the heme is limited by a small barrier (around 2 kcal/mol).

Figure 9

The PMF profile obtained for migration through tunnel 2 in MPgb-deoxy is flatter than that obtained for tunnel 1, with a minimum at around 7 Å from the heme and small barriers (around 1 kcal/mol) that should be easily surpassed due to thermal fluctuations. On the other hand, the most noticeable feature in the PFM profile derived for DPgb-deoxy is the appearance of a free energy minimum at around 5 Å, which can be accessed from the bulk solvent through a barrierless migration.

Overall, SMD computations demonstrate that, even in the absence of heme-bound O₂, dimerization has a large effect in regulating the migration of ligands through tunnels 1 and 2, which should thus lead to an enhanced accessibility in the dimer compared to the isolated monomer.

Functional implications. Inspection of the X-ray crystallographic structures available for oxygenated (2VEB) and deoxygenated (2VEE) MaPgb revealed a number of relevant structural features that are absent in prototypical globins such as Mb,¹⁴ including the unusual V-shaped tunnel leading from the heme cavity to the aqueous solvent. It can be anticipated that the two branches of the tunnel are expected to be functionally relevant since the heme group is completely buried in the interior of MaPgb. Therefore, even though the biological role of MaPgb in Archaea is still unknown, its biochemical function is likely related to the ability of the branched tunnel to permit the migration of specific ligands. Therefore, understanding the factors that regulate the accessibility of ligands can be of utmost importance to shed light into the biological function of Pgb. On the other hand, since gel filtration experiments have shown that the protein exists as a dimer in solution,¹⁴ it can be argued whether dimerization has a direct implication on the migration of ligands through the V-shaped tunnel and is thus relevant

for the physiological role of the protein.

In the absence of robust evidences about the biological role, the peculiar biochemical behavior of MaPgb has been clearly established by kinetic measurements of the migration of O₂, which revealed a very low O₂ dissociation rate (0.092-0.0094 s⁻¹). This is unusual, since inspection of the X-ray structure for the protein with heme-bound O₂ shows that the ligand is not stabilized by hydrogen bonds with distal residues. For the sake of comparison, the low k_{off} values of 0.004 s⁻¹ and 0.2 s⁻¹ determined for *Ascaris* Hb and MtHbN can be related to the stabilization of the heme-bound ligand by two hydrogen-bond donors with TyrB10 and GlnE7 in AscHb, and with TyrB10 and GlnE11 in MtHbN).^{33,34} In contrast, HisE7 is the only hydrogen-bond donor in Mb, this fact being reflected in a higher k_{off} (12 s⁻¹).³³ An even larger value (1600 s⁻¹) has been reported for the HisE7Val mutant.³⁵

The lack of hydrogen bonding interactions that stabilize the bound O₂ in the MaPgb distal cavity is confirmed by the results derived from the extended MD simulations run for the oxygenated dimer. In this case, very transient hydrogen bond interactions are formed with Tyr(61)B10, which is mainly hydrogen-bonded to the carbonyl oxygen of LeuE4. On the other hand, the low k_{off} value cannot be attributed to the occlusion of the tunnels. In fact, tunnel 2 is permanently open along the trajectory of the dimer. Regarding tunnel 1, there is a balance between open and closed states, which in our simulations are defined according to the conformation adopted by Phe(145)G8, and there is a small barrier for egression of O₂. Therefore, ligand escape from the distal cavity cannot be considered to make a significant contribution to the dissociation rate. In turn, the preceding discussion reinforces the potential role played by the remarkable distortion of the heme in MaPgb on the affinity for ligands, as recently pointed out by our previous studies.³² How MaPgb can induce such a structural heme distortion is a challenging question that remains to be answered.

Another relevant finding arising from present simulations is the effect played by both dimerization and O₂ binding on the accessibility through tunnel 1, which is regulated by a balance between open and closed states defined according to the conformation of the Phe(145)G8 side chain. In the deoxygenated dimer the closed form is the major conformational state of tunnel 1. However, there is a significant increase in the population of the open state (which populates around 25%) upon O₂ binding to the heme. This effect reflects the net balance of three factors. First, the direct influence exerted by dimerization on the structural arrangement of helix G (and at less extent

helix H), which contributes to the interface of the interacting monomers. Moreover, in conjunction with helix B, helix G delineates the walls of tunnel 1. Second, the structural and dynamical alteration of helix B upon ligand binding, which is sterically sensed by Phe(93)E11. Finally, the steric destabilization of the closed conformation of Phe(145)G8 by the heme-bound O₂.

On the basis of the preceding discussion, it can be hypothesized that MaPgb might be involved in a ligand-controlled bimolecular chemical process, where loading of the protein with a first ligand would facilitate the entry of the second reagent to the heme cavity. Noteworthy, this hypothesis is supported by recent experimental findings derived from kinetic studies of CO binding to MaPgb which indicate that i) the protein exists in two states characterized by fast and slow rebinding conformations, and that ii) CO ligation favours the fast rebinding conformation.³⁶ On the basis of present results, those fast and slow rebinding conformations can be interpreted in terms of the open and closed conformational states of Phe(145)G8, which regulates the access through this tunnel, and by the ligand sensing properties of Phe(93)E11, which contributes to facilitate the transition from closed to open states.

