



Universidad Nacional Autónoma de México

Facultad de Química

**“Caracterización *in silico* del mecanismo de
unión de ligandos nativos a la proteína LAO”**

Tesis

Que para obtener el grado de:

Doctor en Ciencias

Presenta:

Daniel Adriano Silva Manzano

Tutor: Dr. Alejandro Sosa Peinado

México, D. F.

2011



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



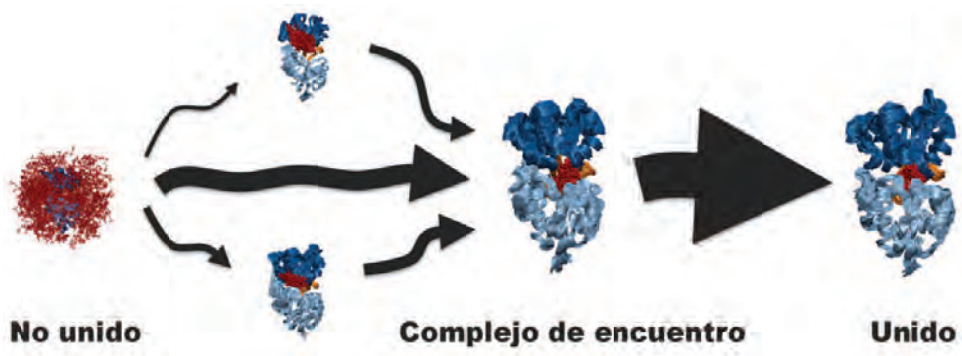
UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

A mi Mamá y Papá





PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS BIOQUÍMICAS

ENTIDADES ACADÉMICAS CAMPUS CIUDAD DE MÉXICO

PMDCB/438/2010

M. en C. DANIEL ADRIANO SILVA MANZANO
Alumno del Doctorado en Ciencias Bioquímicas
P r e s e n t e

Los miembros del Subcomité Académico en reunión ordinaria del día 19 de abril del presente, conocieron su solicitud de asignación de JURADO DE EXAMEN para optar por el grado de DOCTOR EN CIENCIAS (BIOQUÍMICAS), con la réplica de la tesis "**Caracterización in silico del mecanismo de unión de ligandos nativos a la proteína LAO**", dirigida por el Dr. Alejandro Sosa Peinado.

De su análisis se acordó nombrar el siguiente jurado:

PRESIDENTE	Dr. Arturo Rojo Domínguez
VOCAL	Dr. Juan Pablo Pardo Vázquez
SECRETARIO	Dra. Nina Pastor Colón
SUPLENTE	Dr. León Patricio Martínez Castilla
SUPLENTE	Dr. Rogelio Rodríguez Sotres

Sin otro particular por el momento, aprovecho la ocasión para enviarle un cordial saludo.

Atentamente
"POR MI RAZA HABLARÁ EL ESPÍRITU"
Cd. Universitaria, D.F., a 20 de abril de 2010.
EL COORDINADOR DE LA ENTIDAD ACADÉMICA


DR. ROGELIO RODRÍGUEZ SOTRES

C.c.p. Archivo

RRS^{slgg}

Reconocimientos

Este trabajo doctoral se realizó bajo la dirección del Dr. Alejandro Sosa Peinado en el laboratorio de Físicoquímica e Ingeniería de Proteínas en el departamento de Bioquímica de la Facultad de Medicina, UNAM.

El comité tutorial que asesoró el desarrollo de este trabajo estuvo formado por los siguientes miembros:

Dr. Alejandro Sosa Peinado	Facultad de Medicina, UNAM
Dr. Arturo Rojo Domínguez	Universidad Autónoma Metropolitana, Unidad Cuajimalpa
Dr. Lorenzo Segovia Forcella	Instituto de Biotecnología, UNAM

Durante el periodo de agosto del 2007 a agosto del 2010 gocé de una beca para estudios de doctorado otorgada por el CONACYT.

Durante el año 2010 el Dr. Alejandro Fernández Velasco fungió como mi tutor suplente.

Agradecimientos

Si escribiera los agradecimientos cotidianos terminaría escribiendo una página llena de nombres, donde la mayor parte serían amigos, compañeros, familiares, profesores y demás personas que causal o casualmente se han visto involucradas, ya sea en mi vida cotidiana o en mi trabajo. No es mi intención ser mal agradecido, así que cuando alguna de estas personas lea este párrafo, siéntanse aludida y reciba mi sincera gratitud. Sin embargo, deseo agradecer de manera explícita a cuatro personas:

Ante todo, a Alejandro Sosa Peinado, quien desde hace ocho años ha sido “El Tutor”, que me ha apoyado incondicionalmente en todo momento de mi formación académica y personal, sin cuyo apoyo nada de esto podría haber sucedido. Alejandro: “Has sido un gran tutor, modelo y amigo. Gracias, por esto y por todo lo demás que no puedo expresar con palabras”.

Alejandro Fernández, quien ha jugado en más de una vez el papel de mi cotutor. Su apoyo y consejo han influenciado, siempre de manera positiva, mi formación académica.

Arturo Rojo, quien además de ser un gran amigo, es para mí un ejemplo inigualable de calidad humana y científica. Sin su tutela y apoyo este trabajo no sería lo bueno que es ahora.

Xuhui Huang, por haber confiado y aventurarse conmigo a consolidar este proyecto.

Índice

<i>Abstract</i>	7
<i>Resumen</i>	8
<i>Abreviaturas</i>	9
<i>Introducción</i>	10
<i>Objetivo</i>	16
<i>Metodología</i>	17
Estructuras protéicas	17
Simulaciones de MD.....	18
Medición de los ángulos de apertura y dislocación	19
Cuantificación de la interacción de apilamiento	23
Agrupamiento de los datos para las dinámicas cortas de 20 ns	23
Racionalización de la comparación entre LAO e HisJ	24
Descomposición de los estados de Markov (MSM)	24
Escala de Tiempo Implicadas.....	27
Cálculo de las matrices de transición.....	27
Cálculo del tiempo promedio de la primera transición (MFPT).....	28
Cálculo de la energía libre de unión y la tasa de asociación.....	29
Visualización.....	32
Recursos computacionales	32
<i>Resultados</i>	33
Parte I.....	33
Estabilidad térmica del sistema y análisis mediante ángulos diedros	33
LAO en ausencia del ligando (apo-LAO).....	35
LAO en presencia de los ligandos nativos (holo-LAO).....	38
Simulaciones del estado cerrado de LAO.....	39
Simulaciones del estado abierto de LAO.....	43
Simulaciones en presencia de tautómeros neutros de la histidina.	47
Parte II.....	48
MSM para estudiar el mecanismo de unión de LAO.....	48
El mecanismo de unión de LAO.....	51
<i>Discusión</i>	57
<i>Conclusiones</i>	63
<i>Bibliografía</i>	64

Abstract

Molecular recognition is determined by the structure and dynamics of both a receptor and its ligand, but it is difficult to directly assess the role of each of these players. The molecular basis of multiple ligands binding for amino acids in periplasmic binding proteins (PBPs) is an unsolved question. In this work, using unrestrained molecular dynamic simulations (MD), we study the ligand binding mechanism present in the L-lysine, L-arginine, L-ornithine binding (LAO) protein. For the description of the conformational changes upon ligand binding we developed an analysis based on dihedral angles, such analysis exhibits an excellent correlation with each of the two main movements described by a principal component analysis. The analysis of the hydrogen bonds revealed specific protein interactions for each ligand. We also identified an important cation- π interaction of the ligand with the aromatic residues Tyr-14 and Phe-52. Our results suggest that LAO adopts slightly different conformations to make available specific hydrogen bond interactions for each ligand thus, allowing a single mechanism to attain multiple ligand specificity. On the basis of these results we propose a model in which hydrogen bond interactions place the ligand in the correct orientation to induce a cation- π interaction with Tyr-14 and Phe-52, thereby stabilizing the bound state. Then, in a more ambitious project, we combined MD and Markov State Models (MSMs) to elucidate the complete binding mechanism by which the LAO protein captures its ligands. We show that our model can predict the bound state, binding free energy, and association rate with reasonable accuracy, and then we use the model to dissect the binding mechanism. In the past, this binding event has often been assumed to occur via an induced fit mechanism, although more complex mechanisms have also been hypothesized, but these have remained controversial. Here, with our atomistic model, we have been able to directly observe roles for both the conformational selection and induced fit mechanisms in LAO binding. First, the LAO protein tends to form a partially closed encounter complex via a combination of conformational selection and induced-fit pathways. In a second binding step, the interactions with the ligand can induce a transition to the bound state. Based on these results, we propose that MSMs built from atomistic simulations are a powerful way of dissecting ligand-binding mechanisms and may eventually facilitate a deeper understanding of allostery as well as the prediction of new protein-ligand interactions.

Resumen

El reconocimiento molecular se encuentra determinado por la estructura y dinámica tanto del receptor como del ligando, pero es difícil definir el papel de cada uno de estos componentes. Las bases moleculares de la unión múltiple a ligandos en proteínas periplásmicas de unión (PBPs) es una pregunta abierta. En este trabajo, utilizando dinámica molecular no restringida (MD) estudiamos el mecanismo de unión a ligandos en la proteína que une L-lisina, L-arginina y L-ornitina (LAO). Con el fin de describir los cambios conformacionales que tienen lugar en la unión del ligando desarrollamos un análisis basado en ángulos diedros, esta metodología presenta una excelente correlación con los dos movimientos más importantes descritos por un análisis de componentes principales. El análisis de los puentes de hidrógeno reveló interacciones proteicas específicas para cada ligando. También encontramos una interacción importante, del tipo catión- π , entre el ligando y los residuos aromáticos Tyr-14 y Phe-52. Nuestros resultados sugieren que LAO adopta conformaciones ligeramente distintas que provienen de interacciones de puente de hidrógeno específicas para cada ligando, lo que hace posible utilizar un solo mecanismo para lograr una afinidad múltiple. Basados en estos resultados, hemos propuesto un modelo en el que las interacciones de puente de hidrógeno posicionan al ligando en la orientación correcta para inducir una interacción de catión- π con los aminoácidos Tyr-14 y Phe-52, estabilizando así el estado unido. Luego, en un proyecto más ambicioso, combinamos MD y Modelos de Estado de Markov (MSMs) con el fin de elucidar el mecanismo completo de unión por el cual LAO captura a sus ligandos. Hemos mostrado que nuestro modelo puede predecir con una precisión razonable el estado unido, la energía libre de unión y la tasa de asociación. Entonces hemos usado el modelo para estudiar el mecanismo de unión. Anteriormente, se ha asumido que este evento de unión ocurre mediante el mecanismo de ajuste inducido, aunque también se ha hipotetizado la existencia de mecanismos de unión más complejos. En este trabajo, usando nuestro modelo atómico para estudiar la unión en LAO, hemos sido capaces de observar directamente papeles tanto para la selección conformacional como para el ajuste inducido. Primero, la proteína LAO tiende a formar un complejo de encuentro parcialmente cerrado a través de la combinación de rutas que emplean la selección conformacional y el ajuste inducido. En una segunda etapa, la interacción con el ligando puede inducir la transición al estado unido. Basados en estos resultados, proponemos que los MSMs construidos a partir de simulaciones atómicas son una poderosa metodología para diseccionar mecanismos de unión y, eventualmente podrían ayudar a entender con mayor profundidad los mecanismos alostéricos y a predecir nuevas interacciones proteína-ligando.

Abreviaturas

ChoX.- Proteína periplásmica de unión que une colina.

EA.- Estado abierto de la proteína LAO.

EC.- Estado cerrado de la proteína LAO.

GGBP.- Proteína periplásmica de unión que une glucosa/galactosa.

HIS-H.- Histidina con carga formal de +1e.

HIS-ND1 .- Tautómero de la histidina con carga neutra y protón en el nitrógeno δ .

HIS-NE2 .- Tautómero de la histidina con carga neutra y protón en el nitrógeno ϵ .

LAO.- Proteína periplásmica de unión que une L-lisina, L-arginina y L-ornitina.

MBP.- Proteína periplásmica de unión que une maltosa.

MD.- Dinámica molecular (Molecular Dynamics).

MFPT.- Tiempo medio de primer paso o transición (Mean First Passage Time).

MSM.- Modelo de estado de Markov (Markov State Model).

PBP.- Proteína periplásmica de unión (Periplasmic Binding Protein).

PCA.- Análisis de componentes principales (Principal Components Analysis).

PDB.- Banco de datos de proteínas (Protein Data Bank: www.rcsb.org).

RMN.- Resonancia magnética nuclear.

RMSD.- Raíz de la desviación media de los cuadrados (Root Mean Square Deviation).

SHC.- Agrupamiento jerárquico estadístico (Statistical Hierarchical Clustering).

Introducción

Entender los mecanismos moleculares a través de los cuales las proteínas desempeñan su función, es uno de los retos que actualmente enfrenta la bioquímica. Se sabe que la función de las proteínas se encuentra intrínsecamente relacionada con su estructura. Sin embargo, las proteínas lejos de ser estructuras estáticas, se encuentran en constante movimiento, y su función también se encuentra determinada, en parte, por su dinámica. Dentro de las funciones de las proteínas, la unión de moléculas es un evento fundamental para muchas de ellas (posiblemente de todas), pero desafortunadamente nuestro entendimiento en cuanto a las bases moleculares de los mecanismos de unión es aún muy limitado.

La explicación mecánica de los procesos de unión proteína-ligando se reduce esencialmente a dos modelos de unión: el “ajuste inducido” (induced fit) (1) y la “selección conformacional” (conformational selection) (2-9). La diferencia entre estos dos modelos se encuentra en la dinámica intrínseca del receptor y el ligando. En el mecanismo de ajuste inducido el proceso se divide en dos etapas: en la primera el ligando interactúa débilmente con la proteína en una conformación intermedia (complejo de encuentro), luego, este evento de pre-unión induce un cambio conformacional hacia el complejo unido. Por el contrario, en el modelo de selección conformacional, la dinámica intrínseca de la proteína permite que ésta visite la conformación unida (o una muy parecida) en la ausencia del ligando y por lo tanto el ligando puede unirse directamente a la proteína en la conformación unida, desplazando así el equilibrio poblacional hacia el estado unido. Por este último motivo, el modelo de selección conformacional también es conocido como “desplazamiento de población” (population shift).

De manera teórica, explicar la diferencia entre ambos mecanismos es relativamente simple, y en los últimos años este tema ha sido abordado desde varias perspectivas, por ejemplo: Nusinov et. al. ha explicado en numerosas ocasiones (2-6, 10) la diferencia entre los distintos paisajes energéticos y conformacionales, derivados de uno u otro mecanismo; Zhou ha propuesto que la escala temporal de la transición entre el estado unido y no-unido es el factor determinante para discernir entre los mecanismos (11), y deduce que la selección conformacional es el mecanismo dominante cuando la transición entre el estado unido y no-unido es lenta, mientras que el ajuste inducido ocurre cuando la transición es rápida. Finalmente, Okazaki et. al. han sugerido que interacciones proteína-ligando fuertes y de largo alcance favorecen un mecanismo

de ajuste inducido, e interacciones débiles y de corto alcance resultarían en un mecanismo de selección conformacional (12).

Durante las últimas cinco décadas, a partir de que Koshland propuso el modelo de ajuste inducido (1), en una gran variedad de sistemas se ha intentado determinar si sus mecanismos de unión son mejor descritos por uno u otro modelo (10-17). Sin embargo, la complejidad experimental para realizar esta distinción ha demostrado ser una limitante, y en la mayor parte de los casos aún no existen pruebas concluyentes de la participación de uno u otro mecanismo. En los últimos años, en más de un estudio se ha señalado que ambos modelos no son estrictamente excluyentes y que una combinación de ambos mecanismos podría ajustarse mejor a los resultados experimentales (11-13, 15, 18, 19).

Independientemente del mecanismo operante, en última instancia el mecanismo de unión de una proteína se encuentra determinado por la dinámica del receptor y la del ligando, y éstas a su vez, se encuentran definidas por su paisaje energético. Dado que poseemos una limitada capacidad experimental para observar la dinámica microscópica de las proteínas, determinar de manera concluyente los mecanismos de unión operantes en diversas proteínas es uno de los retos actuales de la bioquímica. Un ejemplo clásico se encuentra en las proteínas periplásmicas de unión (PBPs). En general, las PBPs son proteínas monoméricas de estructura y función relativamente sencillas. Su topología se compone de dos lóbulos (Lóbulo I y II) conectados por una bisagra. Numerosas PBPs han sido estudiadas bioquímicamente de manera extensa y las estructuras cristalográficas de más de 100 miembros de la familia se encuentran públicamente disponibles. Los ligandos que las diferentes PBPs abarcan, cubre un amplio espectro de formas y tamaños (20). Todos los miembros conocidos de la familia de las PBPs comparten un mismo tipo de plegamiento (plegamiento Rossman), y se ha propuesto que las PBPs y otras proteínas con la misma topología podrían compartir una dinámica similar (21, 22). Para la mayor parte de las PBPs se sabe que existen dos estados conformacionales estables, uno en ausencia de ligando (llamado “estado abierto”) y otro cuando ha unido el ligando (“estado cerrado). En resumen, se posee una extensa información estructural y bioquímica de las PBPs. No obstante, no se ha logrado precisar el mecanismo de unión en ningún miembro de la familia. Muchos han supuesto que el mecanismo de unión en las PBPs debe corresponder al ajuste inducido; esta hipótesis se ha generado principalmente con base en dos observaciones: 1) que el ligando es completamente excluido del solvente en el estado unido (ver Figuras 1 y 2) que el cambio conformacional de las

PBPs ocurre en presencia del ligando (23-29). Sin embargo, algunos estudios experimentales han mostrado que al menos dos PBPs: la proteína que une glucosa (GGBBP; (30) y la proteína que une colina (ChoX; (31) son capaces de visitar la conformación cerrada en ausencia del ligando. Estos hallazgos ponen de manifiesto la posibilidad de que el mecanismo de selección conformacional participe en el mecanismo de ciertas PBPs (27). Más aún, un estudio reciente realizado en la PBP que une maltosa (MBP), ha logrado identificar en ausencia de ligando una población menor de una conformación “parcialmente cerrada” (~5% población), en equilibrio con la conformación no-unida (conformación abierta; (32). El estudio no encontró evidencia de que en ausencia de ligando MBP sea capaz de visitar la conformación correspondiente al estado unido (conformación cerrada). Basados en estas observaciones, los autores han propuesto que la conformación parcialmente cerrada se encuentra disponible para la unión del ligando y, dado que su conformación es cercana al estado unido, podría actuar como un intermediario que facilita la unión de ligando. De ser ciertas, estas suposiciones implicarían que, en MBP tanto la selección conformacional como el ajuste inducido participan en la unión del ligando. Desafortunadamente estos experimentos se realizaron en la ausencia del ligando, por lo que los mecanismos propuestos son especulativos y no pueden precisar los papeles y contribuciones de las distintas vías al mecanismo global de unión de MBP.

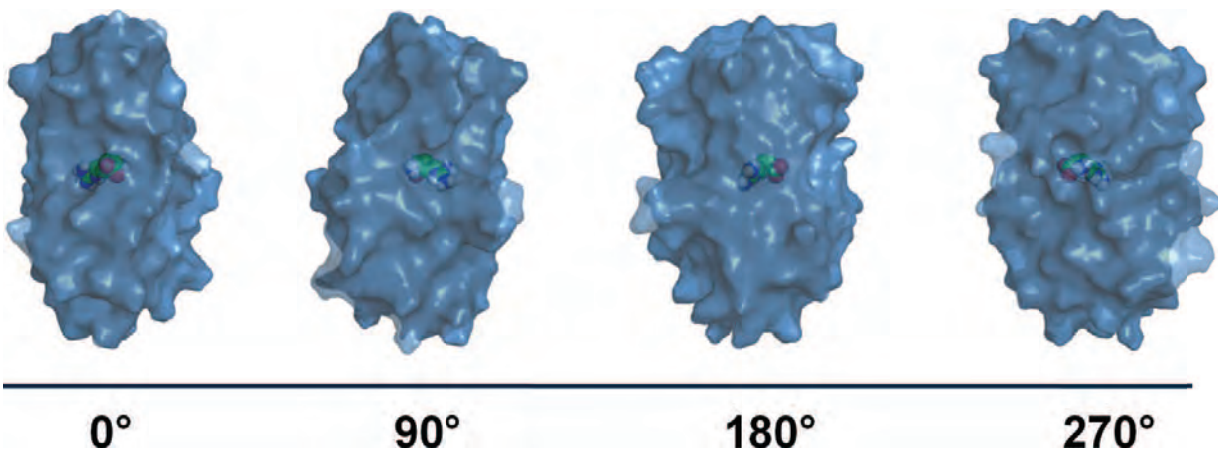


Figura 1. La estructura cristalográfica de LAO en el estado cerrado con el ligando arginina (PDB ID: 1LAF) desde cuatro ángulos distintos. Una vez cerrada, la proteína engloba al ligando. Por lo tanto, si la proteína vacía alcanza el estado cerrado el ligando no puede alcanzar el sitio de unión.

Las simulaciones de dinámica molecular (MD) tienen el potencial de permitirnos estudiar, con detalle atómico, la dinámica de las proteínas, y por lo tanto podrían constituir una poderosa herramienta para entender los mecanismos de unión presentes en los sistemas biológicos. Sin embargo, dos de los problemas asociados a este enfoque son: (i) la reducida escala temporal de los eventos que se pueden estudiar mediante MD (usualmente en la escala de los nanosegundos) y (ii) la dificultad en extraer datos estadísticos que puedan validar el modelo. Los avances en este campo, respecto a la metodología de simulación y/o su análisis, tienen el potencial de influenciar grandemente nuestra capacidad para entender la relación existente entre la dinámica, estructura y función de las proteínas. El objetivo de este trabajo es el estudio del mecanismo de unión de la PBP que une L-lisina, L-arginina y L-ornitina (LAO) del organismo *Salmonella typhimurium*. Para tal fin, se ha empleado un enfoque computacional basado en simulaciones masivas de MD, modelos matemáticos y análisis estadísticos. LAO es una proteína de 238 aminoácidos (26.0 kDa) con especificidad nanomolar a múltiples ligandos (33): L-arginina, L-lisina, L-histidina y L-ornitina ($K_d = 14, 15, 28$ y 500 nM, respectivamente). Los ligandos de LAO comparten la característica de tener carga positiva, lo que sugiere una importante contribución de la interacción electrostática a el mecanismo de unión. Se conocen las estructuras tridimensionales correspondientes a LAO en la forma libre (**apo**) (24) e interactuando con cada uno de sus ligandos (**holo**) (23). Topológicamente, la proteína está compuesta por dos dominios (o lóbulos) conectados por una bisagra. El lóbulo I contiene residuos tanto de los extremos NH_3 y COOH (a.a. 1-88 y 195-238, respectivamente), mientras que el lóbulo II es continuo y contiene a los residuos 93-185. Los experimentos cristalográficos han revelado la misma conformación de la proteína cuando ésta une a cualquiera de sus cuatro ligandos (RMSD de los $\text{C}\alpha$ para los residuos 5-235 < 0.18 Å, entre cualquier par de estructuras;(34). No obstante, en desacuerdo con la cristalografía, estudios realizados en la proteína HisJ (ver Métodos: racionalización de la comparación entre LAO e HisJ) han mostrado evidencia que sugiere que, en solución, la proteína adopta diferentes conformaciones de acuerdo al ligando con que interactúa (29). La cristalografía de rayos-X y experimentos espectroscópicos también han revelado que, en LAO como en la mayor parte de las PBPs, el evento de unión del ligando involucra un rearrreglo conformacional de gran escala entre dos conformaciones (23, 24): (i) el **estado abierto**, cuando el sitio de unión se encuentra vacío (ver Figura 2) y (ii) el **estado cerrado**, cuando el sitio de unión se encuentra ocupado por el ligando (ver Figura 2).

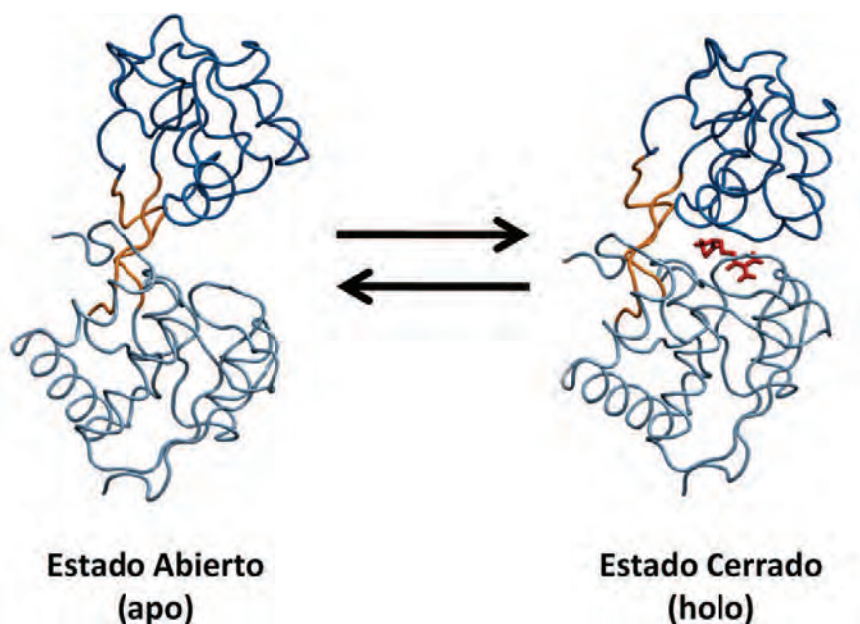


Figura 2. La proteína LAO en el estado abierto (PDB ID: 2LAO) y en el estado cerrado (PDB ID: 1LAF). El ligando arginina se muestra en color rojo en una representación de bastones (sticks).