In the absence of clear evidences about the nature of the true fisiological ligands involved in the biological function of MaPgb, it is also reasonable to expect that a larger ligand (i.e., a triatomic molecule) could lead to a more effective opening of tunnel 1 due to enhanced steric hindrance with the sensing Phe residues. Accordingly, the balance between open and closed states might be regulated by the size/polarity of the ligand, so that opening of tunnel 1 might be facilitated upon binding of an incoming compound through tunnel 2, thus paving the way for the involvement of MaPgb in mediating a bimolecular chemical process. Finally, present results indicate that dimerization might represent a mechanism to regulate the functional role of MaPgb, with the quaternary structure being considered as a necessary step in order to form the biologically relevant form of the protein.

Conclusions. The structural analysis of the MD simulations run for deoxy- and oxygenated states of monomeric and dimeric MaPgb, and the energetic characterization of the tunnels identified in the X-ray structure reveals a number of factors potentially important for the biological role of the protein. First, the absence of stabilizing interactions between the heme-bound O₂ and distal residues reinforces the hypothesis that heme structural distortion (particularly the in-plane breathing deformation) is

playing a key role on the O₂ affinity, which is mainly due to an unusually low dissociation rate in MaPgb. Moreover, the accessibility of ligands through tunnel 1 seems to be regulated by the side chain of Phe(145)G8, which in our simulations assumes two different conformations. Noteworthy, the balance between open and closed states is affected by the synergistic effect of two main factors: i) the quaternary structure of the protein, and ii) the presence of the heme-bound ligand. Thus, our results demonstrate that dimerization of the protein has a profound effect on the shape of the free energy profile for ligand migration through tunnel 1, which is further facilitated by the increased flexibility of helix B triggered upon binding of O₂ to the heme. In contrast, binding of O₂ to a hypothetical monomeric form of the protein is found to have minor impact on the accessibility of tunnel 1. Accordingly, these findings support the biological role of the dimeric form of MaPgb. Based on these results, it might be hypothesized that MaPgb is involved in a ligand-regulated bimolecular process, where binding of a ligand accessing the heme cavity through tunnel 2 would facilitate the migration of a second reagent through tunnel 1, thus leading to an overall increase in the reactivity of the protein.

Acknowledgements

Prof. C. Viappiani is kindly acknowledged for sharing his experimental results of ligand binding and for insightful discussions. This work was partially supported by the University of Buenos Aires, ANPCyT, CONICET, the Spanish Ministerio de Innovación y Ciencia (SAF2008-05595 and PCI2006-A7-0688), the Generalitat de Catalunya (2009-SGR00298), and the EU FP7 program (project NOstress). FF is fellowship from the Spanish Ministerio de Innovación y Ciencia. Computational facilities provided by the Barcelona Supercomputer Center and the Centre de Supercomputació de Catalunya are acknowledged.

Supporting Information Available. Specific parameters for the heme in Pgb, conformational preferences of sensing residues Phe(93)E11 and Phe(145)G8, and free energy isosurfaces used to delineate tunnels 1 and 2 in Pgb. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

1. The smallest biomolecules, Diatomics and their interactions with heme proteins. Ghosh A (Ed). Elsevier. Amsterdam. 2007.
2. Paoli M, Marles-Wright J, Smith A. Structure-function relationships in heme-proteins. *DNA Cell Biol* **2002**, 21, 271-280.
3. Vinogradov SNM, L. Diversity of globin function, enzymatic, transport, storage, and sensing. *J Biol Chem* **2008**, 283, 8773-8777.
4. Frauenfelder H, McMahon BH, Fenimore, PW. Myoglobin, The hydrogen atom of biology and a paradigm of complexity. *Proc Natl Acad Sci USA* **2003**, 100, 8615-8617.
5. Scott EE, Gibson QH, Olson JS. Mapping the pathways for O₂ entry into and exit from myoglobin. *J Biol Chem* **2001**, 276, 5177-5188.
6. Wittenberg JB, Wittenberg BA. Myoglobin function reassessed. *J Exp Biol* **2003**, 206, 2011-2020.
7. Kapp OH, Moens L, Vanfleteren J, Trotman CN., Suzuki T, Vinogradov SN. Alignment of 700 globin sequences, extent of amino acid substitution and its correlation with variation in volume. *Protein Sci* **1995**, 4, 2179–2190.
8. Nardini M, Pesce A, Milani M, Bolognesi M. Protein fold and structure in the truncated (2/2) globin family. *Gene* **2007**, 398, 2-11.
9. Wittenberg JB, Bolognesi M, Wittenberg BA, Guertin M. Truncated hemoglobins, a new family of hemoglobins widely distributed in bacteria, unicellular eukaryotes, and plants. *J Biol Chem* **2002**, 277, 871-874.
10. Freitas TAK, Hou S, Dioum EM, Saito JA, Newhouse J, Gonzalez G, Gilles-Gonzalez M-A, Alam M. Ancestral hemoglobins in Archaea. *Proc Natl Acad Sci USA* **2004**, 101, 6675-6680.
11. Hou S, Freitas T, Larsen RW, Piatibratov M, Sivozhelezov V, Yamamoto A, Meleshkevitch EA, Zimmer M, Ordal GW, Alam M. Globin-coupled sensors, a class of heme-containing sensors in Archaea and Bacteria. *Proc Nat Acad Sci USA* **2001**, 98, 9353-9358.
12. Vinogradov S, Hoogewijs D, Bailly X, Arredondo-Peter R, Gough J, Dewilde S, Moens L, Vanfleteren J. A phylogenomic profile of globins. *BMC Evol Biol* **2006**, 6, 31-47.
13. Vinogradov SN, Hoogewijs D, Bailly X, Mizuguchi K, Dewilde S, Moens L, Vanfleteren JR. A model of globin evolution. *Gene* **2007**, 398,132-142.
14. Nardini M, Thijs L, Saito JA, Dewilde S, Alam M, Ascenzi P, Coletta M, Ciaccio C, Moens L, Bolognesi M. Archaeal protoglobin structure indicates new ligand diffusion paths and modulation of haem-reactivity. *EMBO Rep* **2008**, 9, 157-163.
15. Marti MA, Crespo A, Capece L, Boechi L, Bikiel DE, Scherlis DA, Estrin DA. Dioxygen affinity in heme proteins investigated by computer simulation. *J Inorg Biochem* **2006**, 100, 761-770.
16. Ruscio JZ, Kumar D, Shukla M, Prisant MG, Murali TM, Onufriev AV. Atomic level computational identification of ligand migration pathways between solvent and binding site in myoglobin. *Proc Natl Acad Sci USA* **2008**, 105, 9204-9209.
17. Tomita A, Sato T, Ichiyangi K, Nozawa S, Ichikawa H, Chollet M, Kawai, F, Park S-Y, Tsuduki T, Yamato T, Koshihara S-Y, Adaci S-I. Visualizing breathing motion of internal cavities in concert with ligand migration in myoglobin. *Proc Natl Acad Sci USA* **2009**, 106, 21612-22616.
18. Jorgensen W, Chandrasekhar J, Madura J, Impey R, Klein M. Comparison of