Utilizando un enfoque teórico, hemos monitoreado el mecanismo de unión de LAO en dos etapas:

Primero, con el fin de describir el paisaje conformacional de la proteína generamos una metodología de análisis basada en la medición de dos coordenadas de reacción: el ángulo de apertura y el de dislocación. Utilizando estas coordenadas de reacción hemos demostrado que mediante MD es posible simular el espacio conformacional de LAO. Es decir, una conformación cerrada en presencia del ligando, una conformación abierta en su ausencia y la inducción del cambio conformacional del estado abierto al cerrado (en presencia del ligando en el sitio de unión). Nuestros resultados muestran que, contrario a lo que la evidencia cristalográfica señala y en acuerdo con experimentos en solución (29), LAO adopta una conformación distinta en presencia de cada uno de los ligandos. También hemos estudiado el papel de los distintos tautómeros de L-his en el ensamble conformacional de LAO. Nuestros resultados, al igual que los resultados experimentales, sugieren que el estado de protonación del ligando L-his juega un papel determinante en la afinidad. En conjunto, estos datos nos han dado la pauta para generar un modelo plausible, con detalles a nivel atómico, del mecanismo que opera durante el cambio conformacional producido por el reconocimiento del ligando.

En la segunda parte, abordando un objetivo más ambicioso, utilizamos un muestreo intensivo de MD para reproducir el evento *ab initio* de la unión del ligando (i.e. desde el estado no-unido –en el solvente– hasta el estado unido –en el sitio de unión–). Luego, para monitorear el mecanismo de unión hemos utilizado un “modelo de estados de Markov” (MSM). Los MSM han mostrado ser un método poderoso para identificar los estados meta-estables en los mecanismos de plegamiento de proteínas y RNAs (35-39). Una vez que en un MSM se han identificado los estados meta-estables –mínimos de energía en el paisaje energético, que se traducen en conformaciones que persisten por periodos extensos de tiempo antes de transitar hacia otro estado– es posible utilizar el modelo para derivar parámetros termodinámicos y cinéticos del sistema simulado (35, 40-42). En este estudio, demostraremos que los MSM también pueden capturar el mecanismo de unión proteína-ligando. Nuestro modelo sugiere que para LAO, tanto la selección conformacional como el ajuste inducido forman parte del mecanismo global de unión. Además de los estados abierto y cerrado, hemos identificado un tercer estado meta-estable. Este tercer estado, obligatorio en todas las rutas de unión, está formado por una conformación *parcialmente*-cerrada que interactúa débilmente con el ligando, por lo tanto, hemos concluido que tal conformación representa al complejo de encuentro. Nuestros resultados indican que, si bien, la interacción con el ligando puede inducir a la proteína a transitar hasta el complejo de encuentro (ajuste inducido), también la dinámica intrínseca de la proteína LAO le permite alcanzar el complejo de encuentro en ausencia del ligando, revelando así un papel importante para la selección conformacional. Una vez en el complejo de encuentro, el mecanismo de ajuste inducido es el responsable de la transición al estado cerrado.

Objetivo

Mediante el uso de herramientas computacionales, generar y validar un modelo que explique el mecanismo general de unión de la proteína LAO.

Metodología

Estructuras protéicas

Las siguientes estructuras de LAO (cristalografía de rayos-X) obtenidas del RCSB Protein data bank (43-45) fueron utilizadas para este estudio: estado abierto PDB-ID: 2LAO, y estado cerrado PDB-IDs: 1LAF, 1LST y 1LAG (ligandos L-arginina, L-lisina y L-histidina). Los átomos faltantes de la estructura 2LAO (3 cadenas laterales: Gln-4, Lys-44 y Lys 225) se construyeron automáticamente con el programa “Swiss PDB Viewer” versión 3.7 (46). Cada uno de los estados cerrados y con ligando fue simulado por triplicado mediante MDs de 20 ns. Se comprobó que en todas estas simulaciones la proteína permaneció en el estado cerrado.

Al no existir la estructura de la proteína LAO en el estado abierto con los ligandos arginina, lisina e histidina, se construyeron dichas estructuras a partir de las estructuras conocidas mediante la siguiente metodología: usando el programa PyMol (versión 0.99rc6) (47) se realizó un alineamiento del lóbulo I entre las estructuras correspondientes a LAO en el estado abierto (2LAO) y en el estado cerrado con los diferentes ligandos (1LAF, 1LST y 1LAG); de estas estructuras se extrajeron las coordenadas de los diferentes ligandos y se insertaron en el archivo de la estructura del estado abierto (2LAO), obteniéndose así los archivos de la estructura en estado abierto con los ligandos: arginina, lisina e histidina: 2LAO-ARG, 2LAO-LYS, 2LAO-HISH, 2LAO-HIS-ND y 2LAO-HIS-NE. Posteriormente se analizaron estos archivos para identificar los impedimentos estéricos. Finalmente se realizó un ajuste manual ($<1 \text{ \AA}$ en x, y, z) de las coordenadas del ligando para eliminar impedimentos estéricos de acuerdo con los radios de van der Waals. Para el estado cerrado, con el fin de evitar cualquier variabilidad asociada a estructuras proteicas distintas, se empleó la estructura proteica de 1LAG y se reemplazó el ligando histidina por arginina o lisina mediante el alineamiento con las estructuras cristalográficas 1LAF o 1LST. Para las 65 simulaciones de 200ns con el ligando L-arg, se utilizó el programa PyMol para colocar el ligando, en una posición arbitraria, a una distancia de 25 Å del sitio de unión de la estructura proteica 2LAO, 10 simulaciones fueron iniciadas en la misma condición, pero con distintas distribuciones aleatorias de las velocidades iniciales, las 55 simulaciones restantes fueron iniciadas a partir de estructuras extraídas aleatoriamente de este primer conjunto de 10 simulaciones.

Simulaciones de MD

Las simulaciones de dinámica molecular (MD) se realizaron usando el programa GROMACS (ver. 3.3.1 y 4.0.x) (48-50). El procedimiento general para cada MD consistió en:

Solvatar la proteína en una caja dodecaédrica de un tamaño de 375.79 nm^3 usando ≈ 11300 moléculas de agua "SPC". El tamaño de la caja fue suficiente para que el límite se encuentre a por lo menos 1 nm de distancia de cualquier átomo de la proteína LAO.

Minimizar el sistema mediante un algoritmo de máxima pendiente (steepest descent), durante el número necesario de pasos para que el cambio máximo en la energía fuera menor a $100 \text{ kJ mol}^{-1} \text{ nm}^{-1}$ o un máximo de 10,000 pasos de minimización, lo que sucediera primero.

Para relajar el solvente en el sistema, se realizaron 250 ps de una MD de posiciones restringidas (1000 kJ mol^{-1}) en los átomos pesados de la proteína y el ligando.

Finalmente, se realizó la dinámica de producción. Inicialmente las velocidades de los átomos del sistema se generaron aleatoriamente a partir de un conjunto basado en una distribución de Maxwell a 318 K. Los parámetros de las MD fueron: campo de fuerza GROMOS_53a6; algoritmo de integración "md" (leapfrog); baño térmico a 318 K; celdas periódicas; electrostática PME; presión constante de 1 atm; termostato isotrópico de Berendsen; barostato isotrópico de Berendsen; compresibilidad del sistema 4.5×10^{-5} bar; enlaces restringidos por LINCS; sistema con carga total de cero mediante la adición de suficientes contra iones de Na^+ ; Pasos de integración de 5 fs mediante el uso de sitios virtuales de interacción para los hidrógenos, a excepción de los hidrógenos de los anillos aromáticos en los que se utilizaron hidrógenos pesados (4 uma). Para validar los pasos de integración de 5fs, se compararon los resultados con los obtenidos para simulaciones con pasos de integración de 2fs y sin el uso de sitios virtuales de interacción. Los resultados demostraron que el sistema visita un espacio conformacional similar en ambos casos (ver Figura 3).

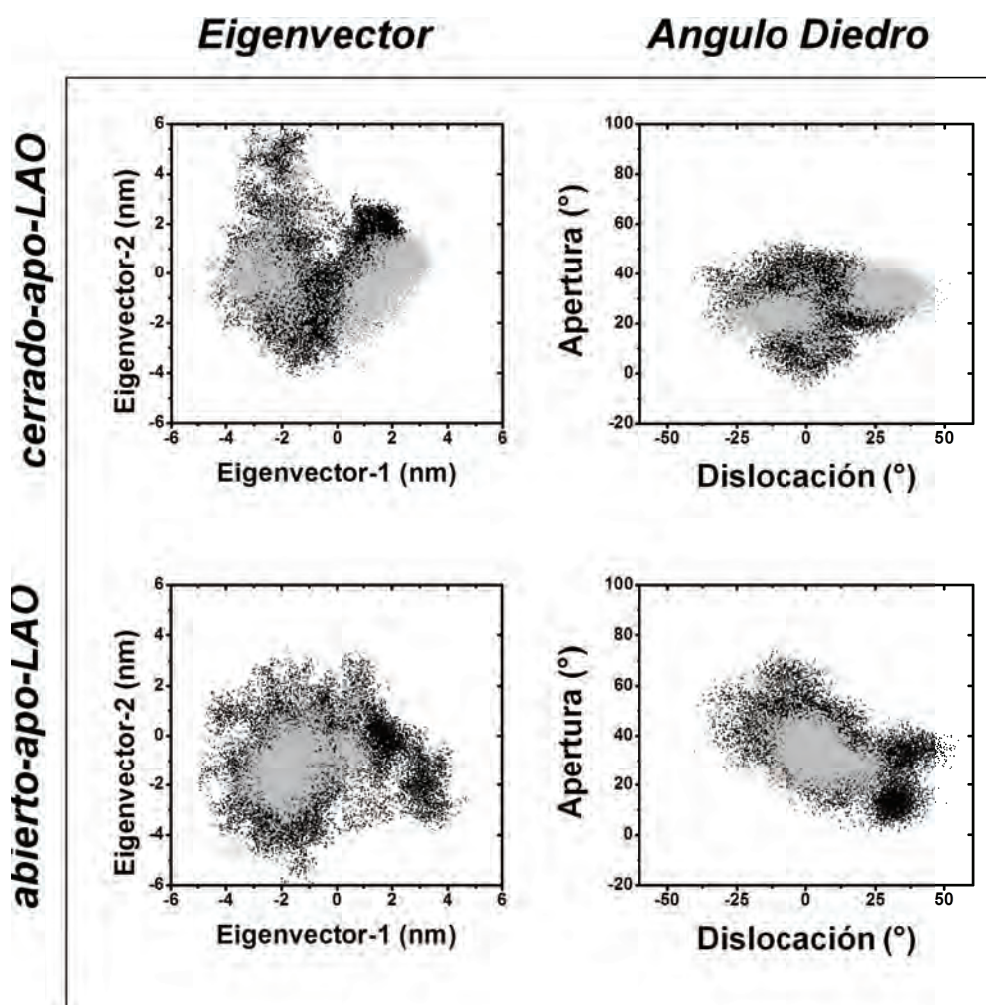


Figura 3. Comparación de los espacios conformacionales visitados por simulaciones usando pasos de tiempo de 2 y 5 fs. Los gráficos muestran proyecciones bidimensionales de i) los dos primeros eigenvectores de PCA (izquierda) y ii) los ángulos diedros de apertura y dislocación. Los gráficos muestran los resultados para simulaciones iniciadas en el estado apo abierto y cerrado. Los puntos color negro corresponden a los pasos de integración de 2 fs (sin sitios virtuales para los hidrógenos) y los grises de 5 fs (con sitios virtuales para los hidrógenos). Estos resultados validan el uso de los pasos de integración de 5 fs con sitios virtuales de interacción.

Medición de los ángulos de apertura y dislocación

Los ángulos diedros que definen la apertura y dislocación de LAO se calcularon como el ángulo entre los vectores normales de dos planos. Cada plano fue definido por tres puntos que corresponden a los centros de masa de tres grupos de carbonos alfa (ver Figura 4). Dado que los movimientos de apertura y dislocación son esencialmente movimientos de bisagra, los ángulos fueron calculados entre dos planos que comparten dos puntos. Los grupos de átomos que definen cada plano fueron seleccionados manualmente y refinados mediante un ensayo de prueba y error

con el objetivo de ajustar los siguientes tres parámetros: i) evitar correlación cruzada, es decir que sean capaces de medir independientemente la apertura y la dislocación; ii) maximizar la diferencia entre los valores para el estado abierto y el estado cerrado cristalográficos; iii) Correlacionar con el primer o segundo eigenvector del PCA encontrado en MD iniciadas en el estado cerrado libre de ligando, donde el primer eigenvector corresponde la dislocación y el segundo eigenvector a la apertura. Como resultado de estos criterios los planos fueron definidos de la siguiente manera:

Apertura:

Plano-A: residuos 6-88 y 195-227; 162-168; 121-127.

Plano B: residuos 92-185; 162-168; 121-127.

Por lo tanto los puntos compartidos entre estos plano son: 162-168 and 121-127.

Dislocación:

Plano-A: residuos 6-88 y 195-227; 83-88 y 194-199; 92-97 y 156-161;

Plano-B: residuos 92-185; 83-88 y 194-199; 92-97 y 156-161.

Por lo tanto los puntos compartidos entre estos plano son: 83-88 y 194-199; 92-97 y 156-161.

Por definición, en la estructura cristalográfica del estado cerrado unido a L-arg (PDB ID: 1LAF) ambos ángulos diedros tienen un valor de cero. La estructura abierta y libre de ligando (PDB ID: 2LAO) después de una minimización de energía tiene un ángulo de apertura de 38.2° y dislocación de -26.2° . Un análisis de la correlación entre estos ángulos diedros y los dos primeros eigenvectores del PCA muestran coeficientes de correlación de 0.83 y 0.84 para la dislocación y la apertura, respectivamente (ver Figura 4).

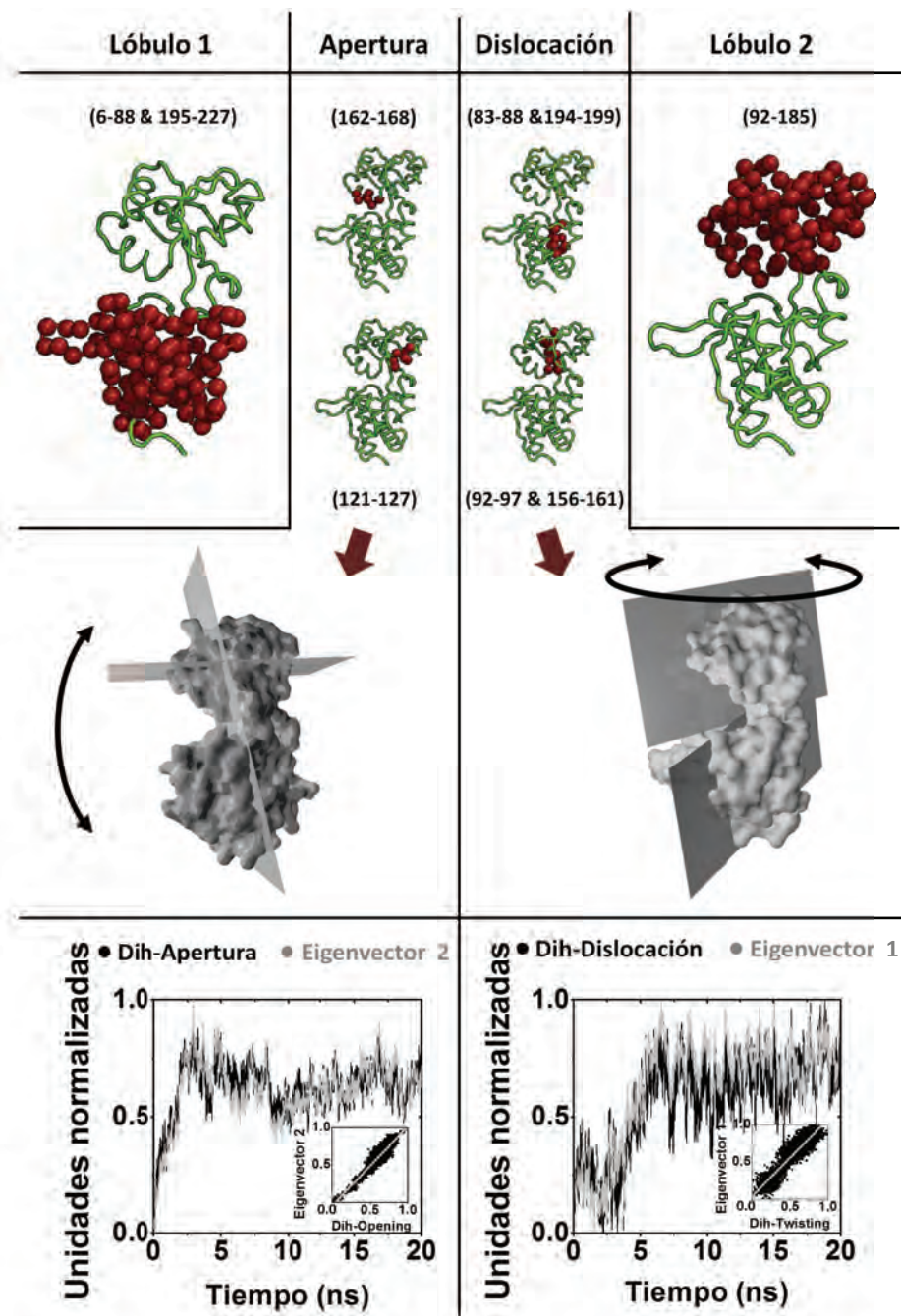


Figura 4. Grupos de C α usados para medir los movimientos de dislocación y apertura. Panel superior: LAO se muestra en una representación de listones. Las esferas indican los C α que se utilizaron para medir los ángulos diedros de dislocación y apertura; Panel medio: representación esquemática de los planos utilizados para medir los ángulos dados los grupos de C α definidos en el panel superior; Panel inferior: gráfico que muestra la correlación de los dos primeros eigenvectores encontrados por PCA de las MDs iniciadas en el estado apo-cerrado. Los diedros son mostrados en color negro y los vectores propios en gris. Dentro de cada gráfica un recuadro muestra un gráfico de correlación del ángulo diedro y el vector propio correspondiente (valor de correlación: R = 0.83 y 0.84, para la dislocación y la apertura, respectivamente).

El análisis de PCA también muestra que los dos primeros componentes principales comprenden más del 60% de los movimientos totales en la proteína. Los ángulos diedros fueron comparados con el RMSD de los C α (ver Figura 5) y se encontró que nuestra definición es lo suficientemente buena como para revelar los mismos cambios conformacionales que el RMSD, pero con ciertas ventajas, dado que: el cálculo de los ángulos diedros son más eficientes computacionalmente, no requieren de una estructura de referencia y son intuitivos (e.g. un valor menor en el ángulo de apertura significa que la proteína está más cerrada y viceversa). En conjunto estos datos demuestran que los ángulos diedros son descriptores razonables del estado conformacional de LAO.

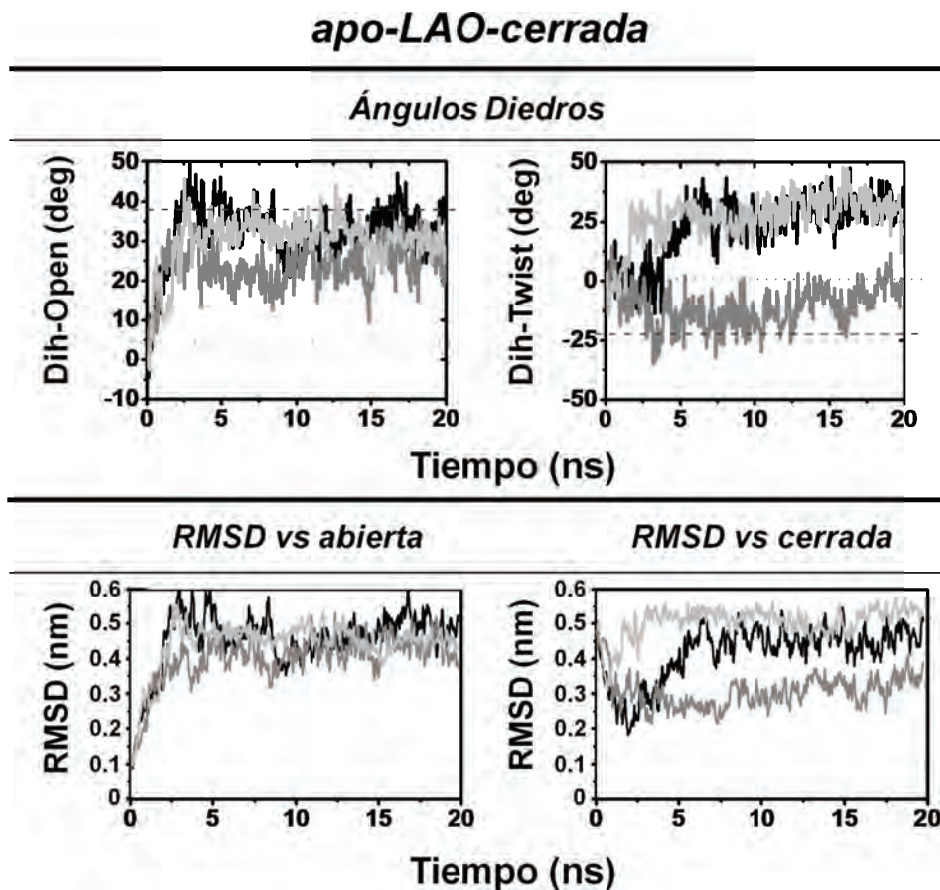


Figura 5. Comparación de los ángulos diedros y el RMSD en trayectorias de apo-LAO iniciadas en el estado cerrado. En el panel superior se muestran los resultados de los ángulos diedros, mientras que en los paneles inferiores se muestran los resultados para los cálculos de RMSD. Tres replicas se muestran en negro, gris y gris-claro. La figura demuestra que los ángulos diedros muestran información similar al RMSD, pero los diedros pueden generar esta información sin utilizar una estructura de referencia; adicionalmente los diedros revelan la magnitud y dirección del movimiento (e.g. 10° en la dirección de la apertura).

Cuantificación de la interacción de apilamiento

Para cuantificar la interacción de apilamiento entre el ligando y los residuos Tyr-14 y Phe-52, primero se definieron los planos geométricos compuestos por tres átomos pesados de las cadenas laterales de los amino ácidos de la siguiente manera: L-arg (NE, NH1 y NH2); L-his: (CG, ND1 and NE2); Tyr-14 (CG, CE1 and CE2); y Phe-52 (CG, CE1 and CE2). Entonces calculamos el ángulo diedro entre los vectores normales de dos de los planos (e.g. entre L-arg y Tyr-14). Definimos la interacción de apilamiento como un evento discreto con valores de 0 y 1 (falso y verdadero, respectivamente), donde verdadero significa que el ángulo entre los planos = $[-180^\circ | 0^\circ | 180^\circ]$, con una tolerancia de $\pm 30^\circ$. De esta manera, sólo aquellas estructuras en las cuales los anillos aromáticos de la Tyr-14 o Phe-52 se encuentran esencialmente paralelos a la cadena lateral de ligando son considerados como una interacción verdadera. Cada conformación del conjunto de datos fue analizada utilizando este método y el promedio fue usado para calcular la fracción de apilamiento (P).

Agrupamiento de los datos para las dinámicas cortas de 20 ns

Para estudiar el mecanismo de unión del ligando en las simulaciones de 20 ns se realizó un agrupamiento de las conformaciones en las trayectorias con transiciones espontáneas al estado cerrado. El agrupamiento utilizó el algoritmo de centroides-K del MSMBuilder (42). Como medida se empleó la distancia euclidiana sobre dos coordenadas de reacción: los ángulos de apertura y dislocación. El agrupamiento fue realizado independientemente en el ensamble de las dinámicas con transiciones espontáneas al estado cerrado para los ligandos L-arg o L-lys (8 y 5 trayectorias, respectivamente). Se probaron agrupamientos de 5, 10, 20, 50 y 100 centroides, y se determinó que 10 centroides podían describir de manera simple las MDs con el ligando L-arg y 5 centroides a las MDs del ligando L-lys. Tras el agrupamiento se identificaron 4 grupos para cada ligando que de manera secuencial representan la transición desde el estado abierto al estado cerrado. Para facilitar la visualización los estados se numeraron secuencialmente: 1, 2, 3 y 4, donde el estado 1 representa el estado inicial (abierto), 2 y 3 representan estados intermedios, y 4 es el estado cerrado y unido.

Racionalización de la comparación entre LAO e HisJ

Las PBP LAO e HisJ tienen una alta similitud estructural y funcional. Ambas proteínas tienen el mismo tamaño (238 a.a.) y un 70% de identidad en la secuencia. De hecho, si se toman en cuenta las mutaciones conservativas, la similitud de la secuencia se incrementa al 83% (51). Estas proteínas se unen al mismo complejo receptor de membrana (HisQ/HisM/HisP)(24) y, también, a los mismos ligandos (L-amino ácidos catiónicos) (52). El RMSD entre las estructuras cristalográficas de LAO e HisJ unidas a L-his (1LAG y 1HBP, respectivamente) es bajo ($C\alpha$ RMSD = 0.62 Å). La afinidad de unión de estas proteínas por sus ligando es similar (nanomolar), aunque para HisJ la L-his es el ligando más fuerte, mientras que para LAO es el menos afín (afinidad inversa por los mismos ligandos). La similitud entre LAO e HisJ ha sido discutida con detalle en un estudio previo de Oh *et. al.* (52). Por lo tanto, creemos que es razonable suponer que LAO e HisJ son comparables en cuanto a su cinética y mecanismos de unión.

Descomposición de los estados de Markov (MSM)

Para construir la descomposición de los estados de nuestro MSM de la unión de ligando de LAO utilizamos el programa MSMbuilder (42) y el algoritmo de SHC (37).