- simple potential functions for simulating liquid water. *J Chem Phys* **1983**, 79, 926-935.
19. Darden T, York D, Pederson L. Particle Mesh Ewald, An $N \cdot \log(N)$ method for Ewald sums in large systems. *J Chem Phys* **1993**, 98, 10089-10092.
 20. Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints, molecular dynamics of n-alkanes. *J Comput Phys* **1977**, 23, 327-341.
 21. Hornak V, Abel, R, Okur, A, Strockbine, B, Roitberg, A, Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **2006**, 65, 712-725.
 22. Boechi L, Martì MA, Milani M, Bolognesi M, Luque FJ, Estrin DA. Structural determinants of ligand migration in Mycobacterium tuberculosis truncated hemoglobin O. *Proteins* **2008**, 73, 372-379.
 23. Boechi L, Manez PA, Luque FJ, Martì MA, Estrin DA. Unraveling the molecular basis for ligand binding in truncated hemoglobins, the trHbO Bacillus subtilis case. *Proteins* **2009**, 78, 962-970.
 24. Case D, Darden TA, Cheatham TE, III, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Pearlman DA, Crowley M, Walker RC, Zhang W, Wang B, Hayik S, Roitberg A, Seabra G, Wong KF, Paesani F, Wu X, Brozell S, Tsui V, Gohlke H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Beroza P, Mathews DH, Schafmeister C, Ross WS, Kollman PA. AMBER, version 9. University of California, San Francisco, CA, 2006.
 25. Cohen J, Arkhipov A, Braun R, Schulten K. Imaging the migration pathways for O₂, CO, NO, and Xe inside myoglobin. *Biophys J* **2006**, 91, 1844-1857.
 26. Jarzynski C. Nonequilibrium equality for free energy differences. *Phys Rev Lett* **1997**, 78, 2690-2693.
 27. Martí MA, Bidon-Chanal A, Crespo A, Yeh S-R, Guallar V, Luque FJ, Estrin DA. Mechanism of product release in NO detoxification from Mycobacterium tuberculosis truncated hemoglobin N. *J Am Chem Soc* **2008**, 130, 1688-1693.
 28. García A. Large-Amplitude Nonlinear Motions in Proteins. *Phys Rev Lett* **1992**, 68, 2696-2699.
 29. Amadei A, Linssen, AB, Berendsen, HJ. Essential dynamics of proteins. *Proteins* **1993**, 17, 412-425.
 30. Crespo A, Martí MA, Kalko SG, Morreale A, Orozco M, Gelpi JL, Luque FJ, Estrin DA. Theoretical study of the truncated hemoglobin HbN, Exploring the molecular basis of the NO detoxification mechanism. *J Am Chem Soc* **2005**, 127, 4433-4444.
 31. Mukai MO, Ouellet Y, Ouellet H, Guertin M, Yeh SR. NO binding induced conformational changes in a truncated hemoglobin from Mycobacterium tuberculosis. *Biochemistry* **2004**, 43, 2764-2770.
 32. Bikiel DE, Boechi L, Nardini M, Luque FJ, Martí MA, Estrin DA. Role of heme distortion on oxygen affinity in heme proteins, the protoglobin case. *J Phys Chem B* **2010**, 114, 8536-8543.
 33. Milani M, Pesce A, Nardini M, Ouellet H, Ouellet Y, Dewilde S, Bocedi A, Ascenzi P, Guertin M, Moens L, Friedman JM, Wittenberg JB, Bolognesi M. Structural bases for heme binding and diatomic ligand recognition in truncated hemoglobins. *J Inorg Biochem* **2005**, 99, 97-109.
 34. Goldberg DE. Oxygen-avid hemoglobin of Ascaris. *Chem Rev* **1999**, 99, 3371-3378.
 35. Kundu S, Snyder B, Das K, Chowdhury P, Park J, Petrich JW, Hargrove MS.

The leghemoglobin proximal heme pocket directs oxygen dissociation and stabilizes bound heme. *Proteins* **2002**, 46, 268-277.

36. Viappiani, C. (personal communication)

Figure 1. View of the V-shaped tunnel observed in MaPgb. The tunnel is composed of two branches, which are mainly delineated by helices B/G (tunnel 1) and B/E (tunnel 2). The heme group is shown as sticks.