Primero utilizamos el algoritmo de centroides-k del MSMBuilder (42) para agrupar nuestros datos en un gran número de micro-estados. El objetivo de este agrupamiento fue el de combinar en un mismo grupo a aquellas conformaciones que son geoméricamente similares y que por lo tanto es razonable suponer (y después verificar) que se encuentran conectadas cinéticamente. Debido a que se deseaba tomar en cuenta tanto la conformación de la proteína como la del ligando realizamos dos agrupamientos separados, uno basado en la conformación global de la proteína (sobre los ángulos de apertura y dislocación) y otro basado en la posición relativa del ligando. Después combinamos estos dos agrupamientos tratándolos como conjuntos independientes de datos. Por ejemplo, P-grupos basados en la conformación de la proteína y L-grupos basados en el ligando dan un total de $P \times L$ “clusters”.

Para el agrupamiento basado en la proteína creamos 50 grupos basados en la distancia euclidiana entre los vectores del ángulo de apertura y dislocación. En el caso del ligando creamos un agrupamiento de 5000 “clusters” basados en su posición relativa. Luego el agrupamiento fue

modificado para tomar en cuenta que en la dinámica existen dos escenarios distintos: uno en el que, debido a la interacción con proteína, el ligando se mueve lentamente y, otro en el que el ligando puede difundir libre y rápidamente en el solvente (ver Figura 6). Los agrupamientos previamente mencionados son adecuados para describir el primer escenario. Sin embargo, cuando el ligando difunde libremente (a más de 5 Å de la proteína), el procedimiento anterior resulta en un gran número de grupos que generan estadísticas pobres (menos de diez transiciones a otros micro-estados). Tratar de muestrear estos micro-estados para obtener estadísticas confiables representaría un desperdicio de recursos computacionales ya que existe una teoría analítica para la difusión de moléculas y un modelo detallado de MSM proveería poca o nula información al respecto. En lugar de esto, elegimos generar un re-agrupamiento de baja resolución de los estados con pobres estadísticas (y a más de 5 Å de la proteína) utilizando las mismas coordenadas de reacción de la proteína pero empleando esta vez el centro de masa del ligando (en oposición a la distancia euclidiana entre todos los átomos pesados del ligando). Para este paso generamos 10 grupos para la proteína y 100 grupos para el ligando.

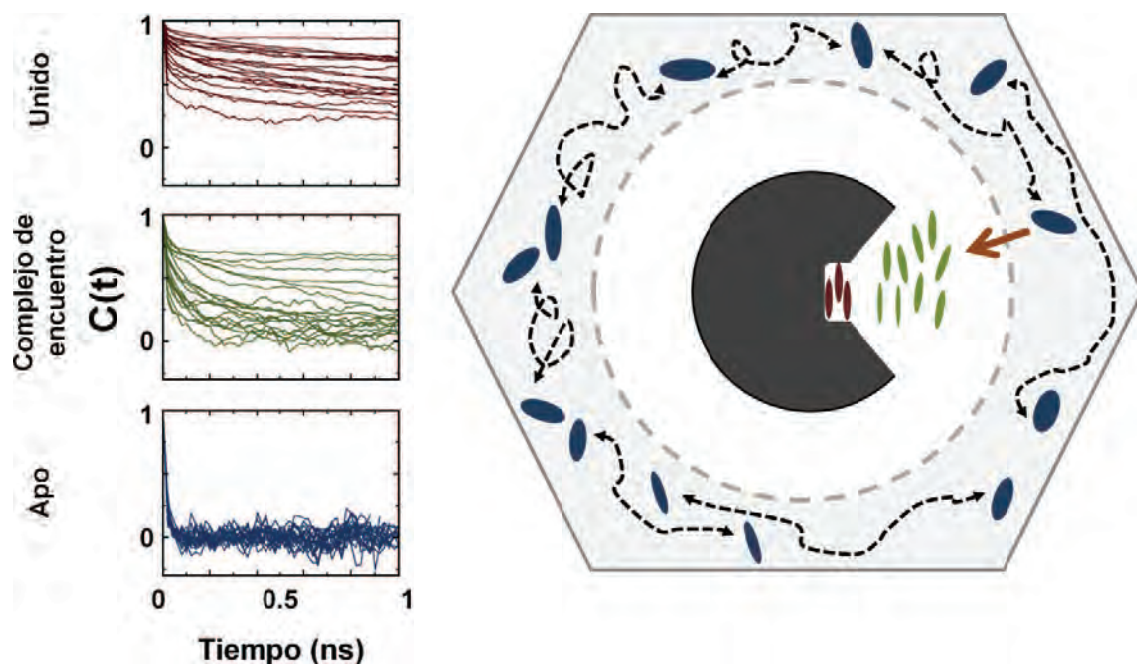


Figura 6. En la izquierda se muestra la función de autocorrelación rotacional del ligando para segmentos de MD en: el estado no unido (azul), el complejo de encuentro (verde) y el estado unido (rojo). Varias funciones de auto-correlación fueron generadas a partir de diferentes trayectorias y combinadas en un mismo gráfico. A la derecha se ejemplifica mediante un esquema que el ligando rota rápidamente cuando se encuentra lejos de la proteína (y por lo tanto puede difundir libremente), pero cuando el ligando se acerca y comienza a interactuar con la proteína su movimiento se restringe. Cuando construimos el MSM consideramos ambos escenarios de manera independiente (ver Métodos).

Después de eliminar los grupos vacíos el procedimiento generó 3730 micro-estados. De éstos, 3290 micro-estados pertenecen al agrupamiento de alta resolución y 440 provienen de los datos que fueron re-agrupados a una baja resolución. Para verificar que el modelo final de micro-estados es válido (Markoviano) se graficaron las escalas de tiempo implicadas en el modelo y se encontró que estas convergen en intervalo de separación de entre 4 y 6 ns (ver Figura 7), implicando que el modelo es Markoviano en intervalo de separación que caen en este rango. Por lo tanto se puede concluir que los micro-estados son suficientemente pequeños como para garantizar que las conformaciones en un mismo micro-estado se encuentran cinéticamente relacionadas.

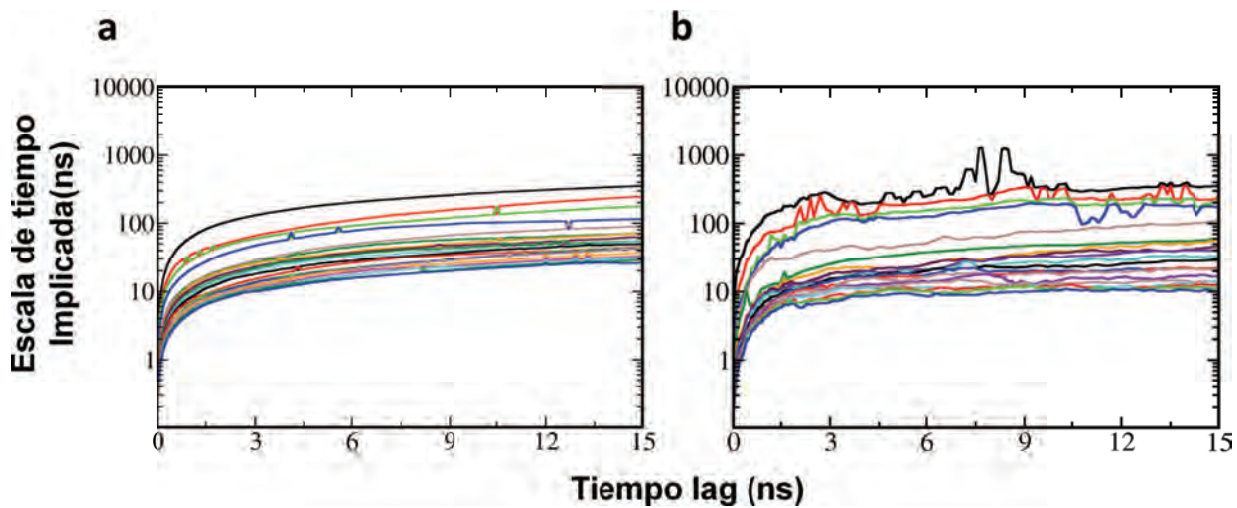


Figura 7. Se muestran las veinte escalas de tiempo implicado más lentas en: a) El modelo conformado por 3730 micro-estados y b) El MSM conteniendo 54 macro-estados. Ambos gráficos convergen en un intervalo de separación de ~ 4 ns, por lo tanto escogimos un tiempo ligeramente mayor (6 ns) para construir el MSM final.

En el paso siguiente, utilizando el algoritmo de SHC (37), agrupamos en macro-estados a aquellos micro-estados relacionados cinéticamente entre sí. En el SHC se realiza un agrupamiento espectral por jerarquías basadas en niveles de densidad, comenzando por el nivel de densidad más alto y por lo tanto garantizando que las regiones meta-estables con poblaciones altas son identificadas antes que las bajamente pobladas. Para el SHC seleccionamos niveles de densidad $L_{\text{high}} = [0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95, 0.99]$ y $L_{\text{low}} = [0.4, 0.95]$, para las regiones de alta y baja densidad, respectivamente. Debido a que los micro-estados tienen diferentes tamaños y por lo tanto es difícil comparar sus densidades, los macro-estados de

alta y baja resolución fueron agrupados de manera separada. Después, combinamos estos dos grupos de macro-estados para construir el MSM final con un total de 54 macro-estados. Una vez más utilizamos las escalas de tiempo implicadas para verificar que el modelo fuera Markoviano. De esta manera se comprobó que el MSM reproduce las escalas de tiempo implicadas en los micro-estados y que con un intervalo de separación de 6 ns el modelo tiene un comportamiento Markoviano (ver Figura 8).

Escalas de Tiempo Implicadas

Examinar el comportamiento de las escalas de tiempo implicadas es una manera de corroborar si el modelo es Markoviano (53), es decir que las transiciones entre los estados A y B se encuentran basadas solo en la probabilidad y que el estado anterior del sistema no es modifica esta probabilidad. Las escalas de tiempo implicadas (τ_k) se puede calcular como:

$$\tau_k = -\frac{\tau}{\ln \mu_k(\tau)} \quad (1)$$

donde μ_k es un eigenvalor de la matriz de transición dado el tiempo de separación τ (distancia temporal entre dos cuadros o estados de la simulación) . Cada escala de tiempo implicada describe una transición agregada entre una subpoblación de macro-estados. Si el modelo es Markoviano y la ecuación (1) se mantiene, la exponenciación del tiempo T debe ser idéntica si el MSM se construye con un tiempo de separación mayor, y por lo tanto las escalas de tiempo implicadas serán independientes del intervalo de separación.

En nuestro modelo las escalas de tiempo implicadas se estabilizan cuando el tiempo de separación es de ~ 4 ns, indicando que el modelo es Markoviano a partir de este intervalo de separación (ver Figura 7). Por lo tanto escogimos un tiempo de separación ligeramente mayor (6 ns) para calcular los parámetros termodinámicos en nuestro sistema.

Cálculo de las matrices de transición

Para calcular las matrices de transición utilizadas para generar la descomposición de los estados, se contó el número de transiciones entre cada par de estados a cierto intervalo de observación

(intervalo de separación = 6 ns). Con las cuentas se generaron las matrices de transición, en las que la celda en el renglón X y la columna Y representa las transiciones observadas desde el estado X hasta el estado Y. En particular utilizamos una ventana deslizante, con el intervalo de separación especificado, en cada una de las 65 trayectorias de 200ns en las cuales las conformaciones fueron almacenadas cada 20 ps (i.e. cada trayectoria contiene 10,000 conformaciones). Debido a que los puntos de corte entre los macro-estados son duros, las conformaciones en el límite de las barreras entre los macro-estados pueden oscilar de un estado a otro, llevando a sobreestimar la tasa de transición entre estados (54). Para mitigar el efecto de estos eventos de recruzamiento, sólo contamos las transiciones del estado X al estado Y si la conformación Y se mantiene por al menos 300 ps. Se comprobó que tiempos mayores a 300 ps generan estadísticas similares, pero disminuyen la cantidad de datos estadísticos ya que eliminan un mayor número de transiciones. Dado que 300 ps es apenas 1/20 del tiempo Markoviano, asumimos que es razonable utilizar esta escala temporal para filtrar el ruido del sistema. Para generar la matriz de probabilidades (donde un campo x,y de la matriz indica la probabilidad de transitar del estado X al Y en un tiempo igual al intervalo de separación elegido) normalizamos cada renglón de la matriz de transiciones.

Cálculo del tiempo promedio de la primera transición (MFPT)

Seguimos el procedimiento descrito en la Ref. (55) para calcular el MFPT del estado inicial i al estado final f. En particular, el MFPT (X_{if}) dado que la transición del estado i al j se realice primero, es el tiempo que toma llegar de i a j más el MFPT de la transición del estado j a f. Por lo tanto el MFPT puede ser definido como:

$$X_{if} = \sum_j P_{ij}(t_{ij} + X_{jf}) \quad (2)$$

donde $t_{ij} = 6 \text{ ns}$ es el intervalo de separación de la matriz de transición T. La condición de límite es:

$$X_{ff} = 0 \quad (3)$$

El conjunto de ecuaciones lineales definidos por (2) y (3) puede ser resuelto para obtener el MFPT X_{if} . Luego utilizamos un método de muestreo aleatorio (bootstrapping) para calcular el error estándar de los MFPTs. El muestreo aleatorio con remplazo estadístico (i.e. Que una misma trayectoria del conjunto de MDs puede ser considerada dos o más veces) generó cien conjuntos de datos, cada uno con 130 trayectorias. Entonces se calculó el MFPT para cada conjunto de datos y se reportó el promedio y la desviación estándar. En los cálculos de MFPT se consideró que el complejo de encuentro estaba compuesto por los estados 11 y 5 (debido a que el estado 5 también tiene las características del complejo de encuentro, pero juega un papel significativamente menor).

Cálculo de la energía libre de unión y la tasa de asociación

Basados en nuestro MSM calculamos la energía libre de unión (ΔG) y la tasa de asociación. Utilizamos el método desarrollado por van Gunsteren *et. al.* (56):

$$K_b = \frac{[P][L]}{[P \bullet L]} \quad (4)$$

$$\Delta G = -k_b T \ln K_b = -k_b T \ln \left(\frac{\alpha_{bound}}{c_0 \alpha_{free}} \right) \quad (5)$$

donde α_{bound} y α_{free} son las fracciones de las especies unidas y libres, respectivamente, y c_0 es la concentración total de ligando. En nuestras MD $c_0 = 1/(N_A V_{box})$, donde N_A es el número de Avogadro y V_{box} es el volumen de la caja de simulación. En nuestro sistema, $c_0 = 0.0049$ mol/L, $T = 318K$. Consideramos todos los estados no unidos como especies libres, de manera que $\alpha_{free} = 1.75\%$. Por lo tanto las especies unidas son: $\alpha_{bound} = 1 - \alpha_{free} = 98.25\%$, que contienen todos los estados donde la proteína y el ligando están en contacto cercano. De la Eq. (5) se deduce que:

$$\Delta G = -8.46 Kcal/mol$$

También podemos derivar la tasa de asociación. Dado que un sistema en equilibrio está descrito por la proteína P , el ligando L y el complejo proteína-ligando $P \bullet L$, la reacción de unión puede escribirse como:



La ecuación de velocidad puede escribirse como:

$$\frac{d[P]}{dt} = -k_{on}[P][L] + k_{off}[P \bullet L] \quad (7)$$

Donde k_{on} y k_{off} son las constantes de asociación y disociación respectivamente. Dado que la reacción de asociación sólo depende de la k_{on} , la ecuación de velocidad para un proceso de asociación se puede escribir como:

$$\frac{d[P]}{dt} = -k_{on}[P][L] \quad (8)$$

La concentración total de la proteína y el ligando en el sistema son constantes:

$$[P_{tot}] \stackrel{def}{=} [P] + [P \bullet L] \quad (9)$$

$$[L_{tot}] \stackrel{def}{=} [L] + [P \bullet L] \quad (10)$$

Por lo tanto, sólo una concentración entre $[P]$, $[L]$, y $[P \bullet L]$ es independiente. Si escogemos $[P]$ como la concentración independiente podemos reescribir la ecuación para una reacción de asociación como:

$$\frac{d[P]}{dt} = -k_{on}[P][L] = -k_{on}([L_{tot}] - [P_{tot}] + [P])[P] \quad (11)$$

Si consideramos que $[P_{tot}] = [L_{tot}]$, que es el caso para nuestras simulaciones:

$$\frac{d[P]}{dt} = -k_{on}[P]^2 \quad (12)$$

Podemos resolver la ecuación (11) con la condición inicial $[P] = [P_0]$ a $t=0$:

$$[P] = \frac{[P_0]}{k_{on}[P_0]t + 1} \quad (13)$$

Definimos el tiempo medio de asociación ($t_{1/2}$) como el tiempo en el que la mitad de las moléculas de proteína ($[P] = [P_0]/2$) se encuentran asociadas al ligando:

$$t_{1/2} = \frac{1}{k_{on}[P_0]} \quad (14)$$

Dado que no existen resultados experimentales para la constante k_{on} de LAO, entonces escogimos como comparación la proteína que une histidina (HisJ, ver Métodos: Racionalización de la comparación entre LAO e HisJ). Para HisJ $k_{on} = 1 \times 10^8 \text{ M}^{-1} \text{ sec}^{-1}$. En nuestras simulaciones, la concentración inicial de proteína es $[P_0] = 0.0049 \text{ mol/L}$, por lo tanto usando la ecuación (14) se puede deducir:

$$t_{1/2} = 2 \mu\text{s}$$

La tasa de asociación experimental es alrededor de ocho veces más lenta que la calculada de nuestro MSM. Sin embargo, es importante notar que k_{on} fue medida a 293 K (57), mientras que nuestras simulaciones fueron realizadas a 318 K y, por lo tanto, su cinética es más rápida. De manera que la diferencia entre la constante experimental y la desprendida de nuestras simulaciones será menor si se considera la misma temperatura. En el caso de la energía libre de unión el valor experimental fue medido a una temperatura aún menor (277 K) (33), por lo que es razonable asumir que la energía libre de unión a la temperatura de nuestras simulaciones tiene un valor aun más cercano al que hemos calculado. Sin embargo, dado que no conocemos el ΔC_p experimental de la asociación, no podemos realizar una estimación más precisa de la energía de unión experimental a 318 K.

Visualización

La visualización y generación de imágenes de los archivos de estructura PDB se realizó mediante el programa PyMol versión 0.99rc6. La visualización y generación de imágenes de las trayectorias de MD se realizó mediante el programa “Visual Molecular Dynamics” (VMD, versión 1.8).

Recursos computacionales

La MD y análisis de los datos se realizaron en los siguientes instalaciones de cómputo de alto desempeño:

- KanBalam, Universidad Nacional Autónoma de México, DGSCA. Procesadores AMD Opteron, 2.6 GHz.
- Aitzaloo, Universidad Autónoma Metropolitana, Campus Iztapalapa. Procesadores Intel Xeon, 3.0 GHz
- Entalpia, Universidad Nacional Autónoma de México, Facultad de Medicina. Procesadores Intel Xeon, 2.5 GHz.
- CCB computational cluster, Hong Kong University of science and Technology. Procesadores Intel Xeon, 2.5 GHz.

Resultados

Los resultados se han dividido en dos secciones complementarias:

En la primera parte, el objetivo principal es el estudio de la dependencia entre el paisaje conformacional de LAO y la presencia de los distintos ligandos en el sitio de unión. De este estudio se deriva la metodología esencial para analizar las simulaciones de MD en términos del estado conformacional de LAO y de sus interacciones con el ligando.

La segunda parte tiene como objetivo el estudio del mecanismo *ab initio* de la unión del ligando.

Parte I

Estabilidad térmica del sistema y análisis mediante ángulos diedros

En teoría, un incremento en la temperatura del sistema podría facilitar el escapar de mínimos energéticos al desestabilizar ciertas interacciones entálpicas. Con el fin de muestrear más rápidamente el espacio conformacional de LAO, realizamos un estudio del efecto de la temperatura en la estabilidad de la proteína. La lógica fue encontrar la temperatura máxima a la que la estructura secundaria de LAO permanece estable. Experimentos de desnaturalización térmica (ver Figura 9), muestran que la desnaturalización de LAO comienza a temperaturas superiores a 318 K. De manera que generamos simulaciones de dinámica molecular a 318 K, de la proteína en el estado abierto y libre de ligando. El análisis de la estructura secundaria confirmó estabilidad a dicha temperatura (ver Figura 9).

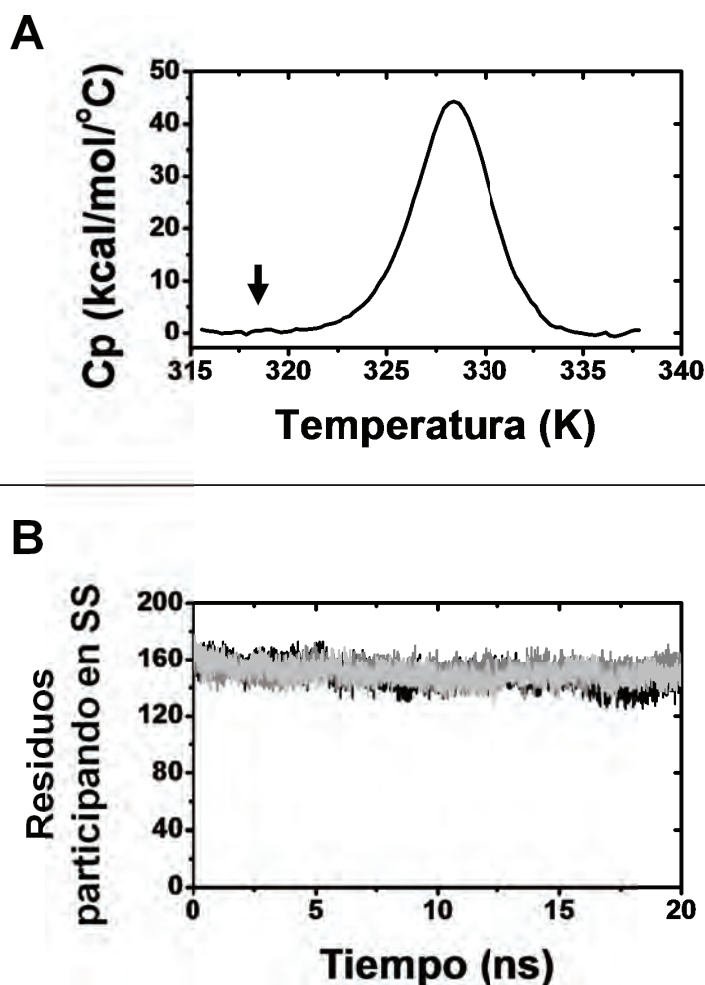


Figura 9. A) Experimento de desnaturalización térmica efectuado en apo-LAO mediante calorimetría diferencial de barrido (a una tasa de calentamiento de $1^{\circ}\text{C min}^{-1}$ 0.5 mg/ml en amortiguador HEPES 10 mM). La flecha indica 318 K, en donde la transición del desplegamiento aun no ha comenzado. B) La estabilidad de la estructura secundaria (SS) de apo-LAO en tres replicas de MD para apo-LAO a 318 K. Los aminoácidos fueron considerados como participando en SS si el programa DSSP los consideró como parte de cualquiera de las siguientes estructuras: "hoja- β , puente- β , loop ó helice- α ".

LAO presenta esencialmente dos cambios globales de gran escala: la apertura (también conocido como "opening") y un movimiento de dislocación entre los lóbulos (también conocido como "twisting"). Con el objetivo de medir estos movimientos, se generó un análisis basado en ángulos diedros (ver Métodos). Los planos de los ángulos diedros (Dih-apertura y Dih-dislocación, para la apertura y la dislocación, respectivamente) fueron definidos utilizando los centros de masa de cuatro secciones de la proteína, que esencialmente describen los lóbulos (I y II) y la bisagra (ver Figura 4). Los resultados se compararon en simulaciones iniciadas en el estado cerrado de apo-

LAO, con los movimientos descritos por los dos primeros modos normales de un análisis de componentes principales (PCA) (ver Figura 4). Por definición, estos ángulos diedros toman un valor de 0° para la estructura cristalográfica/minimizada del estado cerrado, mientras que los valores para la estructura cristalográfica del estado abierto son: Dih-apertura = 38.2 y Dih-dislocación = -26.2.

LAO en ausencia del ligando (apo-LAO)

Se realizaron MD comenzando con las estructuras apo-LAO-abierta y apo-LAO-cerrada. Se encontró que, independientemente de la estructura inicial, el estado alcanzado por la proteína es un estado abierto (ver Figura 10). Sin embargo, en términos de dislocación, se encontraron dos estados: uno similar al estado abierto cristalográfico y otro con aproximadamente 20° más de dislocación (que denominamos estado abierto-dislocado) (ver Figura 10 y Tabla 1).

Tabla 1. Ángulos y distancias del ligando calculados en MD iniciadas en la conformación cerrada de LAO, bajo distintas condiciones iniciales. Los valores corresponden a los promedios durante los últimos 10ns de 23 trayectorias independiente para cada conformación inicial. Se muestra la desviación estándar. Los dos primeros renglones muestran los valores para las estructuras cristalográficas de la proteína la conformación abierta ó cerrada (PDB ID: 2LAO y 1LAG, respectivamente).

Conformación Inicial	Conformación alcanzada	dih-apertura (°)	dih-dislocación (°)	lig-dist (Å)
<i>Abierta (Cristalográfica)</i>	–	38.22	-26.22	–
<i>Cerrada (Cristalográfica)</i>	–	0.0	0.0	0.0
<i>apo-abierta⁽¹⁾</i>	<i>Abierta</i>	36.95 ± 10.1	7.19 ± 14.4	–
<i>apo-cerrada⁽¹⁾</i>	<i>Abierta</i>	32.56 ± 9.5	-2.67 ± 10.1	–
	<i>Abierta-dislocada</i>	30.58 ± 6.0	28.06 ± 9.6	–
<i>arg- cerrada⁽¹⁾</i>	<i>Cerrada</i>	10.97 ± 9.7	10.75 ± 9.4	0.005 ± 0.220
<i>lys- cerrada⁽¹⁾</i>	<i>Cerrada</i>	12.12 ± 8.2	15.06 ± 10.5	0.113 ± 0.554
<i>his-H- cerrada⁽¹⁾</i>	<i>Cerrada</i>	14.50 ± 11.2	9.64 ± 10.7	0.508 ± 0.873

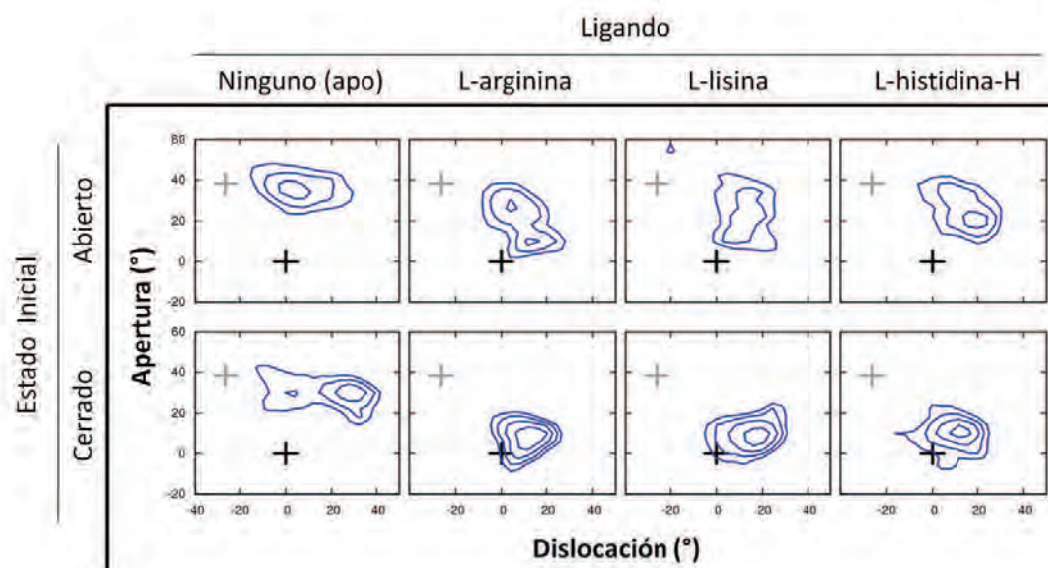


Figura 10. Los mapas de contorno del espacio conformacional (apertura vs. dislocación) visitados por LAO en el conjunto de 23 replicas para cada condición inicial y ligando. Para facilitar la comparación, la escala está transformada mediante el negativo del logaritmo natural. Para cada eje el tamaño de los cajones del histograma es 4° . El punto de corte es $-\log P = 5$ y la separación entre contornos es de -0.5 . Como referencia, las cruces grises y negras muestran los valores para los estados cristalográficos abierto y cerrado, respectivamente.

El análisis de los puentes de hidrógeno reveló una interacción específica del estado abierto entre los aminoácidos 78 y 188. De la misma manera, el estado abierto-dislocado reveló dos interacciones específicas: 76_188 y 77_190 (Tabla 2).

Tabla 2. Fracción de existencia de puentes de hidrógeno importantes. Se muestra la fracción promedio de puentes de hidrógeno (H-bonds) en trayectorias de MD que fueron iniciadas en los estados apo-abierto, apo-cerrado u holo-cerrado (distintos ligandos). Los datos provienen de los últimos 10ns de 23 trayectorias independientes para cada conformación inicial. Solo se muestran los puentes de hidrógeno que se determinó muestran cambios significativos entre las distintas conformaciones/condiciones estudiadas (ver Métodos).

		Conformación (Inicial → Final)					
		Apo-abierta ↓ Apo-abierta	Apo-cerrada ↓ Apo-abierta/o/ abierta-dislocada	L-arg-cerrada ↓ L-arg-cerrada	L-lys-cerrada ↓ L-lys-cerrada	L-his-H-cerrada ↓ L-his-H-cerrada	
Puentes de Hidrógeno Característico	<i>Del estado abierto</i>	78 - 188	0.195	0.110	0.035	0.049	0.0722
	<i>Del estado dislocado</i>	76 - 188	0.093	0.384	0.00004	0.070	0.004
		77 - 190	0.039	0.230	0.016	0.113	0.051
	<i>Del estado cerrado</i>	13 - 143	0.0008	0.002	0.066	0.040	0.199
		15 - 143	0.052	0.243	1.150	1.330	0.683
	<i>De la proteína con el Ligando</i>	11 - LIG	–	–	0.797	0.140	0.204
		13 - LIG	–	–	0.078	0.136	0.026
		69 - LIG	–	–	0.403	0.156	0.049
		70 - LIG	–	–	1.060	0.621	0.533
		72 - LIG	–	–	1.400	1.130	1.060
77 - LIG		–	–	1.370	0.969	0.989	
119 - LIG		–	–	0.163	0.075	0.014	
121 - LIG		–	–	1.270	1.620	1.560	
161 - LIG	–	–	0.078	0.091	0.633		

Para investigar si la conformación abierta-dislocada pertenece a un intermediario entre el estado cerrado y abierto, se estudiaron los vectores de movimiento de las simulaciones iniciadas en el estado apo-cerrado. Ninguna simulación mostró el tránsito a través del estado abierto-dislocado para llegar al estado abierto (equivalente al cristalográfico); de hecho, algunas simulaciones alcanzaron primero el estado abierto y luego transitaron al abierto dislocado (ver Figura 11).

Estos resultados sugieren que el estado abierto-dislocado no es un intermediario en el mecanismo de apertura de la proteína. La presencia de tal conformación en nuestras simulaciones podría estar relacionada con un efecto de histéresis, causado por iniciar las simulaciones en una conformación apo-cerrada, que, como se deduce de su inestabilidad en la ausencia del ligando (y se demostrará en repetidas ocasiones a lo largo de este trabajo), es una conformación de alta energía.

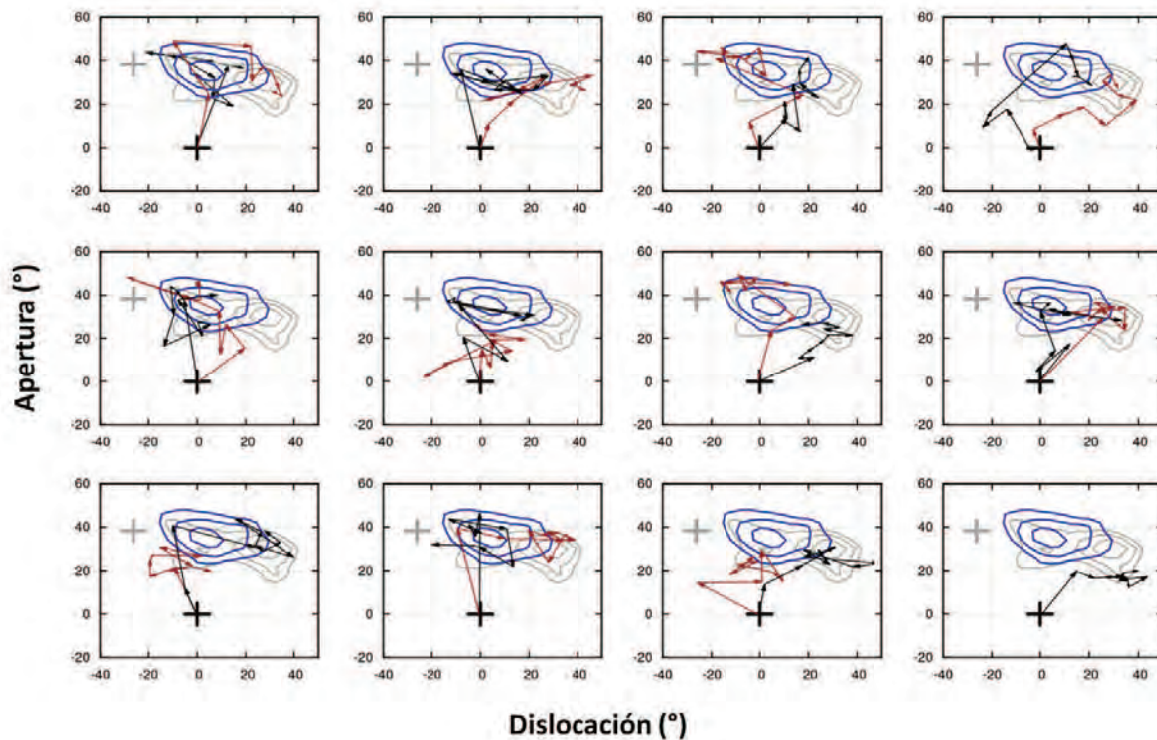


Figura 11. Proyección de las rutas conformacionales de desplazamiento entre el estado cerrado y los distintos estados abiertos encontrados: abierto (semejante al cristalográfico, contorno color azul) y abierto-dislocado (parte derecha del contorno color gris). Los vectores de movimiento (flechas) muestran la ruta seguida en la transición del estado cerrado al abierto. Se muestran dos MD en cada gráfico (flechas roja y negra). Como referencia, los mapas de contorno de las conformaciones abiertas alcanzadas por simulaciones comenzadas en el estado abierto (azul) o cerrado (gris) (los mismos que los mostrados en la Figura 10). El paso de tiempo entre los puntos de los vectores mostrados es 2 ns. Las gráficas revelan que existen dos caminos para alcanzar el estado abierto dislocado: uno directo y otro a través del estado abierto. En ningún caso se encontró evidencia de que el estado abierto-dislocado sea un intermediario entre el estado cerrado y el estado abierto.

LAO en presencia de los ligandos nativos (holo-LAO)

Con el fin de estudiar el efecto de los distintos ligandos en la dinámica de la proteína LAO, realizamos simulaciones de MD en presencia de los “ligandos nativos” (L-arg, L-lys y L-his) . En

el caso de la histidina utilizamos la molécula con carga positiva (L-his-H). En las conformaciones iniciales del estado cerrado, el ligando se posicionó en el sitio de unión de acuerdo a las estructuras cristalográficas (ver Métodos). Para las conformaciones iniciales del estado abierto, el ligando se posicionó cercano al lóbulo I en una conformación lo mas similar posible a la unida (ver Métodos).

Simulaciones del estado cerrado de LAO

En presencia de cualquiera de los ligandos nativos la conformación cerrada permaneció estable (ver Figura 10) y el ligando permaneció unido a la proteína (Tabla 1). Sin embargo, se encontraron distintas combinaciones de los ángulos de apertura y dislocación (ver Figura 10 y Tabla 1), así como de patrones de puentes de hidrógeno (Tabla 2) para cada complejo proteína-ligando. Las cadenas laterales de los ligandos L-arg y L-his-H mostraron una interacción de catión- π con los residuos aromáticos Tyr-14 y Phe-52 (ver Tabla 3 y Figura 12). La interacción mostro ser más débil con el ligando L-his-H que con L-arg. En el caso del ligando L-lys no fue posible medir una interacción de apilamiento, dado que la cadena lateral de la lisina no presenta una conformación planar, pero el análisis de las conformaciones centrales sugiere que la cadena lateral del ligando L-lys alterna una interacción hidrofóbica y otra de catión- π con los ligandos Tyr-14 y Phe-52 (i.e. que interactúa a través de una interacción de catión- π ya sea con Tyr-14 o Phe-52 y la cadena hidrocarbonada establece una interacción hidrofóbica con el otro aminoácido) (ver Figura 12).

Tabla 3. Apilamiento del ligando con los residuos Tyr-14 y Phe-52 en las simulaciones iniciadas en el estado cerrado. La tabla resume la fracción promedio (P) de la interacción de apilamiento (ver Métodos) para simulaciones iniciadas en el estado cerrado. Los cuadros grises muestran a los tautómeros de la histidina con carga neutra. Los valores provienen del análisis de los últimos 10ns de las 23 trayectorias para cada condición.

	L-arg cerrada	L-his-H cerrada	L-his-ND1 cerrada	L-his-ND2 cerrada
Apilamiento <i>Lig <-> Tyr_14</i>	0.59	0.65	0.19	0.14
<i>Lig <-> Phe_52</i>	0.80	0.45	0.17	0.15

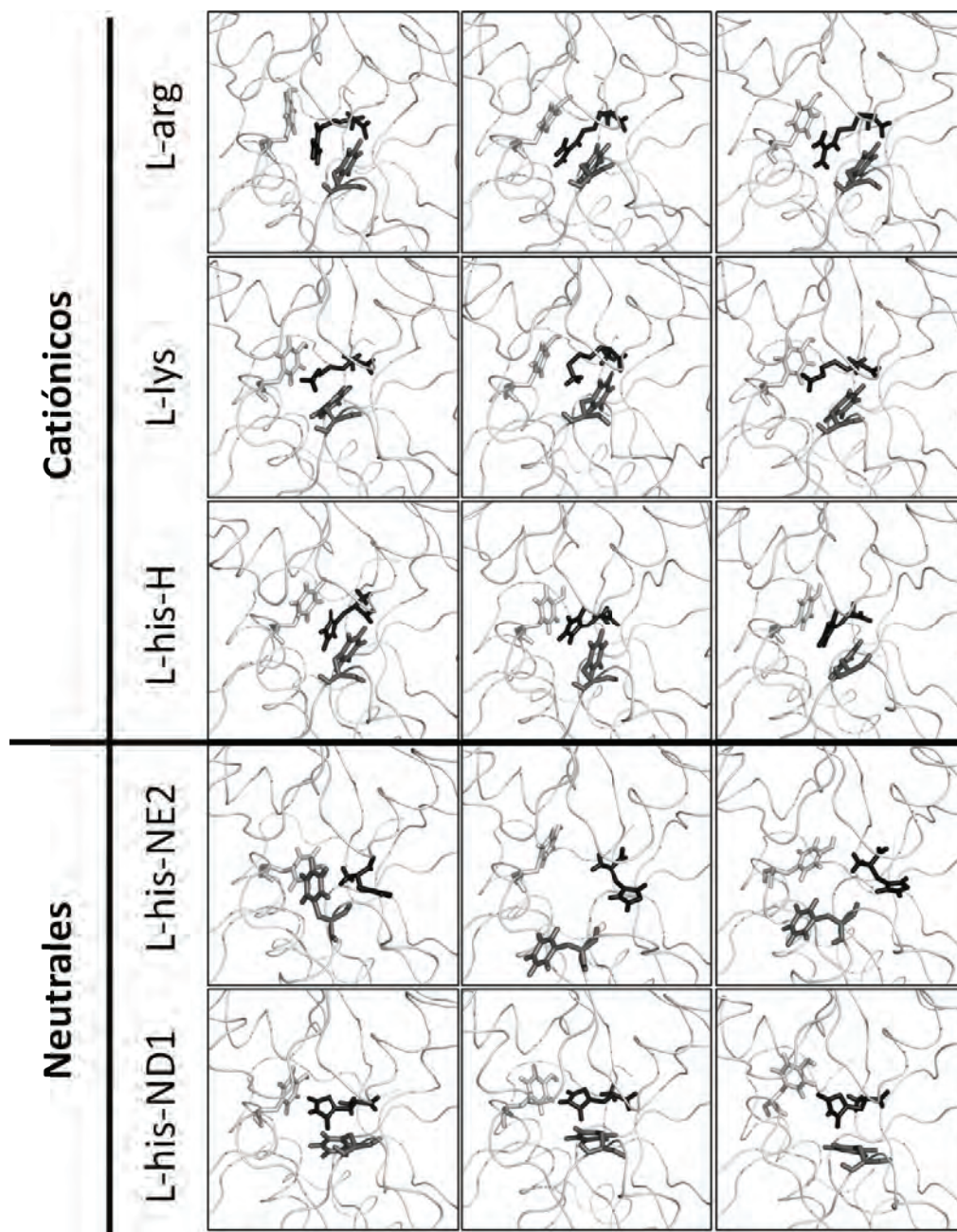


Figura 12. Interacción del ligando con los residuos Tyr-14 y Phe-52 de la proteína LAO. La proteína se muestra en una representación de listones; el ligando en color negro; los residuos Tyr-14 y Phe-52 en gris claro y oscuro, respectivamente. Para evitar introducir un sesgo, las conformaciones fueron seleccionadas de manera aleatoria (tres conformaciones se muestran para cada ligando). Las estructuras se escogieron del mayor conjunto/o/grupo de estructuras de un agrupamiento de centroides-K sobre la distancia euclidiana de la apertura y la dislocación. Se generaron 20 conjuntos para cada ligando, la población total (número de cuadros muestreados fue de 92,000 para cada ligando, y la población total de los agrupamientos mayores fue: L-arg= 22,065, L-lys=34,933, L-his-H= 31,190, L-his-ND1=25,486, y L-his-NE2=21,850. Esto demuestra que los agrupamientos tienen una población significativa y homogénea que garantiza su fiabilidad estadística.

Con el fin de entender si la interacción con el ligando forma un acoplamiento entre Tyr-14 y Phe-52, realizamos un estudio del ángulo entre los planos de estos residuos (ver Métodos). Los resultados, analizados mediante histogramas, muestran que los ligandos nativos inducen la formación de poblaciones bien definidas en los ángulos, más aún, existen distribuciones específicas para cada tipo de ligando (ver Figura 13). Por el contrario, cuando la proteína se encuentra libre de ligando, la distribución es fundamentalmente plana, sugiriendo una alta movilidad (no correlacionada) de los residuos Tyr-14 y Phe-52 (ver Figura 13). En conjunto, estos resultados sugieren que existe un acoplamiento entre los aminoácidos Tyr-14 y Phe-52 mediado por su interacción con la cadena lateral del ligando nativo.

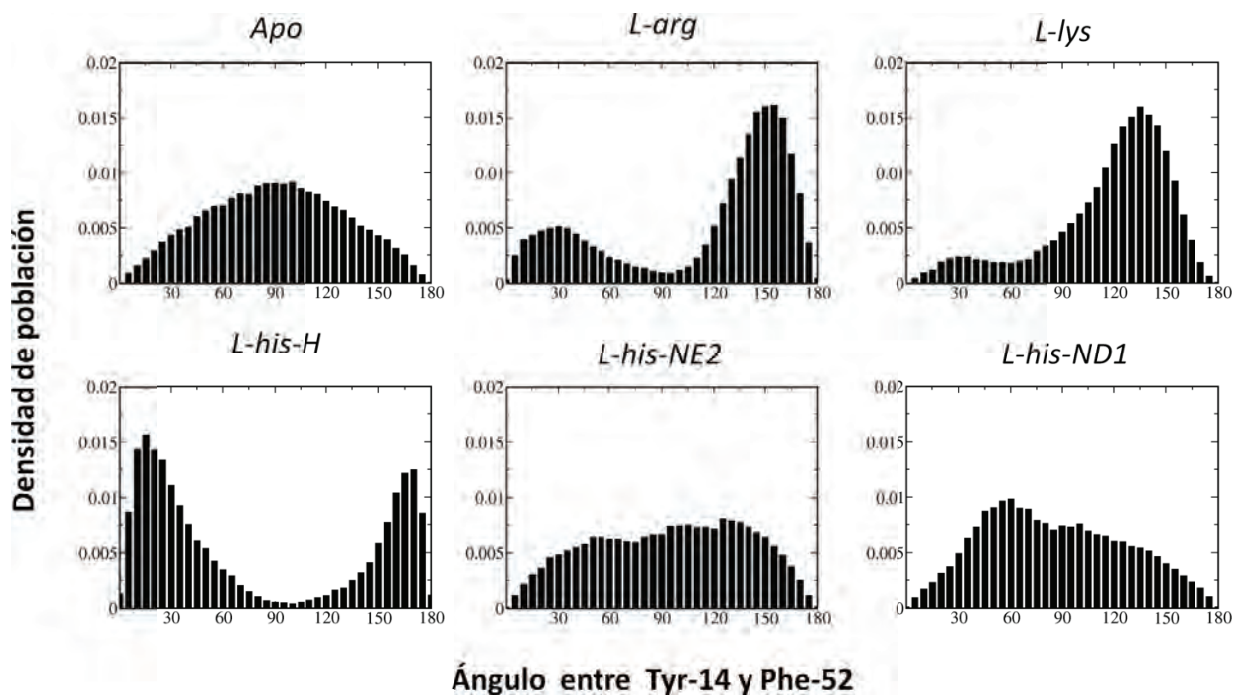


Figura 13. Histogramas de los ángulos entre los anillos de los residuos Tyr-14 y Phe-52 para MDs iniciadas en el estado cerrado. Los histogramas muestran la distribución de los ángulos en la ausencia y presencia de distintos ligandos. Los datos de cada grafica corresponden a los últimos 10ns de 23 trayectorias de MD iniciadas en la condición indicada.

Realizamos una prueba estadística de t de Student para determinar si los angulos de apertura y dislocacion son distintos para la proteína en presencia o ausencia del ligando, asi como para

estudiar si existe alguna diferencia entre los distintos ligandos. Los resultados mostraron que el espacio conformacional es distinto entre el estado apo-abierto y holo-cerrado (en presencia de cualquiera de los ligandos) (Tabla 4). Por otra parte, el ángulo de dislocación es diferente entre holo-LAO-lys y cualquiera de los otros ligandos, mientras que no existe diferencia significativa entre los espacios conformacionales para holo-LAO-his-H y holo-LAO-arginina (Tabla 4).

Tabla 4. Comparación estadística de las simulaciones de MD iniciadas en diferentes estados conformacionales y ligandos. La tabla muestra el resultado de una "prueba estadística de t" para comparar independientemente los ángulos de apertura (tabla superior) y dislocación (tabla inferior) dadas las diferentes condiciones iniciales. El test fue de dos colas y la hipótesis alternativa: "que la diferencia entre las medias no es igual a cero". La varianza fue estimada independientemente para cada grupo y empleamos la modificación de Welch a los grados de libertad. Los datos analizados correspondieron a 23 observaciones para cada condición simulada (los promedios de los últimos 10 ns de cada una de las 23 MD). Los campos en gris denotan cuando la diferencia entre las poblaciones mostro ser estadísticamente relevante a un nivel de significancia del 95%.

Diedro de Apertura		Inicialmente abierta	Inicialmente cerrada		
		<i>(Apo)</i>	<i>(Apo)</i>	<i>L-arg</i>	<i>L-lys</i>
Inicialmente cerrada	<i>(Apo)</i>	0.022	-	-	-
	<i>L-arg</i>	3.9 ⁻¹⁴	7.8 ⁻¹²	-	-
	<i>L-lys</i>	1.9 ⁻¹⁵	1.4 ⁻¹³	0.894	-
	<i>L-his-H</i>	3.7 ⁻¹⁰	7.9 ⁻⁰⁸	0.253	0.260

Diedro de Dislocación		Inicialmente abierta	Inicialmente cerrada		
		<i>(Apo)</i>	<i>(Apo)</i>	<i>L-arg</i>	<i>L-lys</i>
Inicialmente cerrada	<i>(Apo)</i>	0.079	-	-	-
	<i>L-arg</i>	0.258	0.241	-	-
	<i>L-lys</i>	0.011	0.940	0.0318	
	<i>L-his-H</i>	0.552	0.141	0.521	0.014

Mediante el estudio de los puentes de hidrógeno, encontramos que cada ligando produce patrones específicos de interacción. Específicamente, la interacción entre los residuos 15 y 143 se encontró en el complejo holo-cerrado (en presencia de cualquier ligando nativo). Adicionalmente, en presencia del ligando L-his-H se encontró una interacción entre los residuos 13 y 143 (Tabla 2). Como se discutió anteriormente, la conformación abierta-dislocada mostró una interacción de puente de hidrógeno entre los aminoácidos 77 y 190. De la misma manera, en presencia del ligando L-lys la proteína alcanzó un ángulo de dislocación significativamente mayor que con los otros ligandos, al mismo tiempo esta conformación mostró la interacción de puentes de hidrógeno entre los aminoácidos 77 y 190 (Tabla 2). Esta correlación plantea la posibilidad de que dicha interacción establezca de manera específica una conformación con mayor grado de dislocación.

Simulaciones del estado abierto de LAO

Utilizando como ligandos a L-arg y L-lys fuimos capaces de capturar varias transiciones del estado abierto al estado cerrado. Para realizar estos experimentos, el ligando fue previamente localizado cerca del sitio de unión de lóbulo I de la proteína LAO en el estado abierto y la conformación inicial del ligando fue esencialmente idéntica a la conformación que se encuentra en el estado cerrado cristalográfico (ver Métodos). Esta elección fue fundamentada en el trabajo de Oh et. al. (23), que sugieren que la mayor parte de las interacciones del ligando con la proteína tienen lugar con el lóbulo I.

De las 23 simulaciones realizadas para cada ligando, se encontraron 8 trayectorias con transiciones espontáneas al estado holo-cerrado para L-arg y 5 para L-lys, mientras que para el ligando L-his-H, no encontramos ninguna transición completa al estado cerrado (se encontraron 3 transiciones a un estado parcialmente cerrado).

Con el fin de estudiar el mecanismo de la transición del estado abierto al cerrado, generamos un agrupamiento de centroides-k (ver Métodos) para aquellas simulaciones con transiciones espontáneas al estado cerrado. El agrupamiento fue generado de manera independiente para los ligandos L-arg y L-lys. Nuestro análisis (ver Métodos) dividió las transiciones para L-arg y L-lys en 4 grupos (estados) que describen el cambio conformacional del estado abierto al estado cerrado (ver Figura 14). Una vez que los datos fueron agrupados, se extrajeron 100 estructuras aleatorias de cada estado, que entonces fueron usadas para analizar la fracción de puentes de hidrógeno existentes en cada estado-ligando (Tabla 5 y Tabla 6).