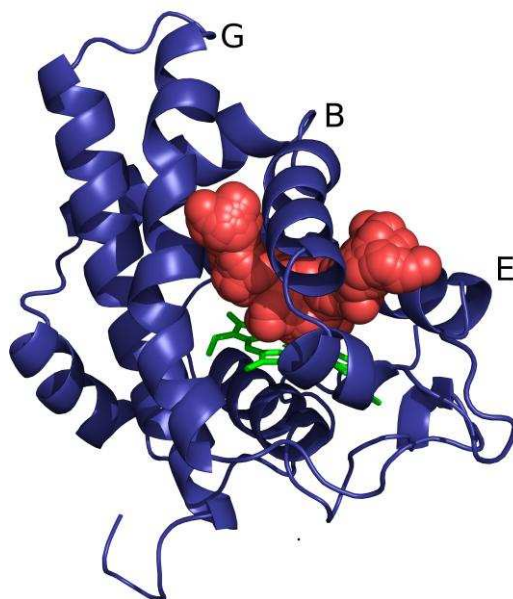


Figure 2. Time (ns) evolution of the rmsd (\AA) of the backbone atoms in selected helical segments of the deoxygenated (top) monomer and (middle) dimer determined relative to the X-ray structure (PDB entry 2VEE). (Bottom) Superposition of helices B, G and H in the final snapshots of simulations run for monomer (green) and dimer (blue) in the deoxygenated state. Phe(145)G8 and heme are shown as red and atom-coloured sticks.

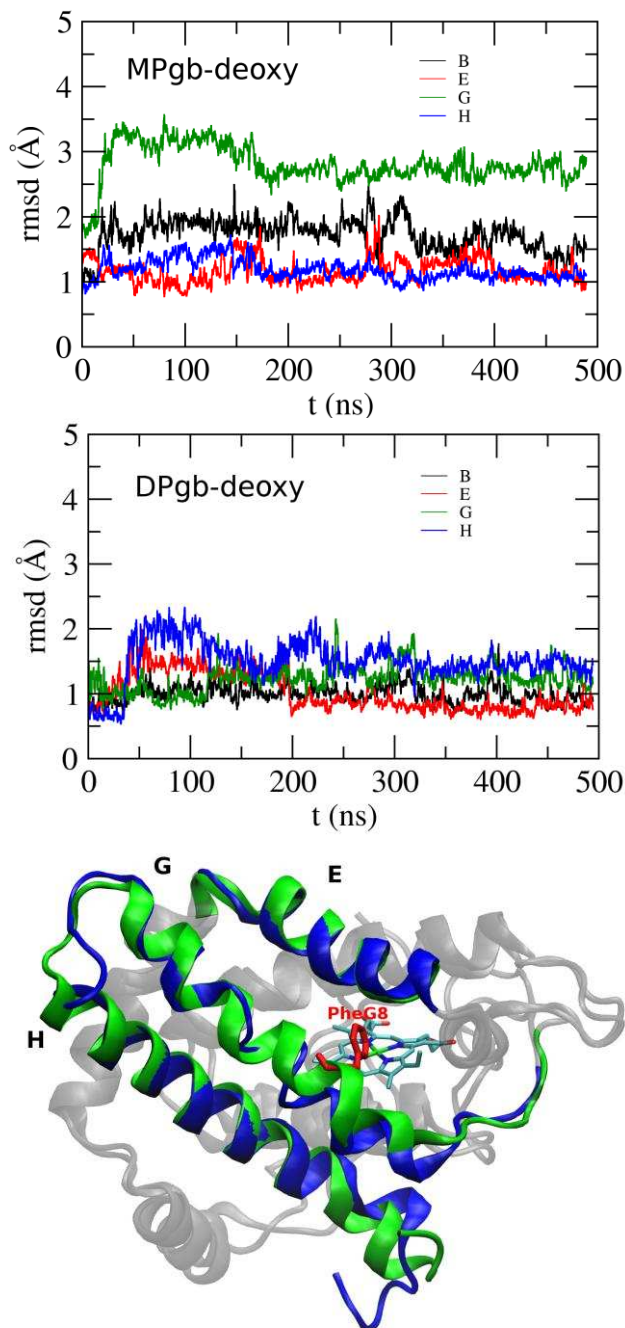


Figure 3. Time (ns) evolution of the rmsd (\AA) of the backbone atoms in selected helical segments of the oxygenated (top) monomer and (middle) dimer determined relative to the X-ray structure (PDB entry 2VEB). (Bottom) Superposition of helices G, B and E in the final snapshots of simulations run for deoxygenated (blue) and oxygenated (cyan) states of the dimer. Phe(93)E11 and heme are shown as red and atom-coloured sticks.