Tabla 5. Fracción de existencia de puentes de hidrógeno importantes en la transición del estado abierto al cerrado para el ligando L-arg. La tabla resume la fracción promedio de puentes de hidrógeno. Los datos fueron analizados en 100 estructuras aleatorias de cada uno de los agrupamientos en la transición del estado abierto al cerrado (ver Figura 14 y Métodos). Los resultados muestran la progresión de la formación y ruptura de los puentes de hidrógeno desde la conformación holo-abierta (estado 1) hasta el estado holo-cerrado (estado 4).

L-arg-Abierta → Cerrada

		Estado 1	Estado 2	Estado 3	Estado 4	
Puente de hidrógeno característico	<i>Del estado abierto</i>	78 - 188	0.32	0.31	0.13	.03
	<i>Del estado abierto y dislocado</i>	76 - 188	0.0	0.0	0.0	0.0
		77 - 190	0.0	0.0	0.0	0.002
	<i>Del estado cerrado</i>	13 - 143	0.0	0.0	0.19	0.06
		15 - 143	0.0	0.0	0.44	0.70
	<i>De la interacción con el ligando</i>	11 - LIG	0.82	0.94	0.92	0.93
		13 - LIG	0.76	0.04	0.07	0.12
		69 - LIG	0.22	0.10	0.08	0.06
		70 - LIG	1.04	1.16	1.34	1.70
		72 - LIG	1.42	1.32	1.29	1.06
77 - LIG		0.89	0.94	0.74	0.80	
119 - LIG		0	0.02	0.09	0.34	
121 - LIG		0	0.54	1.58	1.49	
161 - LIG	0.01	0.02	0.02	0.03		

Tabla 6. Fracción de existencia de puentes de hidrógeno importantes en la transición del estado abierto al cerrado para el ligando L-lys. La tabla resume la fracción promedio de puentes de hidrógeno. Los datos fueron analizados en 100 estructuras aleatorias de cada uno de los agrupamientos en la transición del estado abierto al cerrado (ver Figura 14 y Métodos). Los resultados muestran la progresión de la formación y ruptura de los puentes de hidrógeno desde la conformación holo-abierta (estado 1) hasta el estado holo-cerrado (estado 4).

L-lys-Abierta → Cerrada

		State 1	State 2	State 3	State 4	
Puentes de hidrógeno característico	<i>Del estado abierto</i>	78 - 188	0.14	0.13	0.07	0.01
	<i>Del estado abierto y dislocado</i>	76 - 188	0.0	0.12	0.02	0.0
		77 - 190	0.0	0.03	0.0	0.0
	<i>Del estado cerrado</i>	13 - 143	0.0	0.0	0.12	0.02
		15 - 143	0.0	0.0	0.73	1.12
	<i>De la interacción con el ligando</i>	11 - LIG	0.04	0.16	0.23	0.26
		13 - LIG	0.03	0.11	0.05	0.04
		69 - LIG	0.0	0.0	0.12	0.29
		70 - LIG	0.57	0.37	0.49	0.49
		72 - LIG	0.98	0.78	1.17	1.28
77 - LIG		0.60	0.41	1.23	0.77	
119 - LIG		0.0	0.0	0.05	0.01	
121 - LIG		0.56	0.99	1.45	1.81	
161 - LIG	0.0	0.01	0.03	0.08		

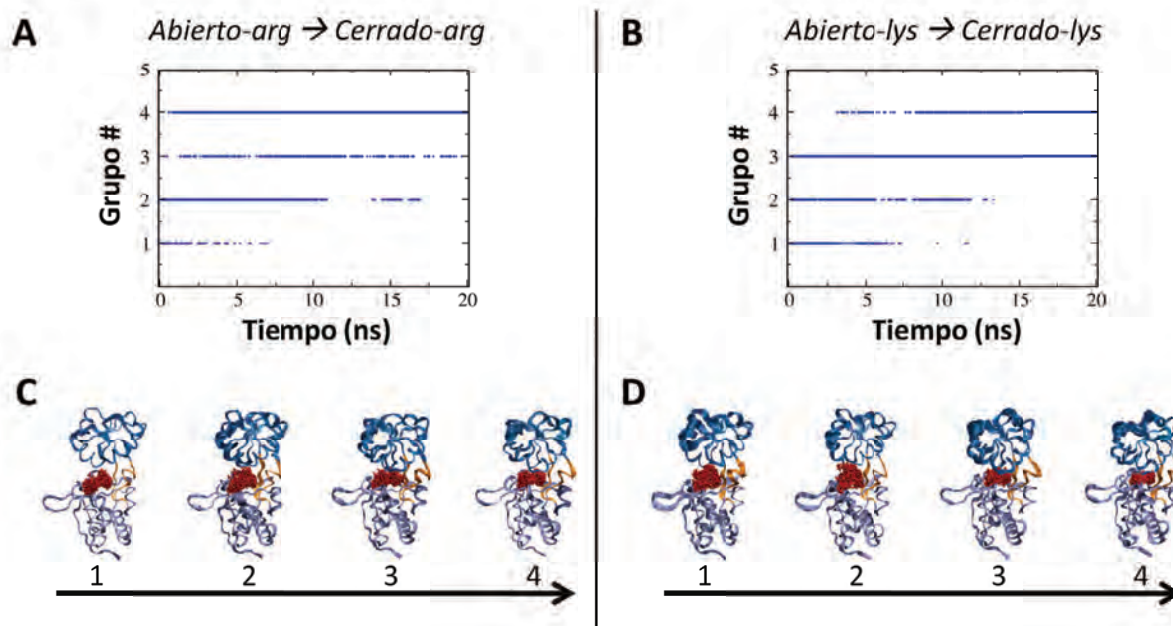


Figura 14. Se muestra la progresión temporal de los cuatro grupos (clusters) de estructuras identificados como representativos (y secuenciales) en las transiciones espontáneas del estado abierto al cerrado para el ensamble cuando el ligando fue: A) L-arg (8 trayectorias) ó B) L-lys (5 trayectorias). Los paneles C) y D) muestran una representación de listones de la proteína para cada uno de los cuatro grupos estructurales. Cada imagen fue generada mediante el sobrelapamiento de 100 estructuras aleatorias de cada grupo (el ligando se muestra en color rojo). Los grupos (1, 2, 3 y 4) se encuentran ordenados de manera progresiva (i.e. 1->2->3->4), donde 1 representa al estado abierto, 4 es el estado cerrado (y unido), mientras que 2 y 3 son intermediarios en el cambio conformacional.

Nuestros resultados sugieren que cuando ocurre una transición del estado abierto al cerrado, la formación de los puentes de hidrógeno se da bajo el siguiente patrón:

1. El ligando genera interacciones con el lóbulo I (residuos 11, 70, 72 ó 77). En el caso específico del ligando L-lys se forma una interacción con el lóbulo II (residuo 121).
2. Incremento o formación de la interacción con el residuo 121.
3. Decremento de la “interacción del estado abierto” (a.a. 78 y 188) y formación de la interacción específica del estado cerrado (a.a. 15 y 143).
4. Pérdida de la interacción del estado abierto e intensificación de las interacciones formadas en los pasos 1, 2 y 3.

Simulaciones en presencia de tautómeros neutros de la histidina.

Estudiamos el efecto del estado de protonación del ligando L-histidina en la dinámica de LAO. Entre los ligandos de LAO, L-his es el único para el cual el estado de protonación podría jugar un papel biológico relevante. Experimentos previos sugieren que en efecto, el estado de protonación de L-his podría estar relacionado con la afinidad, no obstante explicar las bases moleculares de este fenómeno presenta varios retos experimentales. Para abordar esta pregunta, comparamos el espacio conformacional visitado en presencia del ligando con carga positiva L-his-H con los tautómeros de carga neutra (L-his-ND1 [protón unido al N δ] y L-his-NE2 [protón unido al N ϵ]). Contrario a lo observado para el tautómero con carga positiva, los tautómeros con carga neutra presentaron varios eventos de desestabilización del estado cerrado (ver Figura 15), y una fracción de estos eventos fueron seguidos por la pérdida de la interacción con el ligando. En el caso del tautómero L-his-NE2, 3/8 de los eventos de desestabilización fueron seguidos por la pérdida de interacción con el ligando. Para el ligando L-his-ND1, 4/10 de los eventos siguieron el mismo patrón. El análisis de la interacción de apilamiento (ver Figura 12 y Tabla 3) muestra que el efecto de cualquiera de los tautómeros sin carga es similar al encontrado en MDs en ausencia del ligando. En conjunto, estos resultados sugieren que la cadena lateral del ligando L-his sólo puede interactuar correctamente con los residuos 14 y 52 si su carga es positiva (interacción de catión- π), de lo contrario esta interacción se ve afectada y puede producir la desestabilización del estado holo-cerrado.

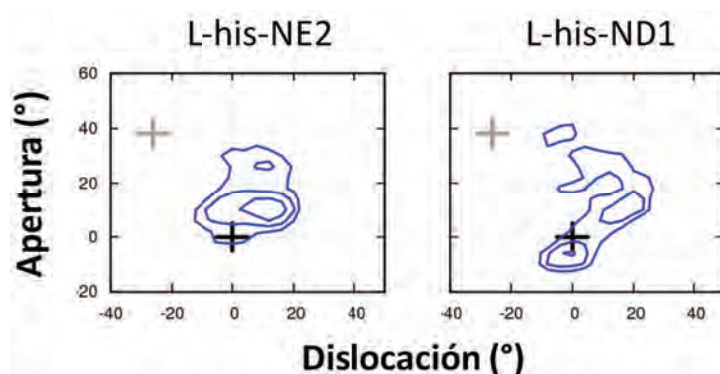


Figura 15. Los gráficos de densidad muestran usando contornos el espacio conformacional (apertura vs. dislocación) visitado por los ensambles de 23 replicas para cada los tautómeros con carga neutra de la L-his: L-his-NE2 y L-his-ND1. Los gráficos muestran, (en una clara diferencia con el ligando con carga positiva L-his-H), que los tautómeros neutros fallan en mantener la conformación cerrada de la proteína. Para facilitar la comparación, la escala está transformada mediante el negativo del logaritmo natural. Para cada eje el tamaño de las clases del histograma es 4°. El punto de corte es $-\ln P=5$ y la separación entre contornos es de -0.5. Como punto de comparación las cruces grises y negras muestran los valores para los estados cristalográficos abierto y cerrado, respectivamente.

Parte II

MSM para estudiar el mecanismo de unión de LAO.

Utilizando un conjunto de MDs no restringidas de la proteína LAO se construyó un modelo de Markov (MSM) del mecanismo de unión del ligando L-arg (ver Métodos). Brevemente, se generó un conjunto inicial de 10 MDs en solvente explícito, utilizando como punto inicial a LAO en el estado abierto con el ligando L-arg (no unido) a una distancia >25 Å del sitio de unión. Luego, 55 conformaciones aleatorias fueron extraídas de las trayectorias de este conjunto inicial y utilizadas para iniciar un segundo conjunto de MDs. El total de 65 MDs (200ns c/u) conformaron el conjunto de datos. Las conformaciones fueron almacenadas cada 20 ps (650,000 conformaciones en total). Finalmente, utilizando el algoritmo de agrupamiento por jerarquías (SHC) previamente descrito por Huang *et. al.* (37) (ver Métodos) se construyó un MSM de 54 estados. Las coordenadas de reacción para construir el modelo fueron: ángulo de apertura, dislocación y la posición relativa del ligando (ver Métodos). En la Figura 16 se muestran los estados conformacionales con poblaciones significativas al equilibrio (> 0.1 %).

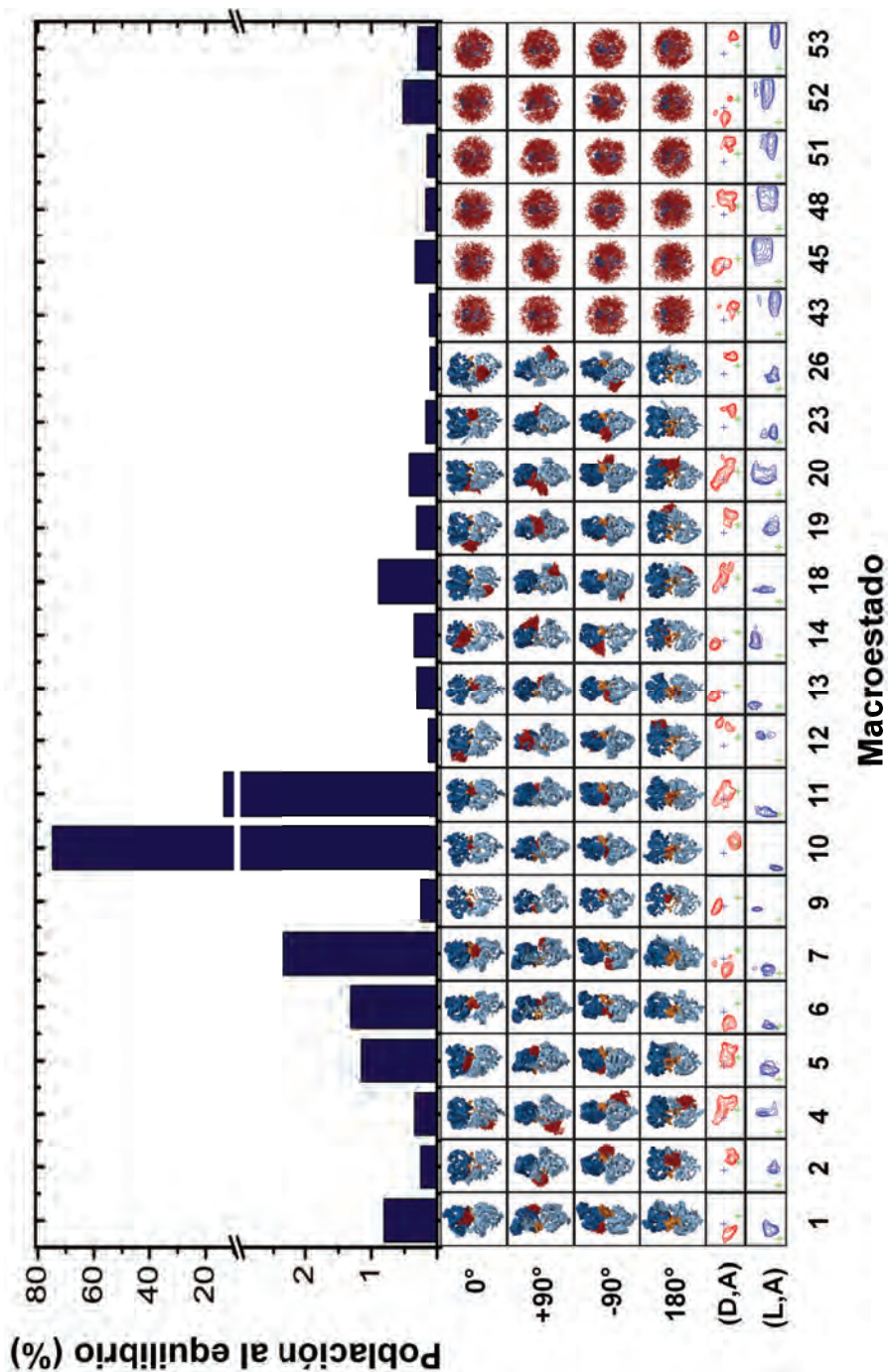


Figura 16. El gráfico muestra la población de cada uno de los 23 macro-estados más poblados en nuestro MSM ($> 0.1\%$). Para generar las figuras (cuatro perspectivas distintas) se eligieron 1,000 conformaciones aleatorias de cada macroestado y se sobrelaparon. Para cada macroestado se muestran gráficos de densidad de población para los ángulos de dislocación y apertura (D,A; color rojo) y para la distancia del ligando al sitio de unión vs la apertura (L,A; color azul). Las gráficas de densidad se encuentran en una escala de energía libre: $-\ln P$. El tamaño de los cajones para generar los histogramas fue de 5° para "A", 5° para "D" y 1.5 \AA para "L". El intervalo entre contornos es 0.5 kT . Las cruces verde y azul corresponden a las estructuras cristalográficas de LAO (estado cerrado y abierto, respectivamente).

De manera previa a la extracción de conclusiones de nuestro MSM, determinamos si el modelo era consistente con los resultados experimentales. La Figura 17 muestra que nuestro MSM es capaz de predecir *ab initio* el estado unido. Es importante considerar que en ningún momento las MD utilizadas para construir el modelo fueron alimentadas o restringidas con información del estado unido. Nuestro MSM también fue capaz de calcular la energía de unión y la constante de asociación del ligando con una precisión razonable. El modelo predice una energía de asociación de -8.46 kcal/mol y un tiempo de asociación del ligando de 0.258 ± 0.045 μ s (ver Métodos), que es un valor bastante cercano a la energía de asociación experimental de -9.95 kcal/mol. En conjunto, nuestros resultados correlacionan con los datos experimentales y, por lo tanto, el modelo puede ser usado para proponer hipótesis del mecanismo de unión de LAO.

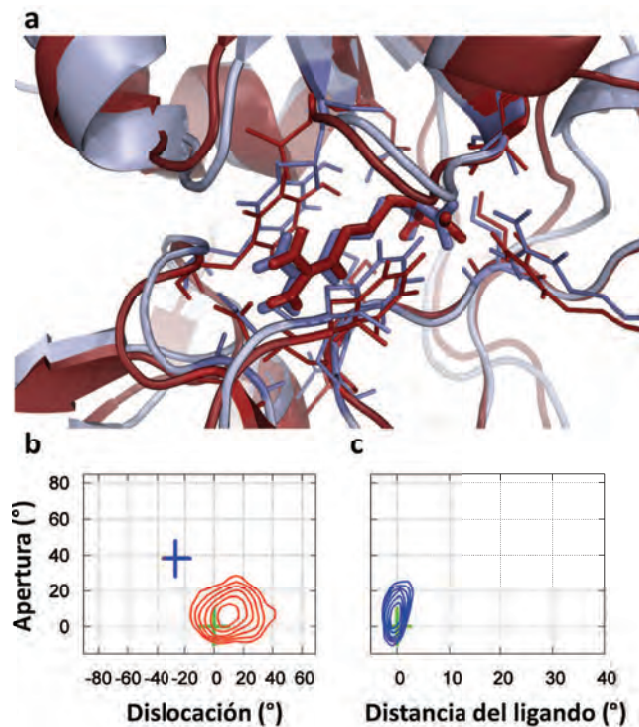


Figura 17. El estado unido identificado por nuestro MSM (macro-estado 10). En A) la proteína LAO en una representación de “listones” y el ligando en varillas. La estructura en color rojo corresponde a un cuadro de nuestras simulaciones y la estructura azul al estado unido cristalográfico (PDB ID: 1LAF). El RMSD de los $C\alpha$ entre estas dos estructuras es tan bajo como 1.8 Å y si sólo se toman en cuenta solo los residuos a una distancia de 8 Å del sitio de unión el RMSD es 1.2 Å (Residuos: 9-15, 17-19, 30, 50-53, 55-56, 67-74, 77, 83, 88, 90-92, 117-124, 141-143, 159-162, 164, 190-191 y 194-196). En B) se muestran los gráficos de densidad del macro-estado 10 para la apertura y la dislocación, mientras que en C) se muestran el grafico de densidad para la distancia del ligando al sitio de unión vs la apertura. Las graficas se encuentran en una escala de energía libre: $-\ln P$. El tamaño de las clases para generar los histogramas fue de 5° para “O” y “T” y 1.5 Å para “L”. El intervalo entre contornos es 0.5 KT. Las cruces verde y azul corresponden a las estructuras cristalográficas de LAO en el estado cerrado y abierto, respectivamente.

El mecanismo de unión de LAO

Utilizando nuestro MSM hemos estudiado las transiciones entre las conformaciones de LAO que tienden a permanecer por largos periodos de tiempo (meta-estables). Nuestros resultados muestran que los diez caminos con mayor contribución en la ruta de unión (desde cualquier estado no-unido [macroestados 45-53] hasta el estado unido [macroestado 10]) se conjuntan en un estado previo al estado-unido (ver Figura 18). Llamamos a este intermediario “el complejo de encuentro” (Macroestado 11), ya que es un estado meta-estable en el que la proteína está en una conformación semi-abierta e interactuando débilmente con el ligando (ver Figuras 19 y 20). En el complejo de encuentro, al comparar la estructura independiente de los dos lóbulos de LAO, encontramos una estructura muy similar a la existente en los cristales (del estado cerrado o abierto) (Tabla 7). Por lo tanto, el cambio conformacional entre las estructuras cristalográficas y el complejo de encuentro puede ser descrito por una simple rotación de cuerpo rígido. El análisis de las conformaciones centrales del complejo de encuentro, mostró que la cadena lateral del ligando L-arg establece una interacción de apilamiento (catión- π) con los residuos Tyr-14 y Phe-52 (Lóbulo I) y, al mismo tiempo, el extremo NH_3COO^- del ligando protruye para interactuar con el residuo Thr-121 (lóbulo II) (ver Figura 21). Estos resultados son consistentes con los presentados en la primera parte de este trabajo. Para sustentar la conclusión de que el complejo de encuentro es un paso obligatorio en la unión del ligando, hemos calculado las escalas de tiempo para transitar desde los estados no-unidos hasta el complejo de encuentro ($0.190 \pm 0.037 \mu\text{s}$), que es más corta que el tiempo necesario para transitar desde el estado no-unido hasta el estado completamente unido ($0.258 \pm 0.045 \mu\text{s}$) (ver Métodos). Por lo tanto, típicamente la proteína en el estado no-unido presentará una transición al complejo de encuentro antes de alcanzar el estado unido.

Tabla 7. RMSD de los lóbulos I y II entre el complejo de encuentro y las estructuras cristalográficas abierta y cerrada. El resultado es el promedio de tres conformaciones representativas (centrales) del complejo de encuentro. El RMSD fue calculado de manera independiente para el lóbulo I (residuos 6-88 & 195-227) y para el lóbulo II (residuos 92-185). Para calcular el RMSD el alineamiento estructural se realizó independientemente para cada lóbulo.

Lóbulo I		Lóbulo II	
Abierta (Apo) (PDB: 2LAO)	Cerrada (PDB: 1LAF)	Abierta (Apo) (PDB: 2LAO)	Cerrada (PDB: 1LAF)
1.61Å	1.71Å	1.94Å	1.92Å

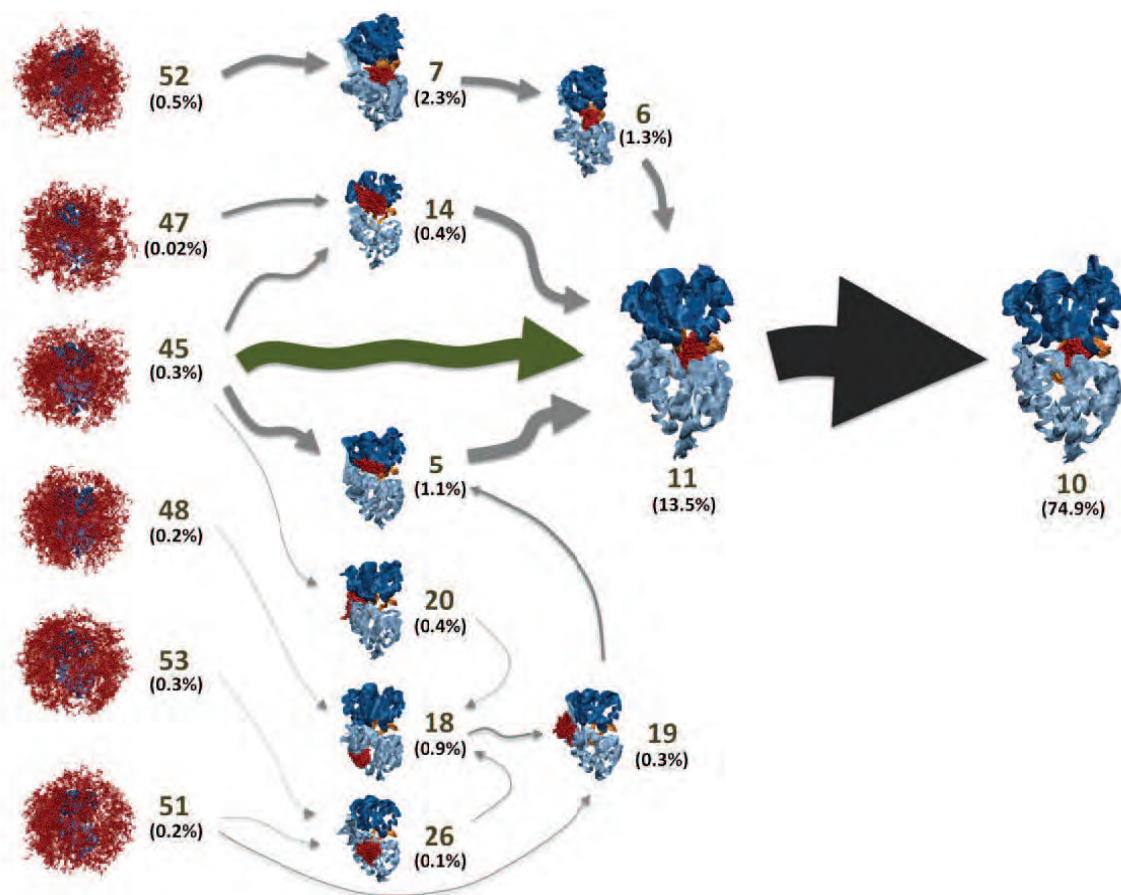


Figura 18. Superposición de los diez caminos con la contribución más grande en la transición del estado solvatado (no-unido) al estado unido. La proteína LAO se muestra como una representación de listones y el ligando en color rojo en una de varillas. 1,000 estructuras aleatorias en cada macro-estado fueron sobrelapadas para mostrar la conformación global de la proteína en cada macro-estado. El flujo fue calculado usando un algoritmo de muestreo regresivo codicioso (ver Métodos) en nuestro MSM. Estos caminos contribuyen con el 35% del flujo total desde los estados no unidos hasta el estado unido. El tamaño de las flechas es proporcional al flujo entre estados. Se muestra el número de macro-estado y la población al equilibrio calculada a partir del MSM. La flecha verde muestra el camino que sigue el mecanismo de selección conformacional, las flechas grises muestran los caminos de ajuste inducido y la flecha negra muestra el segundo paso en la vía de unión: transitar desde el complejo de encuentro hasta el estado unido mediante un mecanismo de ajuste inducido.

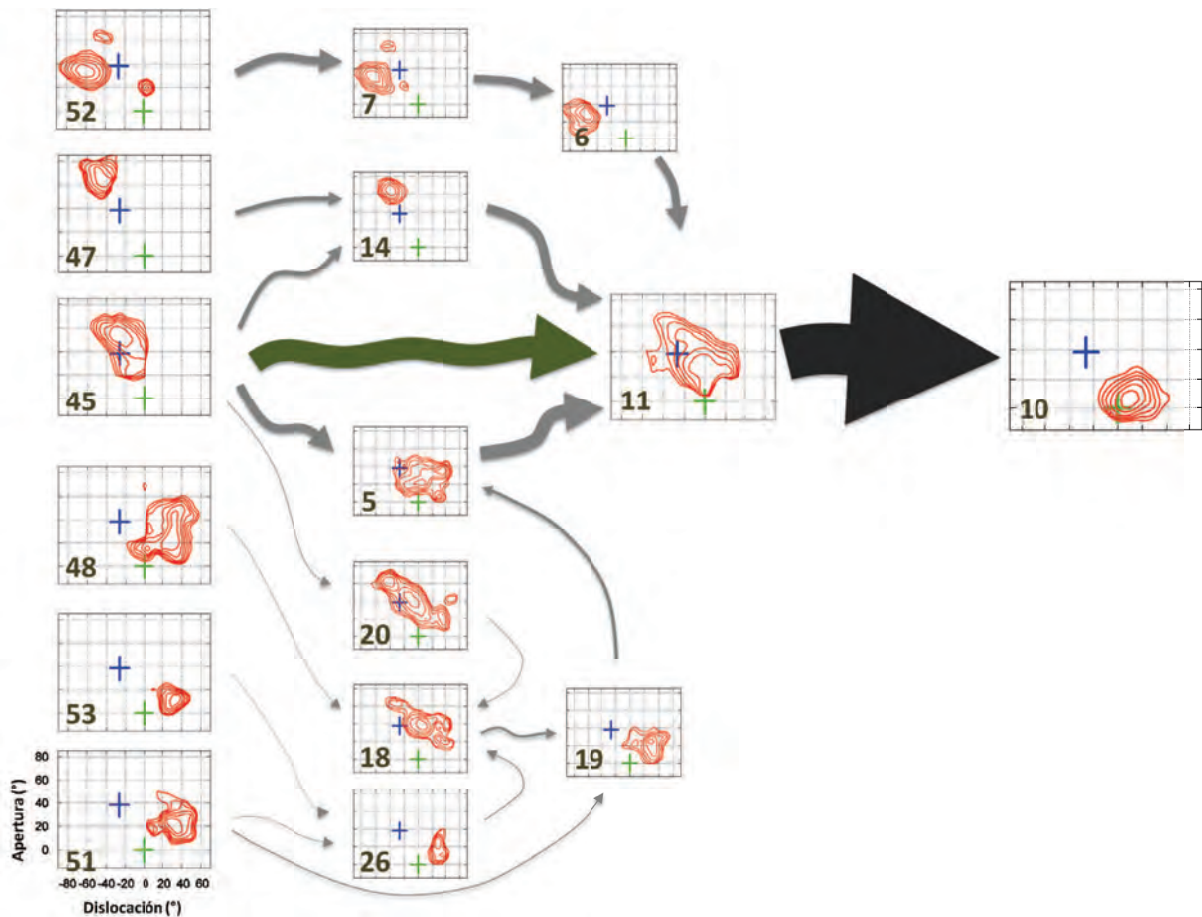


Figura 19. Superposición de los diez caminos con la contribución más grande en la transición del estado solvatado (no-unido) al estado unido. El espacio conformacional de la proteína en cada macro-estado se muestra mediante los gráficos de apertura vs. dislocación. El número de macro-estado se muestra. La flecha verde muestra el camino que sigue el mecanismo de selección conformacional, las flechas grises muestran los caminos de ajuste inducido y la flecha negra muestra el segundo paso en la vía de unión, transitar desde el complejo de encuentro hasta el estado unido (ajuste inducido). Las gráficas se encuentran en una escala de energía libre: $-\ln P$. El tamaño de las clases para generar los histogramas fue de 5° para la apertura y dislocación. El intervalo entre contornos es 0.5 KT . Las cruces verde y azul corresponden a las estructuras cristalográficas de LAO en los estados cerrado y abierto, respectivamente.

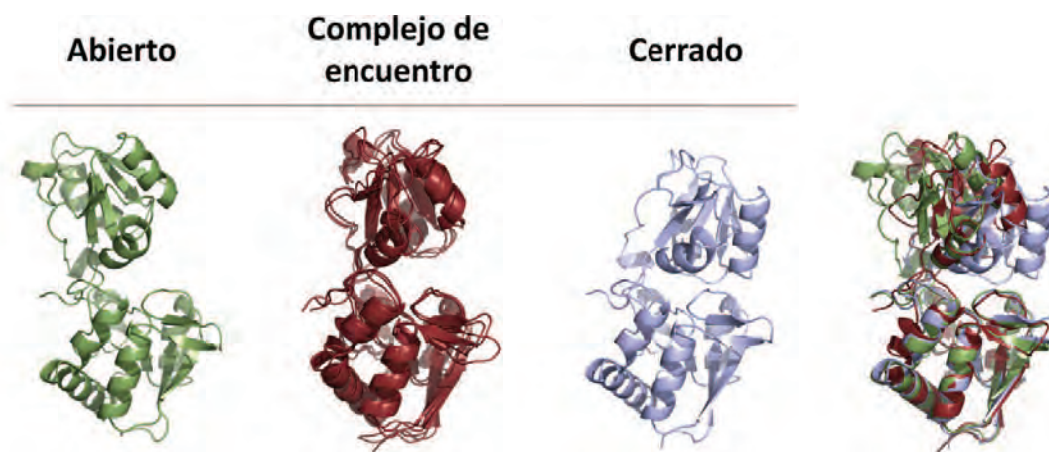


Figura 20. Comparación estructural entre el complejo de encuentro (rojo) y las estructuras cristalográficas del estado abierto (verde) y cerrado (azul) (PDB ID: 2LAO y 1LAF, respectivamente). Se muestran tres conformaciones representativas del complejo de encuentro. Las conformaciones son las estructuras centrales (de una muestra aleatoria de 10,000 conformaciones) del complejo de encuentro. A la derecha se muestra un solapamiento de la conformación abierta, cerrada y el complejo de encuentro.

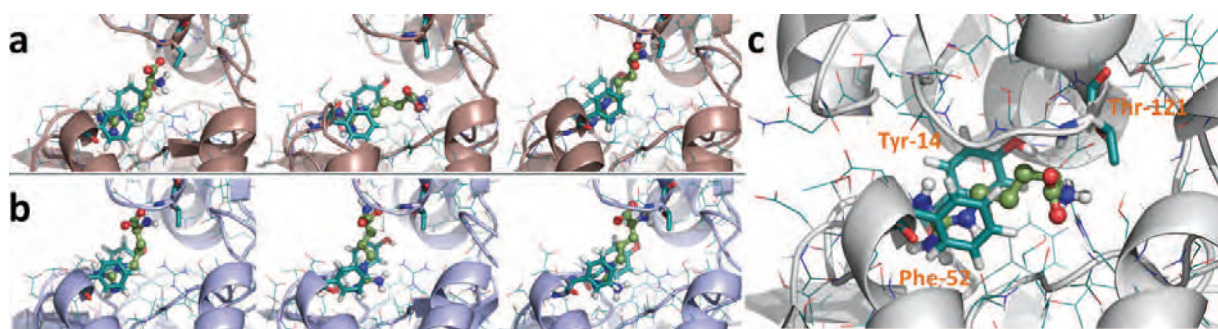


Figura 21. Tres conformaciones representativas (centrales) del complejo de encuentro. Las estructuras en a) fueron obtenidas en base al RMSD de todos los Ca de la proteína, mientras que las estructuras en B) se basan en el RMSD del sitio de unión (aminoácidos a 8Å del sitio de unión). En c) se muestra la estructura cristalográfica del estado cerrado (PDB ID: 1LAF). La proteína se muestra en una representación de esquema, los aminoácidos a 8Å del sitio de unión se muestran en líneas delgadas. Encontramos evidencia de que la cadena lateral del ligando interactúa con los aminoácidos Tyr-14 y Phe-52 a través de una interacción de catión- π . En cinco de seis conformaciones, también observamos que el grupo NH₃COO⁻ protruye para interactuar con el residuo Thr-121 en el lóbulo II, esta interacción también se observa en el estado cerrado.

Los diez caminos (o vías) con una mayor contribución a la unión del ligando pueden ser divididos en dos conjuntos: el primero, transita directamente desde el estado no-unido hasta el complejo de encuentro, en esta vía de unión el ligando interactúa con una estructura pre-existente en el ensamble termodinámico (el complejo de encuentro), y por lo tanto el mecanismo de unión

se ajusta a la descripción de un modelo de “selección conformacional” (ver flecha verde en las Figuras 18 y 19); el segundo conjunto está representado por diversas rutas que unen al ligando en conformaciones cercanas al complejo de encuentro, pero en las cuales la proteína se encuentra con ángulos de apertura y/o dislocación que son diferentes al complejo de encuentro. Estos estados transitan a través de intermediarios que eventualmente llevan al ligando al complejo de encuentro (más rápidamente de lo que difunden al solvente) (ver flecha negra en las Figuras 18 y 19). Por lo tanto este segundo conjunto de vías de unión es descrito por un modelo de “ajuste inducido. En las diez vías de unión más transitadas, la contribución de estos dos mecanismos de unión es similar, dado que la selección conformacional contribuye con un 45 % y el ajuste inducido con el 55%.

El último paso en el mecanismo de unión de LAO es la transición desde el complejo de encuentro hasta el estado cerrado. Esta transición queda mejor descrita por el modelo de ajuste inducido, ya que cuando el sistema entra al complejo de encuentro la proteína se encuentra generalmente en una conformación *cuasi*-abierta, pero cuando el ligando deja el complejo de encuentro para transitar al estado cerrado la proteína se encuentra en una conformación *cuasi*-cerrada (ver Figura 22). Por lo tanto la interacción con el ligando induce el cambio conformacional de LAO al estado cerrado. Además, nuestro MSM muestra que con excepción del macro-estado 10 (estado unido), ningún otro estado alcanza los valores de apertura y dislocación presentes en la estructura cristalográfica del estado cerrado (ver Figura 16). El modelo predice que la transición desde el complejo de encuentro hasta el estado unido ($0.090 \pm 0.015 \mu\text{s}$) es mucho más rápida que la transición del complejo de encuentro al solvente ($1.927 \pm 0.499 \mu\text{s}$), por lo tanto el complejo de encuentro típicamente se convierte en el estado unido en lugar de permitir la difusión del ligando hasta el solvente. Estas observaciones concuerdan con los resultados presentados en la “parte I” e indican que el mecanismo de ajuste inducido domina la transición del complejo de encuentro al estado cerrado.

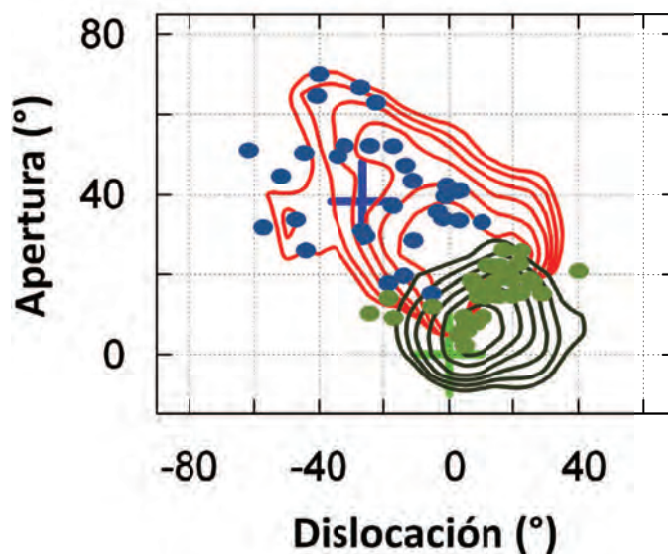


Figura 22. Sobreapamiento de los gráficos de los espacios conformacionales (apertura vs. dislocación) del complejo de encuentro (macro-estado 11, contornos color rojo) y el estado cerrado (macro-estado-10, contornos color azul). Los puntos azules corresponden a las conformaciones cuando tienen lugar las transiciones de los estados no-unidos al complejo de encuentro y los puntos verdes corresponden a las conformaciones cuando tienen lugar las transiciones del complejo de encuentro al estado cerrado. Sólo se muestran las transiciones sin "recruzamiento" (con una residencia mínima en el estado final > 6ns).

Estados incorrectamente unidos

El MSM también identificó algunos estados meta-estables minoritarios en los cuales el ligando se encuentra incorrectamente unido a la proteína (ver Figura 23). En general en estos estados el ligando interactúa con la proteína fuera y lejos del sitio de unión. Por ejemplo, en el estado 20 (población 0.4%) el ligando se encuentra interactuando con la parte posterior de la bisagra. La transición desde un estado incorrectamente unido hasta el estado unido generalmente involucra la difusión del ligando al solvente (ver Figura 23). Estos estados se encuentran usualmente fuera del camino de unión y su efecto global en el ensamble es el de hacer más lenta la unión del ligando.

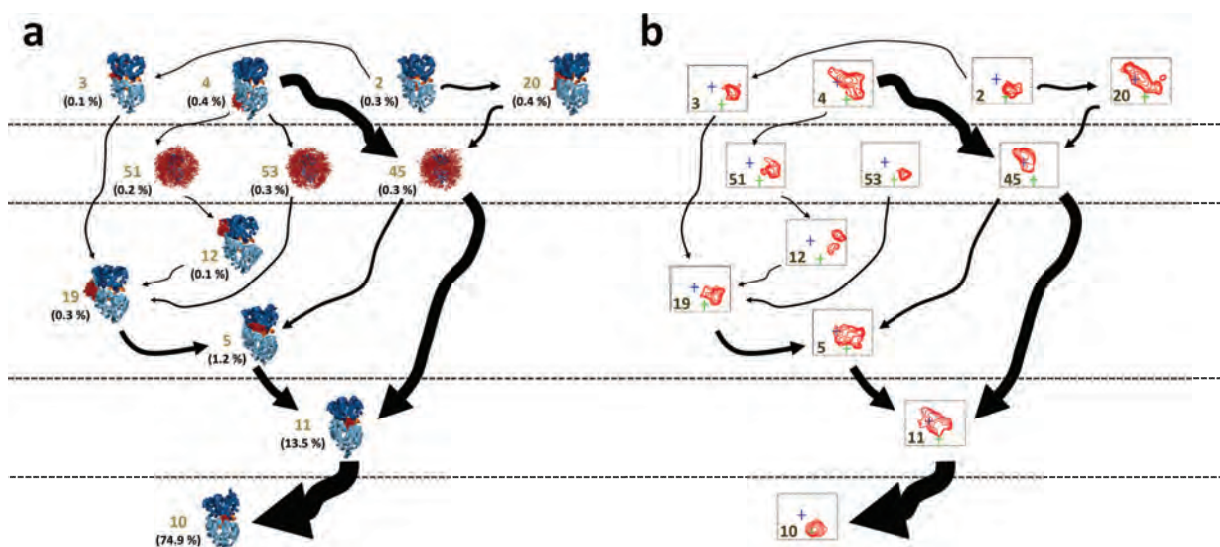


Figura 23. Las 5 rutas con mayor contribución a la transición de los estados mal-unidos al unido . El flujo fue calculado usando un algoritmo de muestreo regresivo ambicioso (ver Métodos) en el MSM generado a partir de los 53 macro-estados. El tamaño de las flechas es proporcional al flujo. a) Muestran la estructuras y poblaciones al equilibrio de los macro-estados y b) muestra los espacios conformacionales (apertura vs. dislocación) usando gráficos de contornos. Las gráficas se encuentran en una escala de energía libre: $-lnP$. El tamaño de las clases para generar los histogramas fue de 5° para “O” y “T”. El intervalo entre contornos es 0.5 KT . Las cruces verde y azul corresponden a las estructuras cristalográficas de LAO (estado cerrado y abierto, respectivamente).

Discusión

La proteína LAO es un miembro de las PBP que participan en el mecanismo de transporte conocido como “ABC” (20, 33, 58). Este mecanismo es responsable de transportar ligandos de bajo peso molecular desde el periplasma hasta el interior de las bacterias gram-negativas. Estudios estructurales y bioquímicos de las PBP han mostrado que el cambio conformacional del estado abierto al cerrado es fundamental para su función. Nuestro estudio amplía este conocimiento revelando un panorama más complejo, en el que el cambio conformacional de LAO (y posiblemente de otras PBP) está compuesto por, al menos, dos movimientos: apertura y dislocación. La combinación de estos movimientos (esencialmente de cuerpo rígido) da lugar a un paisaje conformacional que contiene múltiples estados meta-estables.

La mayor parte de los trabajos teóricos para estudiar los cambios conformacionales en PBP han sido analizados por dos métodos: RMSD y PCA. Sin embargo, ambas metodologías presentan ciertos inconvenientes. El RMSD refleja los cambios globales y, por lo tanto, los movimientos relevantes pueden ser oscurecidos por las fluctuaciones globales o movimientos combinados;

por otra parte en el PCA, el orden de los eigenvectores depende de la magnitud (y velocidad) de los movimientos muestreados en las trayectorias, por lo tanto entre distintas MDs no se puede garantizar que en el análisis se obtenga el mismo resultado. En nuestro trabajo hemos diseñado una metodología mejorada para analizar el estado conformacional de LAO. Nuestro método, basado en simples medidas geométricas (ángulos diedros), captura de manera precisa el estado conformacional de la proteína con sólo medir dos coordenadas de reacción que son intuitivas: ángulo de apertura y dislocación. Debido a la alta similitud estructural entre los miembros de las PBPs, este método se podría extender fácilmente para analizar otros miembros de la familia o proteínas con un plegamiento similar, y probablemente se podría adaptar para estudios equivalentes en otras familias de proteínas.

Nuestros resultados, en concordancia con los experimentales, muestran que la conformación abierta es preferida en la ausencia del ligando. Sin embargo, nuestro trabajo muestra que el llamado “estado abierto” es un ensamble de conformaciones que abarca un rango amplio de ángulos de apertura y dislocación, en el cual la estructura “cristalográfica abierta” es sólo un punto en el extremo del ensamble. Esta diferencia entre las simulaciones y los datos experimentales, podría deberse a que la celda unitaria de nuestras simulaciones (419.6 nm^3) es considerablemente más grande que la de los cristales de apo-LAO ($2\text{LAO} = 286 \text{ nm}^3$) y holo-LAO ($1\text{LAF} = 261 \text{ nm}^3$). Por lo tanto en el cristal podrían existir restricciones (contactos cristalográficos) que favorecen las conformaciones encontradas experimentalmente.

Cuando el movimiento de dislocación fue tomado en consideración, el estado abierto y libre de ligando exhibió dos poblaciones distintas y cada una de estas presentó un patrón específico de puentes de hidrógeno. Un análisis de las trayectorias que alcanzaron el estado abierto-dislocado nos permitió descartar la posibilidad de que sea un intermediario obligatorio en el mecanismo de la transición del estado cerrado al abierto. En el mecanismo inverso, nuestros resultados concuerdan, ya que nuestras simulaciones en presencia del ligando y con transiciones espontáneas al estado cerrado tampoco transitan por un intermediario abierto-dislocado. De manera que, los resultados sugieren que si bien es posible estudiar los cambios conformacionales de apo-LAO en simulaciones cortas (20 ns), también es importante tener en cuenta que podría generarse un efecto de histéresis debido a un punto inicial artificial (e.g. una conformación de muy alta energía).

Realizamos MDs del estado cerrado en presencia de los ligandos nativos. De esta manera fuimos capaces de reproducir el resultado experimental, es decir, una proteína que permaneció en el estado cerrado y unida al ligando. Encontramos que el estado cerrado, en lugar de ser una conformación rígida, es una conformación que puede muestrear segmentos amplios de combinaciones de ángulos de apertura y dislocación. Dicho de otra manera, el estado cerrado es flexible y puede inclusive muestrear conformaciones relativamente abiertas y/o con dislocadas ($\sim +20^\circ$). A diferencia de la evidencia cristalográfica (23), hallamos que la conformación cerrada de cada uno de los complejos proteína-ligando es distinta. Creemos que una interpretación plausible de estos resultados es que en conjunto los diferentes complejos cerrados representan lo que macroscópicamente llamamos el estado cerrado, y el efecto de la interacción con cada uno de los diferentes ligandos desplaza el equilibrio hacia un región particular del espacio de fase. Por lo tanto, una hipótesis razonable es que la flexibilidad del estado cerrado de LAO podría estar asociada a la capacidad para unir múltiples ligandos. Esta variabilidad podría explicar, al menos parcialmente, el por qué la fluorescencia intrínseca de los triptófanos y tirosinas muestra diferentes espectros de emisión para cada complejo proteína-ligando (33). Esta hipótesis sobre la flexibilidad del estado cerrado también podría explicar algunos de los resultados experimentales que Wolf *et. al.* (29) encontraron en la proteína HisJ (ver Métodos: racionalización de la comparación entre LAO e HisJ), en el que se ha mostrado que HisJ adopta diferentes conformaciones al interactuar con cada uno de los distintos ligandos en solución.

Utilizando como ligando a los tautómeros de la L-histidina, abordamos una pregunta que es difícil de contestar experimentalmente: ¿Cuál es el efecto del estado de protonación del ligando en la afinidad? Si bien se ha sugerido que en LAO el estado de protonación del ligando histidina podría ser importante para la afinidad, ningún experimento lo ha demostrado. Nuestros resultados concuerdan con esta hipótesis y proveen de una explicación. De manera opuesta al tautómero con carga de L-histidina, los tautómeros sin carga tienden a desestabilizar el estado cerrado. La explicación a este fenómeno se encuentra en la interacción del ligando con los residuos Tyr-14 y Phe-52. Mientras que la cadena lateral de cualquiera de los ligandos con carga positiva puede generar una interacción de catión- π con estos residuos, no es posible formar esta interacción cuando la cadena lateral del ligando tiene carga neutra. Por esto, nuestra hipótesis es que la interacción de catión- π juega un papel decisivo en la unión del ligando. Encontramos interesante que la naturaleza eligió este tipo de interacción en lugar, por ejemplo, de una mediada por

puentes de hidrógeno. Una explicación plausible podría encontrarse en las siguientes dos razones: i) la interacción del ligando mediante catión- π con los residuos 14 y 52 es extremadamente selectiva a la carga pero no a la forma de la cadena lateral, por lo que puede estabilizar a cualquiera de los ligandos catiónicos que se encuentren en la posición correcta, ii) dada la naturaleza hidrofóbica de los residuos 14 y 52, cuando éstos no se encuentran interactuando con el ligando, o cuando un ligando incorrecto o mal colocado se encuentra en el sitio de unión, la movilidad de las cadenas laterales de los residuos 14 y 52 puede aumentar, generando así una contribución entrópica desfavorable al estado cerrado.

En un intento por capturar transiciones espontáneas del estado abierto al estado cerrado realizamos MDs cortas (20 ns) con el ligando en el sitio de unión. La idea detrás del experimento fue la de tratar de rescatar, en el estado abierto inicial, la mayor parte de las interacciones que existen en el estado cerrado. Como más tarde lo demostró nuestro mapeo exhaustivo del mecanismo completo de unión, nuestra lógica fue acertada. Al colocar un ligando en el sitio de unión de la proteína abierta (en una conformación en la que la cadena lateral de ligando se encuentre correctamente posicionada) aumenta considerablemente la posibilidad de éxito en la captura de transiciones al estado cerrado. Utilizando las trayectorias con transiciones del estado abierto-unido al cerrado-unido fuimos capaces de observar la formación secuencial de puentes de hidrógeno. Este patrón de formación progresivo se observó, tanto en simulaciones para el ligando L-arg como para el ligando L-lys, sugiriendo así que un mecanismo de ajuste inducido opera en la transición del estado abierto al estado cerrado. Estos estudios sentaron las bases para realizar un segundo experimento abordando un objetivo más complejo: entender el mecanismo completo de unión operante en LAO.

Las estructuras cristalográficas muestran que el sitio de unión de LAO se encuentra completamente inaccesible en el estado cerrado, lo que hace imposible que el ligando entre al sitio de unión una vez que la proteína ha adoptado la conformación cerrada (ver Figura 1). Se ha sugerido frecuentemente que el mecanismo de unión operante en LAO es el ajuste inducido.

Sin embargo, nuestras observaciones indican que al igual que el plegamiento de proteínas (59), la unión de ligandos es un proceso multi-estado con vías paralelas. Las diez vías con mayor contribución a la unión del ligando alcanzan un estado común antes del estado unido: “el complejo de encuentro”. El sistema llega al complejo de encuentro utilizando tanto un mecanismo de ajuste inducido como uno de selección conformacional (ver Figura 24). El

complejo de encuentro es un estado meta-estable, en el que la proteína se encuentra parcialmente cerrada e interactuando débilmente con el ligando. Una vez en el complejo de encuentro, el ligando puede inducir rápidamente el cambio conformacional al estado cerrado (ver Figura 24).