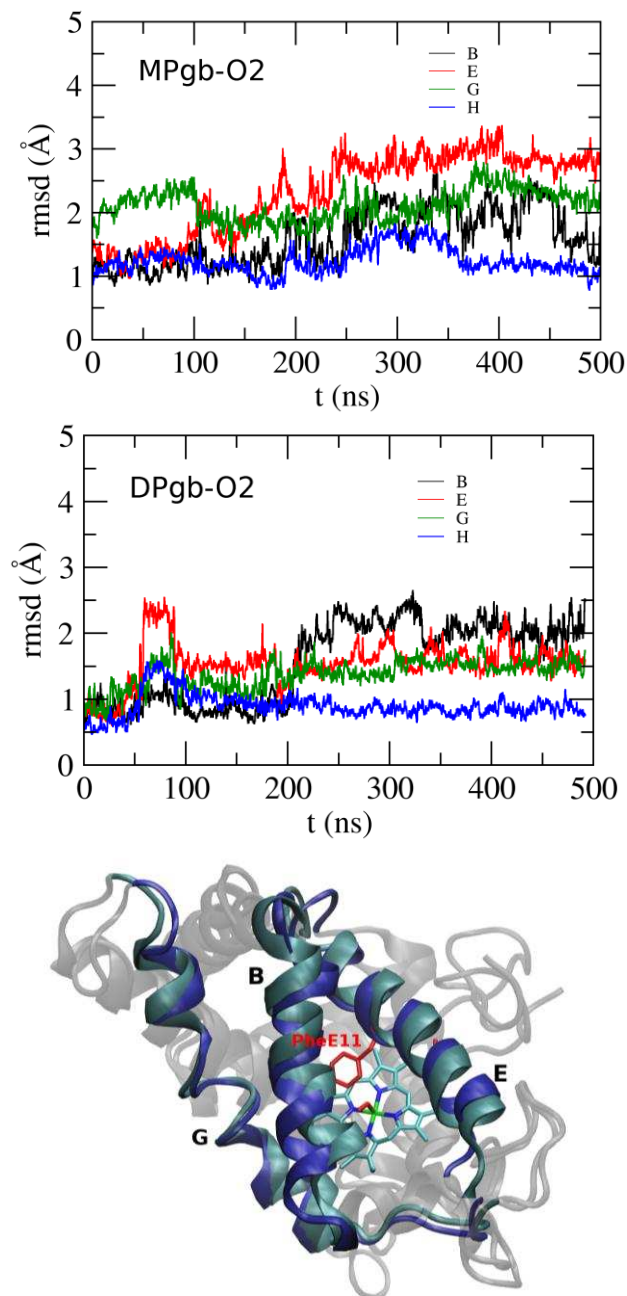


Figure 4. Representation of pairs of snapshots that represent the deformations of the protein backbone due to the first essential mode for (A) MPgb-O₂, (B) DPgb-O₂, (C) MPgb-deoxy, and (D) DPgb-deoxy forms of MaPgb.

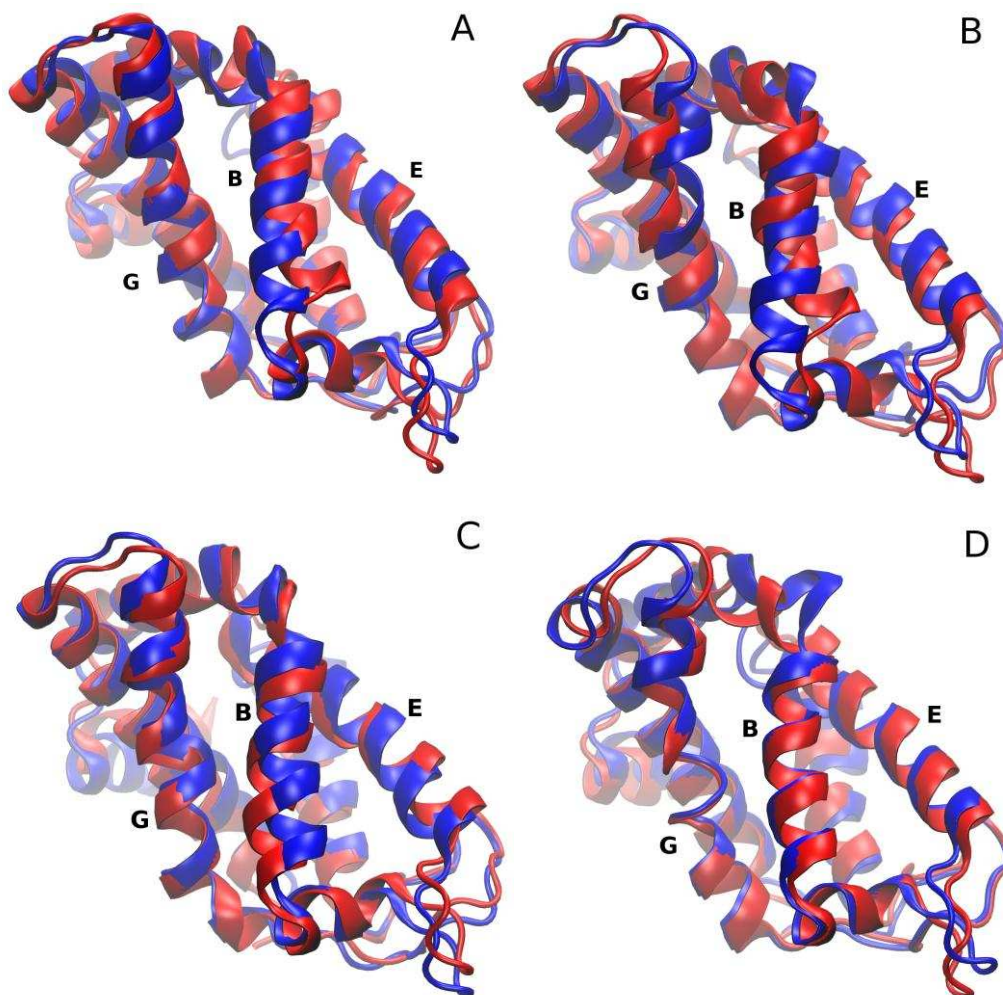


Figure 5. Time (ns) evolution of OH(TyrB10)---O(O₂) (black) and OH(TyrB10)---O(LeuE4) (grey) distances (Å) for the oxygenated forms of (left) monomeric and (right) dimeric MaPgb. For the dimer only the distances determined for subunit A are shown. Similar results stand for subunit B.