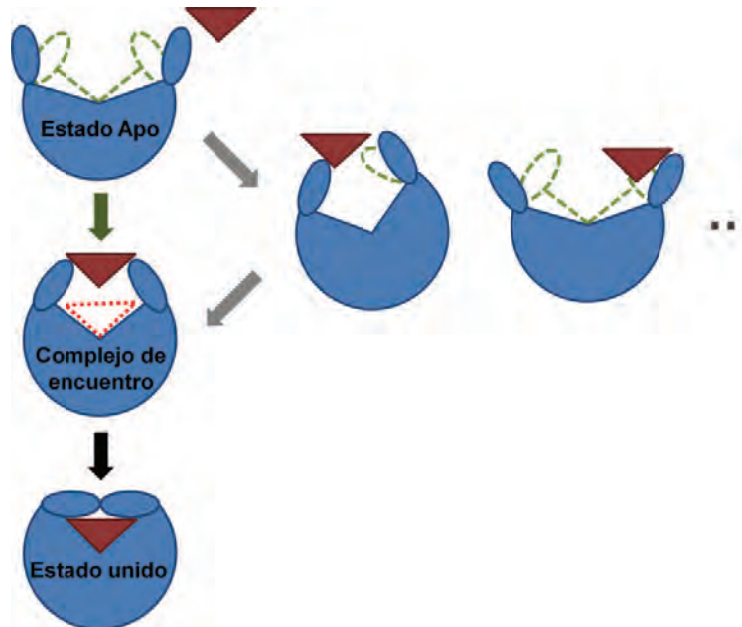


Figura 24. Diagrama que describe esquemáticamente el mecanismo de unión de dos etapas para la proteína LAO. La proteína es representada por la estructura azul y el ligando por el triángulo rojo. La primera etapa es la transición del estado no-unido (apo) al complejo de encuentro. En este paso existen múltiples vías y, tanto la selección conformacional, como el ajuste inducido juegan papeles importantes. En el segundo paso el sistema transita del complejo de encuentro al estado unido y un mecanismo de ajuste inducido es el responsable de este fenómeno.

Es razonable pensar que otros sistemas operan mediante mecanismos similares, tal como las proteínas de la familia de las PBP. Efectivamente, nuestro modelo es consistente con resultados experimentales que muestran que la PBP que une maltosa (MBP) se encuentra en equilibrio entre una forma abierta y una parcialmente cerrada en ausencia de ligando (32). En la conformación parcialmente cerrada el sitio de unión se encuentra accesible al ligando y por lo tanto disponible para unirlo mediante un mecanismo de selección conformacional. Se ha postulado que la unión a este estado semi-cerrado podría, tras la unión del ligando, facilitar la transición al estado unido mediante un mecanismo de ajuste inducido. Sin embargo, este modelo para MBP fue especulado a partir de experimentos de resonancia magnética nuclear (RMN) en ausencia del ligando. No obstante, nuestras simulaciones en presencia del ligando y con resolución atómica han capturado un escenario similar en LAO. Mediante nuestro modelo hemos podido definir que la transición

del estado abierto al complejo de encuentro ocurre utilizando una combinación de selección conformacional y ajuste inducido. En el caso de la transición desde el complejo de encuentro hasta el estado cerrado hemos confirmado que un mecanismo de ajuste inducido es el responsable del último paso de asociación. Es interesante mencionar que mediciones utilizando experimentos de cinética rápida en solución (stopped flow), han mostrado que para varias PBPs la tasa de asociación es proporcional y cambia en forma monotónica con la concentración del ligando, indicando un mecanismo de unión simple de dos estados (no-unido y unido). Este escenario se ha observado en HisJ y también para MBP. Sin embargo, como ha sido sugerido por Tang *et. al.* (32) existe una explicación simple para esta discrepancia con el modelo de ajuste inducido: “si la escala temporal de la transición es extremadamente veloz, los experimentos de cinética rápida carecerían de la resolución temporal para capturar al intermediario”. En el caso de LAO, nuestro MSM sugiere que la transición al complejo de encuentro tiene un tiempo medio de transición (MFPT) de apenas unos cuantos nanosegundos.

En un enfoque más amplio, el mecanismo de unión que hemos propuesto para LAO (ver Figura 24) también podría existir en otras proteínas con sitios de unión (o sitios activos) cerrados, pero la contribución relativa de la selección conformacional y el ajuste inducido podrían variar dependiendo de diversos factores, como la intensidad de las interacciones proteína-ligando y las fuerzas de atracción y repulsión de corto y largo alcance.

Hasta donde sabemos, este trabajo representa el primer ejemplo en el que simulaciones de MD sin restricciones reproducen el mecanismo completo de unión de un ligando a una proteína. La capacidad de extraer detalles de estas simulaciones utilizando MSM es un paso importante hacia un mejor entendimiento de la interacción entre las PBPs y sus ligandos. Aplicaciones futuras de esta metodología a otros sistemas podrían inclusive llevar a la identificación de los principios generales que guían las interacciones proteína-ligando y alostéricas. Este conocimiento podría ayudar enormemente en el diseño de compuestos farmacéuticos mediante computadora. Trabajos futuros con campos de fuerza mejorados y muestreos más amplios podrían aumentar nuestro entendimiento de estas interacciones, pero afirmamos que este trabajo tiene el potencial de sentar las bases que se emplearan para realizar tales investigaciones.

Conclusiones

Aunque las estructuras cristalográficas de LAO sugieren que la conformación unida es esencialmente idéntica para todos los ligandos, el estudio dinámico muestra que la proteína adopta conjuntos conformacionales específicos para cada ligando. Este ajuste es el resultado de la combinación de dos movimientos: la apertura y la dislocación. Hemos mostrado que es razonable proponer que el reconocimiento del ligando está dado por una combinación de puentes de hidrógeno y una interacción de catión- π con los residuos Tyr-14 y Phe-52.

Mediante una combinación de MDs no restringidas, un muestreo extensivo del espacio conformacional de LAO y MSM, hemos reproducido el paisaje conformacional del mecanismo de unión del ligando L-arg. Así, demostramos el poder que los MSM ofrecen para entender las interacciones proteína-ligando. Con nuestro modelo, hemos definido que el mecanismo de unión de LAO está formado por varios estados y vías paralelas, que se pueden resumir esencialmente en dos pasos. En el primer paso, el ligando se une a una conformación parcialmente cerrada para formar el complejo de encuentro, y tanto la selección conformacional como el ajuste inducido juegan un papel en este proceso. En el segundo paso, el sistema transita desde el complejo de encuentro hasta el estado unido mediante un mecanismo de ajuste inducido. Creemos que este mecanismo de dos pasos podría operar en otros sistemas, tales como PBPs, enzimas con sitios activos cerrados y sistemas en los cuales la proteína libre de ligando/sustrato raramente visita la conformación unida/activa.

La metodología desarrollada tiene el potencial de sentar las bases para generar un mejor entendimiento de los mecanismos de unión. Aplicaciones futuras de los MSM, utilizando mejores campos de fuerza, muestreos más amplios y estudios en otros sistemas proteína-ligando, revelarán que tan generales son los mecanismos que hemos planteado y de esta manera podrán ayudar a generar un diseño de fármacos más eficiente.

Bibliografia

1. Koshland, D. E. (1958) Application of a Theory of Enzyme Specificity to Protein Synthesis, *Proc Natl Acad Sci U S A* 44, 98-104.
2. Kumar, S., Ma, B., Tsai, C. J., Sinha, N., and Nussinov, R. (2000) Folding and binding cascades: dynamic landscapes and population shifts, *Protein Sci* 9, 10-19.
3. Ma, B., Kumar, S., Tsai, C. J., and Nussinov, R. (1999) Folding funnels and binding mechanisms, *Protein Eng* 12, 713-720.
4. Ma, B., Shatsky, M., Wolfson, H. J., and Nussinov, R. (2002) Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations, *Protein Sci* 11, 184-197.
5. Tsai, C. J., Kumar, S., Ma, B., and Nussinov, R. (1999) Folding funnels, binding funnels, and protein function, *Protein Sci* 8, 1181-1190.
6. Tsai, C. J., Ma, B., and Nussinov, R. (1999) Folding and binding cascades: shifts in energy landscapes, *Proc Natl Acad Sci U S A* 96, 9970-9972.
7. James, L. C., Roversi, P., and Tawfik, D. S. (2003) Antibody multispecificity mediated by conformational diversity, *Science* 299, 1362-1367.
8. Arora, K., and Brooks, C. L., 3rd. (2007) Large-scale allosteric conformational transitions of adenylate kinase appear to involve a population-shift mechanism, *Proc Natl Acad Sci U S A* 104, 18496-18501.
9. Bahar, I., Chennubhotla, C., and Tobi, D. (2007) Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation, *Curr Opin Struct Biol* 17, 633-640.
10. Csermely, P., Palotai, R., and Nussinov, R. (2010) Induced fit, conformational selection and independent dynamic segments: an extended view of binding events, *Trends Biochem Sci* 35, 539-546.
11. Zhou, H. X. From induced fit to conformational selection: a continuum of binding mechanism controlled by the timescale of conformational transitions, *Biophys J* 98, L15-17.
12. Okazaki, K., and Takada, S. (2008) Dynamic energy landscape view of coupled binding and protein conformational change: induced-fit versus population-shift mechanisms, *Proc Natl Acad Sci U S A* 105, 11182-11187.
13. Hammes, G. G., Chang, Y. C., and Oas, T. G. (2009) Conformational selection or induced fit: a flux description of reaction mechanism, *Proceedings of the National Academy of Sciences of the United States of America* 106, 13737-13741.
14. Sullivan, S. M., and Holyoak, T. (2008) Enzymes with lid-gated active sites must operate by an induced fit mechanism instead of conformational selection, *Proc Natl Acad Sci U S A* 105, 13829-13834.
15. Wlodarski, T., and Zagrovic, B. (2009) Conformational selection and induced fit mechanism underlie specificity in noncovalent interactions with ubiquitin, *Proc Natl Acad Sci U S A* 106, 19346-19351.
16. Swift, R. V., and McCammon, J. A. (2009) Substrate induced population shifts and stochastic gating in the PBCV-1 mRNA capping enzyme, *J Am Chem Soc* 131, 5126-5133.
17. Weikl, T. R., and von Deuster, C. (2009) Selected-fit versus induced-fit protein binding: kinetic differences and mutational analysis, *Proteins* 75, 104-110.

18. Formanek, M. S., Ma, L., and Cui, Q. (2006) Reconciling the "old" and "new" views of protein allostery: a molecular simulation study of chemotaxis Y protein (CheY), *Proteins* 63, 846-867.
19. Bakan, A., and Bahar, I. (2009) The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding, *Proc Natl Acad Sci U S A* 106, 14349-14354.
20. Dwyer, M. A., and Hellinga, H. W. (2004) Periplasmic binding proteins: a versatile superfamily for protein engineering, *Curr. Opin. Struct. Biol.* 14, 495-504.
21. Pang, A., Arinaminpathy, Y., Sansom, M. S., and Biggin, P. C. (2005) Comparative molecular dynamics-similar folds and similar motions?, *Proteins* 61, 809-822.
22. Keskin, O., Jernigan, R. L., and Bahar, I. (2000) Proteins with similar architecture exhibit similar large-scale dynamic behavior, *Biophys J* 78, 2093-2106.
23. Oh, B. H., Ames, G. F., and Kim, S. H. (1994) Structural basis for multiple ligand specificity of the periplasmic lysine-, arginine-, ornithine-binding protein, *J. Biol. Chem.* 269, 26323-26330.
24. Oh, B. H., Pandit, J., Kang, C. H., Nikaido, K., Gokcen, S., Ames, G. F., and Kim, S. H. (1993) Three-dimensional structures of the periplasmic lysine/arginine/ornithine-binding protein with and without a ligand, *J. Biol. Chem.* 268, 11348-11355.
25. Ames, G. F. (1986) Bacterial periplasmic transport systems: structure, mechanism, and evolution, *Annu Rev Biochem* 55, 397-425.
26. Stockner, T., Vogel, H., and Tieleman, D. (2005) A salt-bridge motif involved in ligand binding and large-scale domain motions of the maltose-binding protein, *Biophys. J.* 89, 3362-3371.
27. Bermejo, G. A., Strub, M. P., Ho, C., and Tjandra, N. (2010) Ligand-free open-closed transitions of periplasmic binding proteins: the case of glutamine-binding protein, *Biochemistry* 49, 1893-1902.
28. Loeffler, H. H., and Kitao, A. (2009) Collective dynamics of periplasmic glutamine binding protein upon domain closure., *Biophys J* 97, 2541-2549.
29. Wolf, A., Lee, K. C., Kirsch, J. F., and Ames, G. F. (1996) Ligand-dependent conformational plasticity of the periplasmic histidine-binding protein HisJ. Involvement in transport specificity., *J Biol Chem* 271, 21243-21250.
30. Flocco, M. M., and Mowbray, S. L. (1994) The 1.9 Å x-ray structure of a closed unliganded form of the periplasmic glucose/galactose receptor from *Salmonella typhimurium*, *J Biol Chem* 269, 8931-8936.
31. Oswald, C., Smits, S. H., Hoing, M., Sohn-Bosser, L., Dupont, L., Le Rudulier, D., Schmitt, L., and Bremer, E. (2008) Crystal structures of the choline/acetylcholine substrate-binding protein ChoX from *Sinorhizobium meliloti* in the liganded and unliganded-closed states, *J Biol Chem* 283, 32848-32859.
32. Tang, C., Schwieters, C. D., and Clore, G. M. (2007) Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR, *Nature* 449, 1078-1082.
33. Nikaido, K., and Ames, G. F. (1992) Purification and characterization of the periplasmic lysine-, arginine-, ornithine-binding protein (LAO) from *Salmonella typhimurium*, *J. Biol. Chem.* 267, 20706-20712.
34. Oh, B. H., Ames, G. F., and Kim, S. H. (1994) Structural basis for multiple ligand specificity of the periplasmic lysine-, arginine-, ornithine-binding protein, *The Journal of biological chemistry* 269, 26323-26330.

35. Buchete, N. V., and Hummer, G. (2008) Coarse master equations for peptide folding dynamics, *J Phys Chem B* 112, 6057-6069.
36. Noe, F., Schutte, C., Vanden-Eijnden, E., Reich, L., and Weikl, T. R. (2009) Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations, *Proc Natl Acad Sci U S A* 106, 19011-19016.
37. Huang, X., Yao, Y., Bowman, G. R., Sun, J., Guibas, L. J., Carlsson, G., and Pande, V. S. (2010) Constructing multi-resolution markov state models (msms) to elucidate RNA hairpin folding mechanisms, *Pac Symp Biocomput*, 228-239.
38. Voelz, V. A., Bowman, G. R., Beauchamp, K., and Pande, V. S. Molecular simulation of ab initio protein folding for a millisecond folder NTL9(1-39), *J Am Chem Soc* 132, 1526-1528.
39. Morcos, F., Chatterjee, S., McClendon, C. L., Brenner, P. R., Lopez-Rendon, R., Zintsmaster, J., Ercsey-Ravasz, M., Sweet, C. R., Jacobson, M. P., Peng, J. W., and Izaguirre, J. A. (2010) Modeling conformational ensembles of slow functional motions in Pin1-WW, *PLoS Comput Biol* 6, e1001015.
40. Noe, F., and Fischer, S. (2008) Transition networks for modeling the kinetics of conformational change in macromolecules, *Curr Opin Struct Biol* 18, 154-162.
41. Chodera, J. D., Singhal, N., Pande, V. S., Dill, K. A., and Swope, W. C. (2007) Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics, *J Chem Phys* 126, 155101.
42. Bowman, G. R., Huang, X., and Pande, V. S. (2009) Using generalized ensemble simulations and Markov state models to identify conformational states, *Methods* 49, 197-201.
43. Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) The Protein Data Bank. A computer-based archival file for macromolecular structures, *Eur J Biochem* 80, 319-324.
44. Bourne, P. E., Address, K. J., Bluhm, W. F., Chen, L., Deshpande, N., Feng, Z., Fleri, W., Green, R., Merino-Ott, J. C., Townsend-Merino, W., Weissig, H., Westbrook, J., and Berman, H. M. (2004) The distribution and query systems of the RCSB Protein Data Bank, *Nucleic Acids Res* 32, D223-225.
45. Deshpande, N., Address, K. J., Bluhm, W. F., Merino-Ott, J. C., Townsend-Merino, W., Zhang, Q., Knezevich, C., Xie, L., Chen, L., Feng, Z., Green, R. K., Flippen-Anderson, J. L., Westbrook, J., Berman, H. M., and Bourne, P. E. (2005) The RCSB Protein Data Bank: a redesigned query system and relational database based on the mmCIF schema, *Nucleic Acids Res* 33, D233-237.
46. Guex, N., and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling, *Electrophoresis* 18, 2714-2723.
47. DeLano, W. L. (2002) The PyMOL Molecular Graphics System, en WWW <http://www.pymol.org>
48. Lindahl, E., Hess, B., and van der Spoel, D. (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis, *Journal of Molecular Modeling* 7, 306-317.
49. van der Spoel, D., Lindahl, E., Hess, B., van Buuren, A. R., Apol, E., Meulenhoff, P. J., Tieleman, D. P., Sijbers, A. L. T. M., Feenstra, K. A., van Drunen, R., and Berendsen, H. J. C. (2005) GROMACS User Manual, version 3.3.

50. van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. (2005) GROMACS: fast, flexible, and free, *Journal of computational chemistry* 26, 1701-1718.
51. Kang, C. H., Shin, W. C., Yamagata, Y., Gokcen, S., Ames, G. F., and Kim, S. H. (1991) Crystal structure of the lysine-, arginine-, ornithine-binding protein (LAO) from *Salmonella typhimurium* at 2.7-Å resolution, *J. Biol. Chem.* 266, 23893-23899.
52. Oh, B. H., Kang, C. H., De Bondt, H., Kim, S. H., Nikaido, K., Joshi, A. K., and Ames, G. F. (1994) The bacterial periplasmic histidine-binding protein. structure/function analysis of the ligand-binding site and comparison with related proteins, *J. Biol. Chem.* 269, 4135-4143.
53. Swope, W. C., Pitera, J. W., and Suits, F. (2004) Describing Protein Folding Kinetics by Molecular Dynamics Simulations. 1. Theory., *J. Phys. Chem. B* 108, 6571-6581.
54. Berezhkovskii, A., Hummer, G., and Szabo, A. (2009) Reactive flux and folding pathways in network models of coarse-grained protein dynamics, *J Chem Phys* 130, 205102.
55. Singhal, N., Snow, C. D., and Pande, V. S. (2004) Using path sampling to build better Markovian state models: predicting the folding rate and mechanism of a tryptophan zipper beta hairpin, *J Chem Phys* 121, 415-425.
56. Hunenberger, P. H., Granwehr, J. K., Aebischer, J. N., Ghoneim, N., Haselbach, E., and vanGunsteren, W. F. (1997) Experimental and theoretical approach to hydrogen-bonded diastereomeric interactions in a model complex, *Journal of the American Chemical Society* 119, 7533-7544.
57. Miller, D. M., 3rd, Olson, J. S., Pflugrath, J. W., and Quiocho, F. A. (1983) Rates of ligand binding to periplasmic proteins involved in bacterial transport and chemotaxis, *J Biol Chem* 258, 13665-13672.
58. Quiocho, F. A. (1990) Atomic structures of periplasmic binding proteins and the high-affinity active transport systems in bacteria, *Philos Trans R Soc Lond B Biol Sci* 326, 341-351; discussion 351-342.
59. Bowman, G. R., and Pande, V. S. (2010) Protein folded states are kinetic hubs, *Proc Natl Acad Sci U S A* 107, 10890-10895.

Conformational dynamics of L-lysine, L-arginine, L-ornithine binding protein reveals ligand-dependent plasticity

Daniel-Adriano Silva,¹ Lenin Domínguez-Ramírez,¹ Arturo Rojo-Domínguez,² and Alejandro Sosa-Peinado^{1*}

¹Departamento de Bioquímica, Facultad de Medicina, Universidad Nacional Autónoma de México, D.F. C.P. 04510, México

²División de Ciencias Naturales e Ingeniería, Universidad Autónoma Metropolitana-Unidad Cuajimalpa, C.P. 11950, México, D.F.

ABSTRACT

The molecular basis of multiple ligand binding affinity for amino acids in periplasmic binding proteins (PBP) and in the homologous domain for class C G-protein coupled receptors is an unsolved question. Here, using unrestrained molecular dynamic simulations, we studied the ligand binding mechanism present in the L-lysine, L-arginine, L-ornithine binding protein. We developed an analysis based on dihedral angles for the description of the conformational changes upon ligand binding. This analysis has an excellent correlation with each of the two main movements described by principal component analysis (PCA) and it's more convenient than RMSD measurements to describe the differences in the conformational ensembles observed. Furthermore, an analysis of hydrogen bonds showed specific interactions for each ligand studied as well as the ligand interaction with the aromatic residues Tyr-14 and Phe-52. Using uncharged histidine tautomers, these interactions are not observed. On the basis of these results, we propose a model in which hydrogen bond interactions place the ligand in the correct orientation to induce a cation- π interaction with Tyr-14 and Phe-52 thereby stabilizing the closed state. Our results also show that this protein adopts slightly different closed conformations to make available specific hydrogen bond interactions for each ligand thus, allowing a single mechanism to attain multiple ligand specificity. These results shed light on the experimental evidence for ligand-dependent conformational plasticity not explained by the previous crystallographic data.

Proteins 2011; 79:2097–2108.
© 2011 Wiley-Liss, Inc.

Key words: cation- π interaction; dihedral angle; molecular dynamics; periplasmic binding protein; principal component analysis.

INTRODUCTION

The periplasmic-binding proteins (PBPs) belong to a structurally conserved family of proteins that bind diverse ligands and are members of the bacterial ABC transporters. Experimental evidence had shown that PBPs display a principal conformational change associated to ligand binding: the transition from open to closed state.^{1–3} This change precedes the interaction of the PBP with the membrane complex.^{4–6}

Previous X-ray crystallography diffraction studies had provide more than 300 crystallographic structures of PBPs in the presence and absence of ligands.⁴ However, the molecular mechanism behind ligand binding,^{7,8} and the role of the conformational change in the ligand recognition remain elusive. Since X-ray crystallographic data lacks the protein dynamic behaviour in solution, including the pathways between open and closed states, molecular dynamics (MD) simulations can provide valuable information to better understand the conformational changes present in PBPs.^{2,3} Therefore, we used unbiased MD simulations to study the multiple ligand affinity L-lysine, L-arginine, L-ornithine binding protein (LAO-BP) of *Salmonella typhimurium*.⁷ LAO-BP displays a high affinity to four amino acids: L-lysine, L-arginine, L-ornithine, and L-histidine (Kd values: 15, 14, 30, and 500 nM, respectively).⁹ LAO-BP topology is characterized by a bilobed structure joined by a hinge region and the binding site is located at the interface of the two lobes. The lobe-1 is discontinuous containing the residues 1–88 and 195–238, whereas the lobe-2 is continuous and contains the residues 93–185. PDB structures for the free LAO-BP protein

Additional Supporting Information may be found in the online version of this article.

Abbreviations: LAO-BP, L-lysine, L-arginine, L-ornithine binding protein; PBP, periplasmic binding protein; His-H, histidine molecule with positive charge; His-ND1, histidine molecule with neutral charge, proton bound to the N δ ; His-NE2, histidine molecule with neutral charge, proton bound to the N ϵ .

Grant sponsor: DGAPA-UNAM; Grant number: IN216207; Grant sponsor: CONACYT, México; L. D-R's present address: Department of Molecular and Cellular Biology, College of Biological Sciences, University of California at Davis, CA 95616, USA

*Correspondence to: A. Sosa-Peinado, Departamento de Bioquímica, Facultad de Medicina, Universidad Nacional Autónoma de México, C.P. 04510, México, D.F. E-mail: asosa@bq.unam.mx.
Received 13 October 2010; Revised 18 February 2011; Accepted 3 March 2011
Published online 21 March 2011 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.23030

and in complex with each ligand are available.^{7,10} However, the RMSD between each ligand-bound structure are less than 0.18 Å and does not explain the ligand-dependent conformational plasticity proposed from experiments made in the homologous protein HisJ.¹¹

In addition, understanding the molecular basis of multiple ligand binding of PBPs to amino acids is of great relevance due to its high structural similarity to the soluble portion of several protein receptors that bind amino acids, like: the C class G protein coupled receptors (GPCRs),¹² some family of tyrosine kinases,¹³ and a histidine kinase receptor.¹⁴ Several phylogenetic studies strongly suggest that PBP-like domains that bind basic amino acids were derived from the bacterial LAO-BP, therefore understanding the atomistic details of the binding mechanisms may help in the design of new drugs with targets on cell signaling pathways involving GPCRs or parasite's tyrosine kinases.

The structural homology between PBPs had lead to the proposal of a general mechanism for its conformational changes.¹ However, MD studies had shown that slight differences in the simulation conditions and the methods to measure the conformational states produce significant differences in the observable conformational changes. For example, in a previous studies of LAO-BP the transition from the closed to the open state exhibited a strong dependency on the specific starting closed-apo structure.² Simulations performed with the glutamine-BP had show a similar dependency on the starting conditions.¹⁵ Therefore, it is fundamental to increase the MD sampling and improve the measurement methods in order to observe an overall picture of the conformational transitions present in these systems.