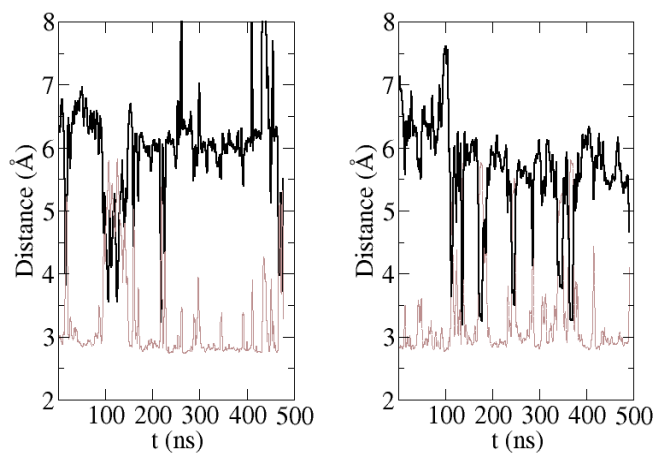


Figure 6. Distribution of closest distances between the side chain of (left) Phe(145)G8 and (right) Phe(93)E11 from the heme iron. The distances are determined for both (top) monomeric (MPgb) and (bottom) dimeric (DPgb) species in deoxygenated and oxygenated states. For the dimers the values determined for the two interacting subunits are denoted with “1” and “2”.

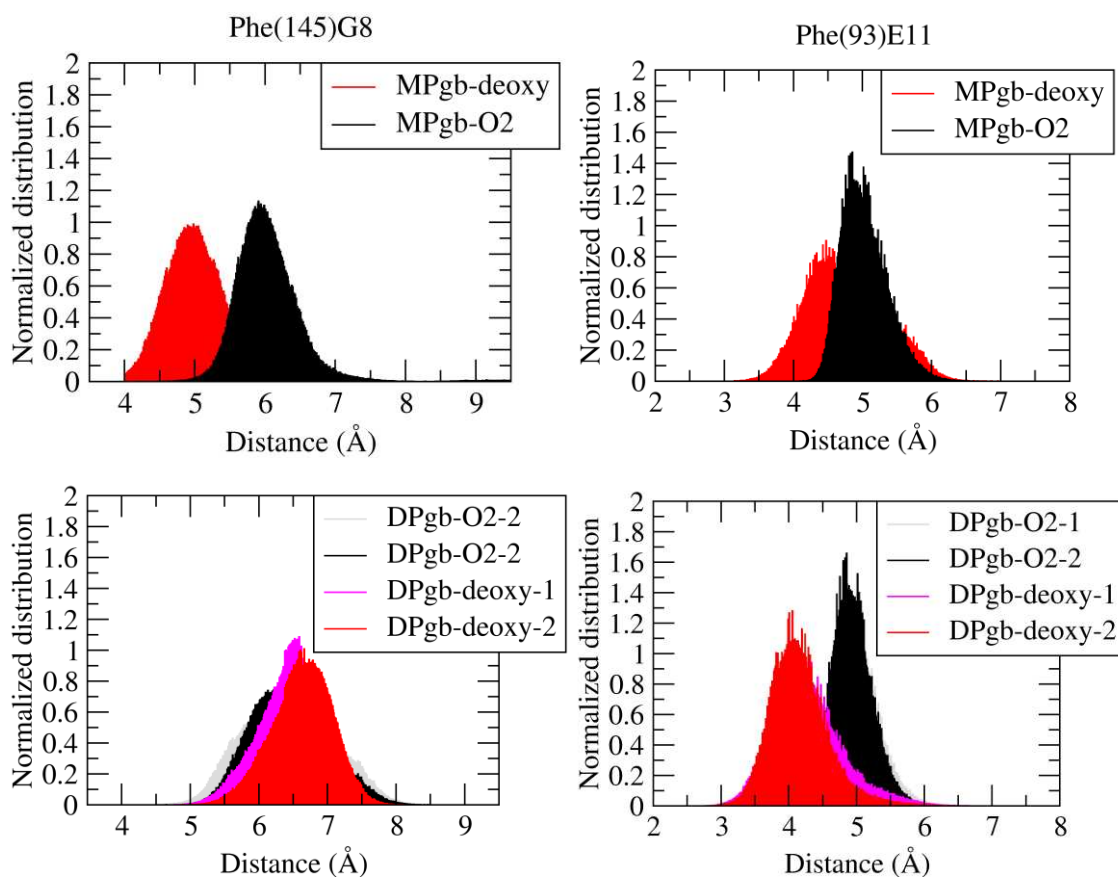


Figure 7. Free energy isosurfaces used to delineate tunnels 1 and 2 for DPgb-deoxy. Whereas tunnel 2 (enclosed in dashed line) is open along the trajectory, tunnel 1 (enclosed in solid line) is found in both open (left) and closed (right) states depending on the conformation of Phe(145)G8 (shown as yellow sticks). Heme is shown in atom-coloured sticks.

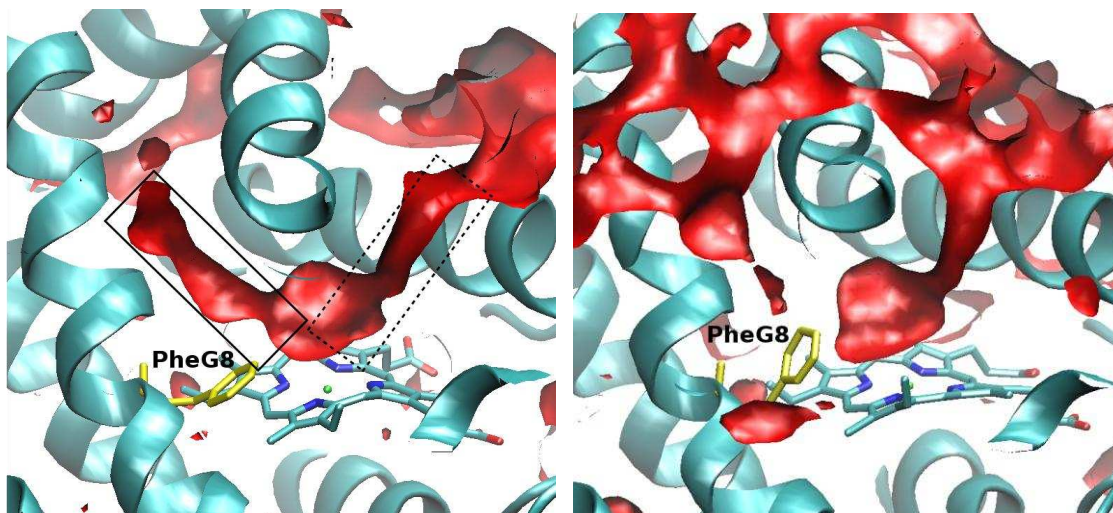


Figure 8. Time (ns) evolution of the Phe(145)G8 N-C $_{\alpha}$ -C $_{\beta}$ -C $_{\gamma}$ torsion for (top-left) the MPgb-deoxy, (top-right) MPgb-O $_2$, (middle) the two subunits (A and B) in the deoxygenated dimer, and (bottom) the two subunits in the oxygenated dimer.

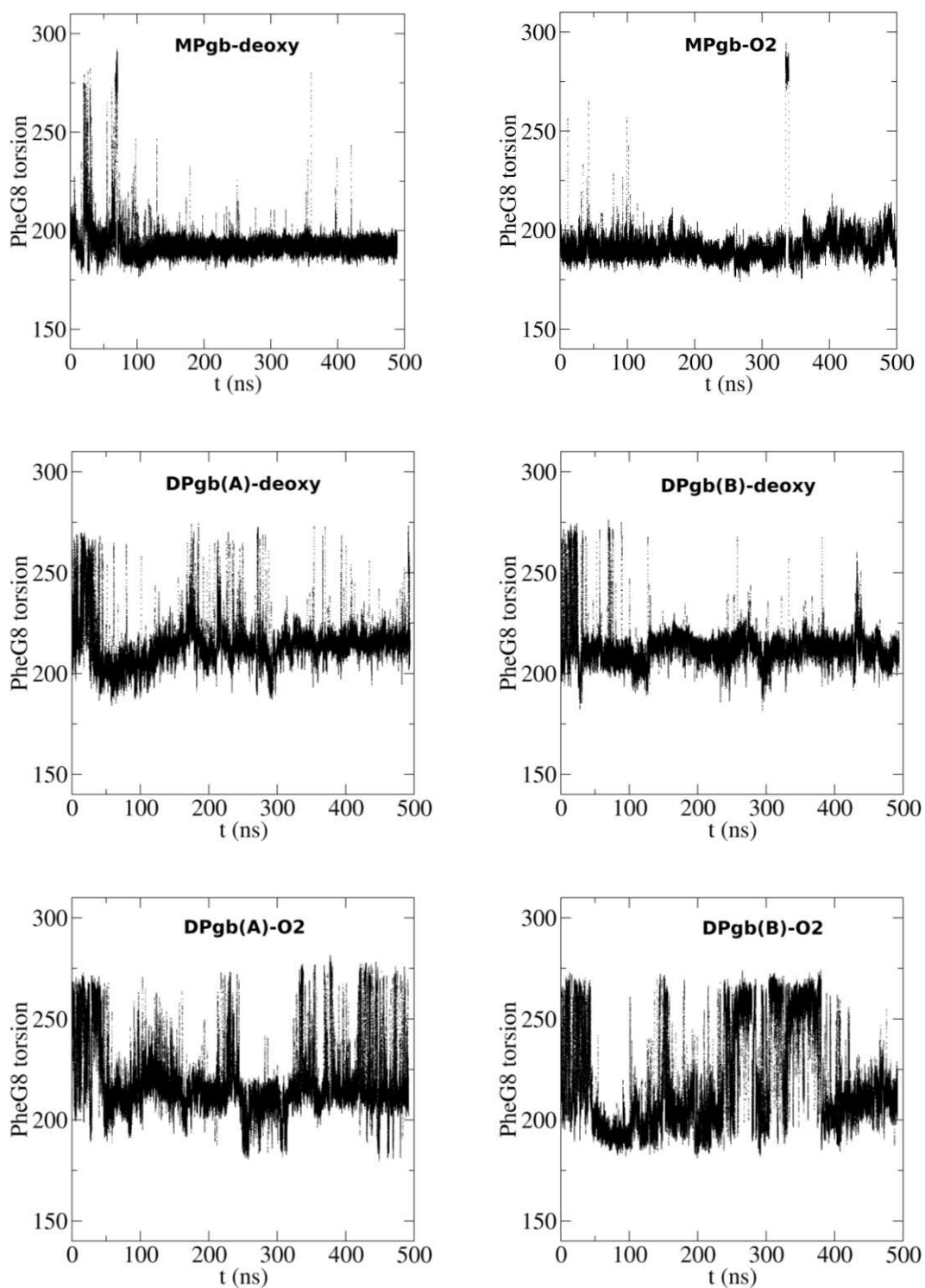
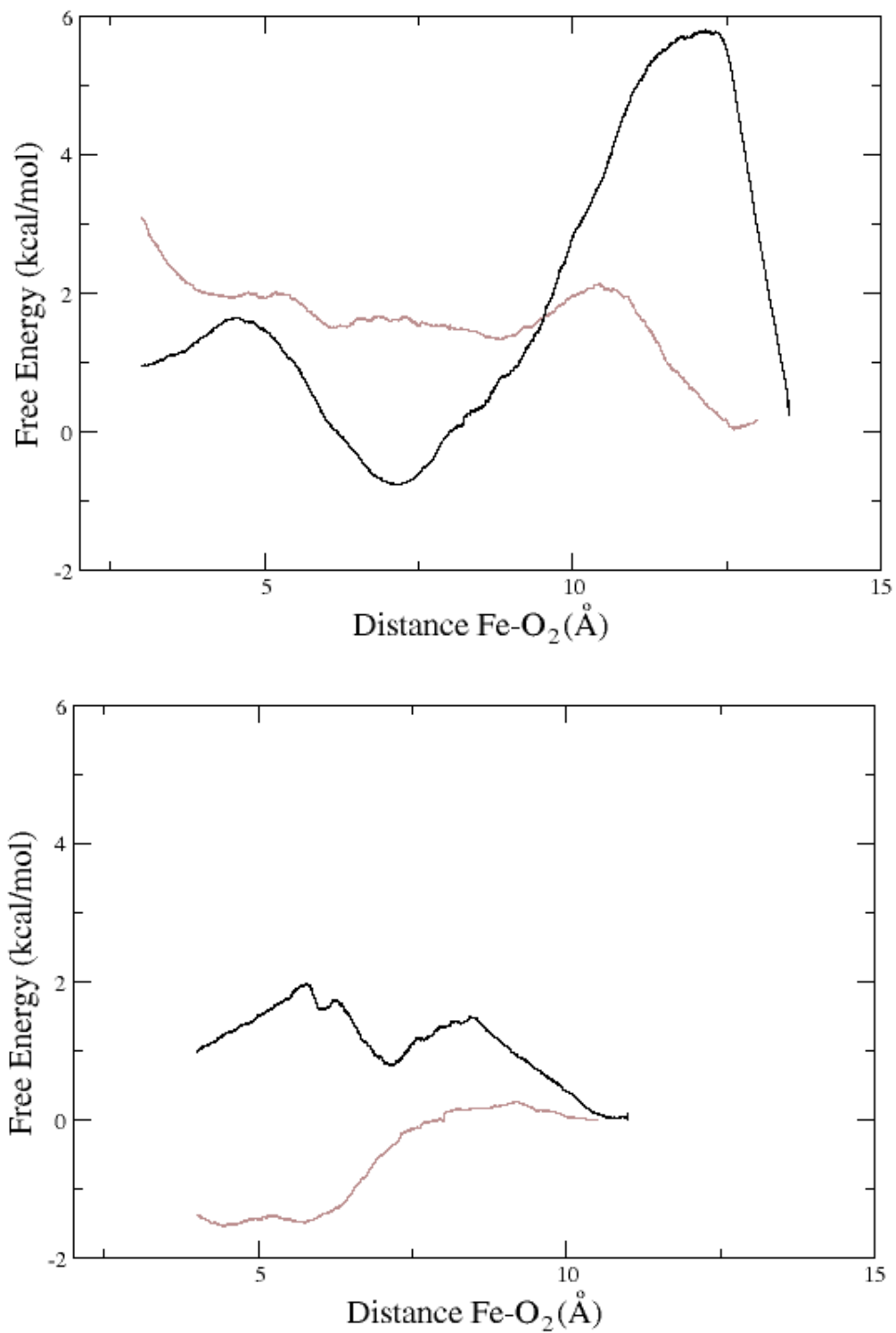


Figure 9. Free energy profiles for O₂ migration through (top) tunnel 1 and (bottom) tunnel 2. The profiles are shown for the deoxygenated protein in monomeric (black) and dimeric (grey) states.



Otras publicaciones

- Forti Flavio, Boechi Leonardo, Estrin Dario A., Marti Marcelo A., Comparing and Combining Implicit ligand sampling with Multiple Steered Molecular Dynamics to study ligand migration processes in heme proteins, J. Comput. Chem., 2011, aceptado
- Forti Flavio, Boechi Leonardo, Novo de Oliveira Ana, Bikiel Damián, Arroyo Pau, Nadra Alejandro, Capece Luciana, Bidon-Chanal Axel, A. Martí Marcelo, Estrín Darío A., Luque F. Javier, Molecular simulations of globins: Exploring the relationship between structure, dynamics and function, Recent Advances in Pharmaceutical Sciences, Eds Diego Muñoz-Torrero, 2011
- Soteras Ignacio, Forti Flavio, Orozco Modesto, Luque F. Javier, Performance of the IEF-MST Solvation Continuum Model in a Blind Test Prediction of Hydration Free Energies, J. Phys. Chem. B, 2009, 113(14), 4501-7
- Svozil Daniel, Šponer Judit, Marchan Ivan, Pérez Alberto, Cheatham Thomas, Forti Flavio, Luque F. Javier, Orozco Modesto, Šponer Jiří, Geometrical and Electronic Structure Variability of the Sugar-phosphate Backbone in Nucleic Acids, 2008, J. Phys Chem B, 112(27), 8188-97