To study the multiple binding affinity mechanism present in the LAO-BP, we carried out a total of 322 MD simulations of 20 ns (6.44 μs of simulation time). Simulations for the open and the closed structures, in the absence (apo-LAO-BP) or presence of ligands (holo-LAO-BP) were performed. The ligands used were as follows: L-arginine, L-lysine, and L-histidine and two uncharged tautomers of histidine (L-his-ND1 and L-his-NE2). Overall our results suggest a common binding mechanism: different combinations of opening and twisting angles make available specific sets of hydrogen bonds interactions for each ligand, while a concerted a cation-π interaction of the ligand with two aromatic residues (Tyr-14 and Phe-52) stabilizes the ligand position and ultimately the closed state.

METHODS

Protein structures

The PDB: 2LAO (empty open state) was used for the open state,¹⁶ whereas the PDB: 1LAG (closed L-histidine bound) was used for closed state.⁷ Structures were checked and repaired for missing lateral chains using the “what if” suite.¹⁷ To eliminate any variability related to

slight differences on the initial structures of the holo-closed states, the 1LAG structure was used to generate the complexes with the ligands: L-arginine and L-lysine. To generate the structures of the different holo-open states, we followed an analogue procedure: the PDB structures 1LAF (arginine bound), 1LST (lysine bound), and 1LAG (histidine bound) were structurally aligned to the lobule I of 2LAO (residues 6–88 and 195–227), and the ligand coordinates were extracted and inserted in the 2LAO structure.

Molecular dynamics

The MD simulation package GROMACS ver. 3.3.1^{18,19} and 4.0.5.¹⁹ were used. Each system was solvated in a dodecahedron with 11,500 SPC water molecules²⁰ and enough counter ions to keep the system neutral. The minimal distance of the protein to the boundary was 1.0 nm. Initial energy minimization was performed using the steepest descent algorithm. A 250 ps position restrained MD was then performed. For each condition, the final configurations from position restrained MD were then used to initiate 23 replicas of 20 ns production NTP MD simulations. For each replica, different initial velocities were assigned. Production MDs used the force field GROMOS96 53a6,²¹ protein and solvent were separately coupled to temperature and pressure using Berendsen's thermostat and barostat²² at 318 K and 1 bar, and coupling constants of 0.1 and 1.0 ps respectively; compressibility of the system was $4.5 \times 10^{-5} \text{ bar}^{-1}$; SETTLE algorithm was used to constrain water molecules²³; a twin cut off of 1.0 nm was used for short and long range forces; PME was used^{24,25}; Neighbours list was updated every 10 steps. A 5 fs integration time step was used (Supporting Information). When the trajectories were analyzed in terms of essential dynamics analysis,²⁶ the PCA was based on the diagonalization of the covariance matrix of the protein α-carbons atomic fluctuations. Further details of the system setup and MD simulation can be found on the SI.

Measurement of open and twist movements of LAO-BP by dihedral angles

The dihedral angles were calculated between the normal vectors of two planes. Each plane was described by three points (the centres of mass of three different α-carbons clusters, see Fig. 1). Since the opening and twisting movements can be essentially described as hinge motions, the angles were defined using “two planes that share two points” (Fig. 1). The α-carbons clusters defining each plane were manually selected and refined by trial and error by the following criteria: (i) Describe either opening or twisting movements but not both, (ii) obtain a large difference for the measured values between the open and closed crystallographic structures; (iii) a high correlation of dih-twist and dih-open to the first and

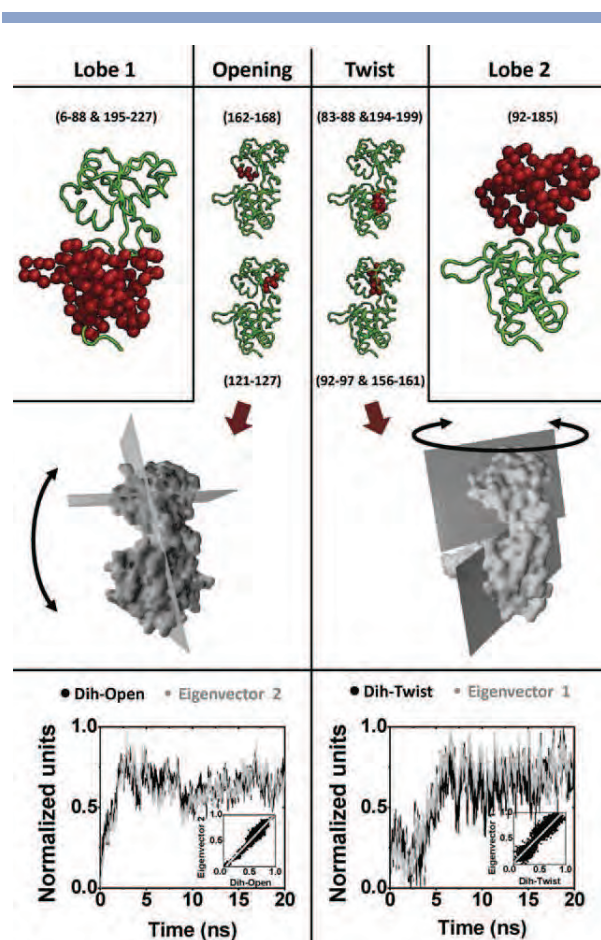


Figure 1

Clusters of C α atoms used to measure the twisting and opening movements. Top-Panel: The protein is shown in ribbons and balls representations. The balls indicate the clusters of C α atoms used to measure the opening and twisting dihedral angles (Methods section: “Measurement of open and twist movements of LAO-BP by Dihedral Angles”); Middle-Panel: Schematic representation of the planes used for the measurement of dihedral angles given the clusters of atoms in the “top panel”: at the left “the opening movement” and at the right “the twisting movements” are shown; Bottom-Panel: Correlation of the opening and twisting dihedral angles to the PCA. The comparisons of normalized data for the dih-open to the second eigenvector or the dihtwist to the first eigenvector are shown. The insets show a plot of dihedrals versus the corresponding eigenvector (correlation values: $R = 0.83$, $R = 0.84$ for opening and twisting, respectively). Dihedrals are shown in black and eigenvectors in gray color. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

second eigenvectors. Given this definition, the planes for the opening angle were as follows:

Plane-A: Residues 6–88 and 195–227; 162–168; and 121–127.

Plane B: residues 92–185; 162–168; and 121–127.

Therefore, the shared points between these two planes are as follows: 162–168 and 121–127.

For the twisting angle the planes were as follows:

Plane-A: residues 6–88 and 195–227; 83–88 and 194–199; and 92–97 and 156–161;

Plane-B: residues 92–185; 83–88 and 194–199; and 92–97 and 156–161.

Therefore, the shared points between these two planes are: 83–88 and 194–199; 92–97 and 156–161.

By definition, the holo X-ray structure (PDB ID: 1LAF) has both opening and twisting angles equal zero, whereas the apo X-ray structure (PDB ID: 2LAO) has opening angle of 38.2° and twisting angle of -26.2° . An analysis of the correlation between these dihedrals and the two slowest eigenvectors from principle component analysis (PCA) shows a correlation factor of 0.84 for the opening to the second eigenvector, and 0.83 for the twisting to the first eigenvector. The PCA also reveals that the first two principal components explain more than the 60% of the total protein movements (Fig. 2). We also investigated the movement described by the third eigenvector and found that it is a previously described movement for PBPs,²⁷ we call this movement wobbling, and it could be interpreted as a low frequency vibration of the protein loops. We also found that the opening and twisting angles are enough to reveal the same overall changes than RMSD (all C α) measurements, but with the advantage of being intuitive and avoid the need of know a “target/reference structure” to achieve the measurement (Supporting Information Fig. S1). Taken together, these data shows that the dihedral angles are a reasonable descriptor of the protein’s conformation.

Measurement of stacking interaction

To determine if stacking interactions were present, we first defined geometrical planes composed by three atoms for each of the side chains of: L-arginine (NE, NH1, and NH2); L-histidine: (CG, ND1, and NE2); Tyr-14 (CG, CE1, and CE2); and Phe-52 (CG, CE1, and CE2).

Then, we computed the dihedral angles between the normal vectors of two planes. We discretized the definition of stacking as: (0) no interaction or (1) interaction. An stacking interaction was defined as: “angle between the planes” = $\{-180^\circ | 0^\circ | 180^\circ\}$, with a tolerance of $\pm 30^\circ$. Thus, only those structures in which the aromatic rings of the Tyr-14 or Phe-52 are essentially parallel to the side chain of the compared molecule are considered as stacking interactions. Each MD trajectory was analyzed using this method and the average was used to compute the fraction of staking (P).

Data clustering

To study the ligand binding mechanism, we performed a K-center clustering on the ensemble generated by the MD trajectories exhibiting complete transitions from the open to the closed state. The clustering was realized over the euclidean distance of the opening and twisting angles and was performed separately on MDs with the ligands: L-Arg or L-Lys (8 and 5 MDs, respectively). The number

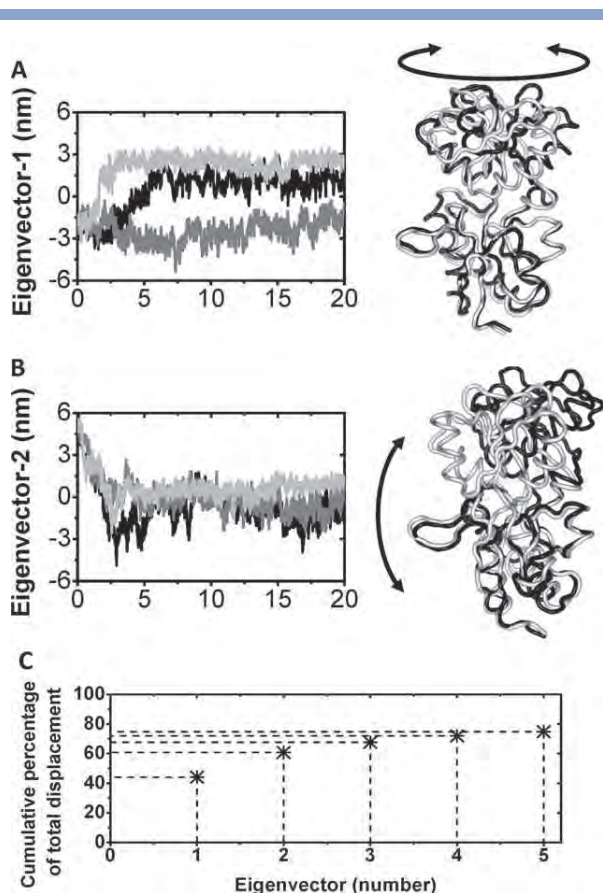


Figure 2

First and second eigenvectors found in PCA analysis of simulations started from closed apo-LAO-BP. **A:** Plot of the first eigenvector versus the time in three independent 20 ns MD simulations started in the close-apo structure (each simulation is shown in a different color). A ribbon representation of the movement related to the first eigenvector: “the twisting” is shown. **B:** Plot of the second eigenvector found in three independent 20 ns MD simulations started in the close-apo structure (each simulation is shown in a different color). A ribbon representation of the movement related to the second eigenvector: “the opening” is shown. **C:** Cumulative percentage of the total movements of LAO-BP explained by each PCA eigenvector. MDs were started in the apo-closed state.

of K-centers used was as follows: 10 for MDs with L-arg and 5 for L-lys. After the clustering, it was easy to identify the states that were present at the beginning, intermediate, and ending parts of the open to closed transition. Essentially, we found that four states contain all the relevant information to dissect the transitions from the open to the closed state. After identifying these four states, the other states were not taken into account, and for easy of visualization, the results (cluster number) were ordered to progressively match the transition from the open state to the closed state (e.g., State 1 is the open state, States 2 and 3 are intermediates, and State 4 is the final closed state).

Computational resources

AMD[®] Opteron processors, running at 2.6 GHz on 64-bit Linux at the supercomputing facility “Kan-Balam,” DGSCA, Universidad Nacional Autónoma de México. Intel[®] Xeon processors, running at 3.0 GHz on 64-bit Linux at the supercomputing facility “Aitzalooa,” Universidad Autónoma Metropolitana, México.

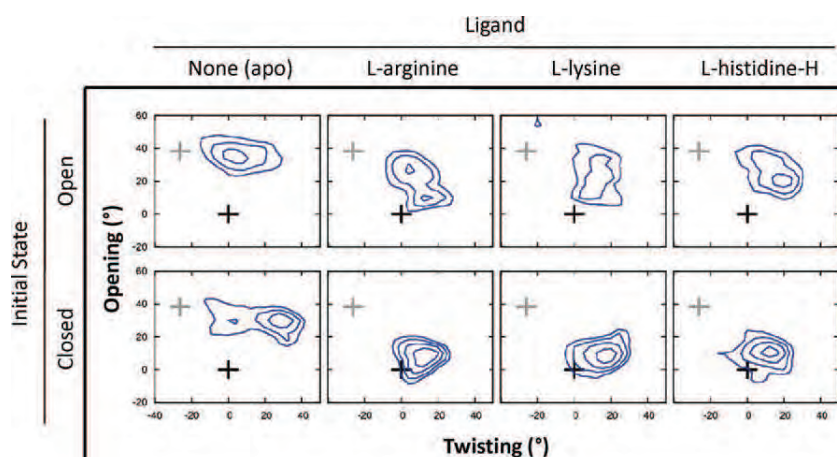
RESULTS

To investigate the conformational changes present in LAO-BP, we first come across how to best measure the conformational state of the protein. We found that in agreement to previous reports, the open to closed state transition was mainly described by the first two eigenvectors from PCA (Fig. 2). The visual projection of eigenvectors strongly suggested a correlation of the first and second eigenvectors to the twisting and opening movements, respectively (Figs. 1 and 2). To quantify the conformational changes, we built a geometrical measurement based on the measurement of dihedral angles describing the twisting (dih-twist) and opening (dih-open) movements (Fig. 1 and Methods). We chose dihedral angles over C α -RMSD measurements because the opening and the twisting dihedrals provided more detailed and intuitive information (Methods). Furthermore, this approach provides an independent measurement of opening and twisting movements, hence allowing an overall the description of the conformational state of the protein.

MD of apo-LAO-BP

When MD simulations were started with the Apo protein either in the open or closed states (23 replicas for each condition), the protein reached and maintained an open conformation (Fig. 3 and Table I). We found that the twisting angle of the conformations in the open ensemble differed significantly from those in the crystallographic open structure 2LAO (Fig. 3 and Table I). This is not unexpected since crystals conditions usually imply large contact areas between the molecules forming the crystals. These contacts could easily stabilize an infrequently populated “open” state. Interestingly, when simulations were started from the apo-closed state, the protein always reached an open conformation. However, two populations with different twisting angles were found (Figs. 3 and 4): one equivalent to simulations started from the apo-open-state and a second ensemble with higher values of twisting that we describe as the “open-twisted” conformation.

A hydrogen bonds analysis revealed one interaction present preferentially in the open state: the amino acids pair 78–188 (Table II). The open-twisted state revealed two hydrogen bond interactions, amino acids 76–188 and 77–190 (Table II).

**Figure 3**

Conformational space visited by LAO-BP in presence or absence of ligands. The contours show the overall conformational space (opening vs. twisting dihedral angles) for 23 MD replicas of each condition. Each contour is proportional to the population density. For ease of comparison plotted values are on a negative logarithmic scale ($-\log P$) such every contour imply greater difference, (for each axis bin size is 4°). The contour cut off is $-\log(P) = 5$ and each inner level is -0.5 . Gray and black crosses show the values for the crystallographic open or closed states, respectively. The data corresponds to the last 10 ns of each MD simulation. The protein initial state (open or closed) and the ligand are indicated. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

To assert whether the open-twisted state represents an intermediate on the pathway from the open to the closed state, we projected the vectors of movement for the MD simulations started in the apo-closed state (Fig. 4). None of the simulations displayed pathways going through the open-twisted state to the fully open state, besides we found that some of the simulations even went first to the fully open state and then transited to the open-twisted state (Fig. 4). This strongly suggests that the open-twisted state is not an obligate intermediate in the closed to open transition and its presence in simulations started from the closed state may be related to some hysteretical effect.

MD of holo-LAO-BP

To study the conformational changes induced by cationic ligands, we placed each of the LAO-BP ligands

(L-arg, L-lys, and L-his) into the binding site of the closed and open conformers. For L-histidine, we used the molecule with positive formal charge (L-his-H). Each condition was examined by 23 replicas of MD.

MD started at the holo-closed state

In presence of any cationic ligand, the closed conformation remained stable (Fig. 3 and Table I). The measurement of the ligand distance to the binding site (lig-dist) confirmed that the ligands remained in the binding pocket (Table I). However, different combinations of opening angle, twisting angle, and H-bonds were found for each ligand (Fig. 3, and Tables I and II). The side chains of the ligands L-arg and L-his-H exhibited a stacking cation- π interaction with the aromatic residues Tyr-14 and Phe-52 (Fig. 5 and Table III), we found that

Table I
Angles and Lig-Distances Reached from MD Started at the Closed Conformation

Starting conformation	Reached conformation	Dih-opening ($^\circ$)	Dih-twisting ($^\circ$)	Lig-distance (\AA)
Open (Crystallographic)	—	38.22	-26.22	—
Closed (Crystallographic)	—	0.0	0.0	0.0
apo-open	Open	36.95 ± 10.1	7.19 ± 14.4	—
apo-closed	Open	32.56 ± 9.5	-2.67 ± 10.1	—
	Open-twisted	30.58 ± 6.0	28.06 ± 9.6	—
arg-closed	Closed	10.97 ± 9.7	10.75 ± 9.4	0.005 ± 0.220
lys-closed	Closed	12.12 ± 8.2	15.06 ± 10.5	0.113 ± 0.554
his-H-closed	Closed	14.50 ± 11.2	9.64 ± 10.7	0.508 ± 0.873

Values are averages over the last 10 ns of 23 independent MD replicas for each starting conformation, the standard deviation is shown. The First two rows show the values found in the Open and Closed crystal structures of LAO (PDB: 2LAO and 1LAG, respectively).

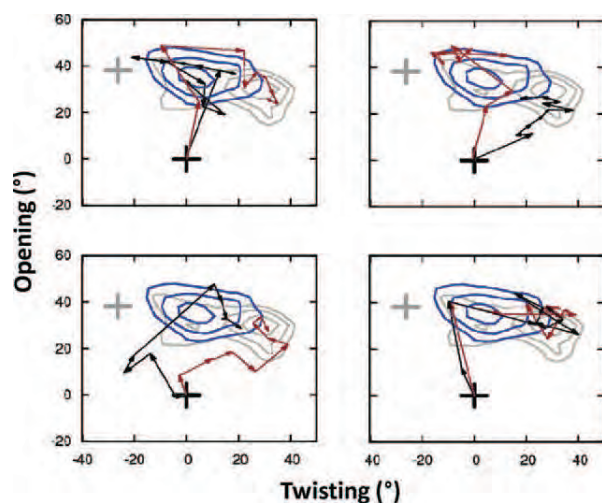


Figure 4

Projections of pathways from the closed to the open state. The movement vectors (arrows, each simulation is shown in a different color) show a total eight representative pathways that the LAO-BP followed in the transition from the closed to the open state. For reference, contour plots are shown for the open ensemble reached by simulations started in the open (Blue) or closed (Black) states (same as the ensembles shown in Fig. 3). The time-step between the vectors data points is 2 ns. These graphs essentially reveal two pathways to reach the open-twisted state: (1) Direct transition, or (2) Transition through the fully open state as intermediate. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

the stacking between L-his-H (Table III) and the protein seems weaker than with the L-arg (Table III). Noncharged histidine tautomers exhibited the lowest fraction of stacking.

In the case of L-lys, since the side chain does not present a planar structure, it was not possible to study the

stacking interaction. However, the analysis of the central conformations for each ligand suggests that the side chain of the L-lys ligand displays an alternating cation- π and hydrophobic interaction with the residues Tyr-14 and Phe-52 (Fig. 6). This means that when the L-lys ligand is interacting through a cation- π either with the Tyr-14 or the Phe-52, it is establishing a hydrophobic interaction with the other amino acid (Fig. 6).

To study a possible role of the ligands as a bridge (or coupling) interaction between the residues Tyr-14 and Phe-52, we studied the angle formed between the planes of the rings of these residues. The histograms revealed that in the presence of any of the ligands a well-defined population arises (Fig. 6). In contrast, in the absence of ligands, the histogram shows a different distribution of the angles (Fig. 6). This evidence suggests that the interaction with any native ligand enhances the coupling of the aromatic residues Tyr-14 and Phe-52.

It is interesting that, in the presence of L-lys, the LAO-BP closed state reached a higher twisting angle than those found for L-arg and L-his-H (Table I).

For all the ligands, we found a distinctive H-bond signature of the closed state: residues pair 15–143. The complex with the ligand L-his-H displayed an extra interaction: residues pair 13–143. Additionally, one of the H-bond signatures of the open-twisted state (residue pairs 77 and 190) was also present in the L-lys simulations (Table II).

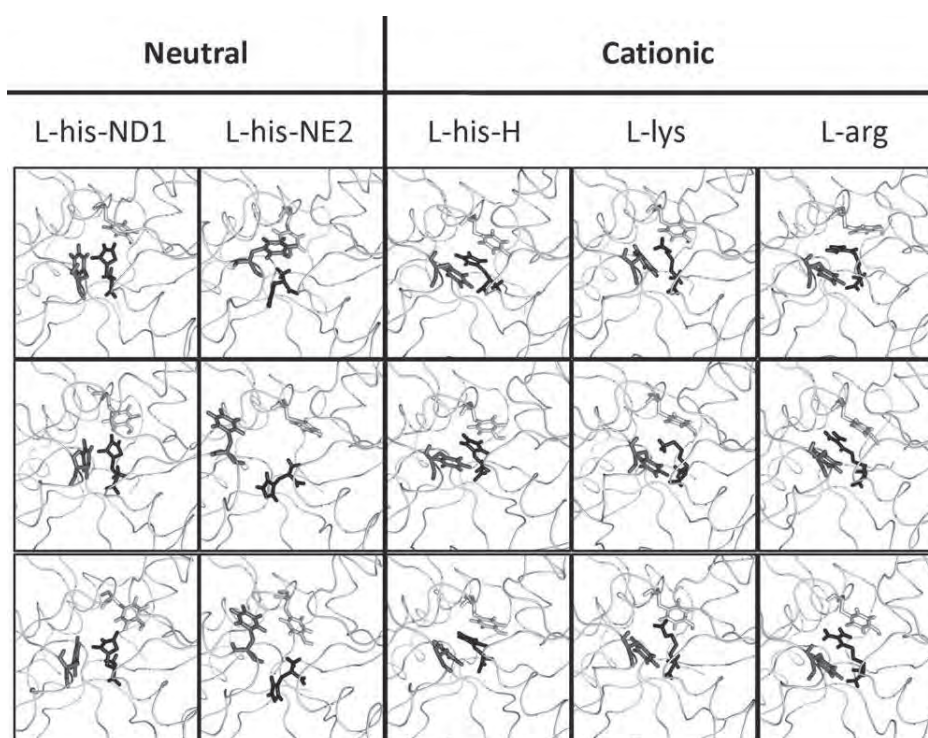
MD started at the holo-open state

Using L-arginine or L-lysine as ligands, we were able to capture several holo transitions from the open to the closed state. From the 23 MDs in each condition, the number of simulations exhibiting transitions to the closed state was as follows: 8 for L-arginine and 5 for

Table II
Fraction of H-Bonds Involved in Ligand Binding and Protein Conformation

		Conformation (Starting → Final)					
			Apo-open ↓ Apo-open	Apo-closed ↓ Apo-open/or/twisted	L-arg-closed ↓ L-arg-closed	L-lys-closed ↓ L-lys-closed	L-his-H-closed ↓ L-his-H-closed
Signature H-bond	Open	78–188	0.195	0.110	0.035	0.049	0.0722
	Twisted	76–188	0.093	0.384	0.00004	0.070	0.004
		77–190	0.039	0.230	0.016	0.113	0.051
Closed	13–143	0.0008	0.002	0.066	0.040	0.199	
	15–143	0.052	0.243	1.150	1.330	0.683	
Ligands	11—LIG	—	—	0.797	0.140	0.204	
	13—LIG	—	—	0.078	0.136	0.026	
	69—LIG	—	—	0.403	0.156	0.049	
	70—LIG	—	—	1.060	0.621	0.533	
	72—LIG	—	—	1.400	1.130	1.060	
	77—LIG	—	—	1.370	0.969	0.989	
	119—LIG	—	—	0.163	0.075	0.014	
	121—LIG	—	—	1.270	1.620	1.560	
	161—LIG	—	—	0.078	0.091	0.633	

The average fraction of H-bonds found in simulations started in the apo-open, apo-closed, and holo-closed systems is shown. The data were analyzed over the last 10 ns of 23 independent MD replicas for each starting conformation. Data with a fraction <0.1 were highlighted.

**Figure 5**

Interaction of ligands with residues Tyr-14 and Phe-52. The backbone of LAO-BP is shown in a ribbon representation. The ligand is in black; Tyr-14 is in light gray; and Phe-52 in dark gray. To avoid any bias, the conformations present in this figure were randomly selected from the major conformational cluster found for each ligand (three random conformations are shown). The clusters were generated separately for each ligand using a K-center clustering algorithm over the euclidean distance of the opening and twisting. 20 K-centers clusters were generated, the total population (number of frames) clustered were 92,000 for each ligand, and the population of the major clusters were: Arg = 22,065, Lys = 34,933, HisH = 31,190, His-ND1 = 25,486, and HisNE2 = 21,850.

L-lysine. For the ligand L-Histidine, we did not find any complete transition to the closed state but we found three transitions to a partially closed state.

To extract useful information from these ensembles, we generated a K-center clustering separately on those simulations with complete transitions to the closed state (Methods). The analysis divided the transitions for L-arg or L-lys in four clusters (Fig. 7). Once the data was clustered, we extracted 100 random conformations from each state and analyzed the previously identified signature hydrogen bonds. The data from this analysis can be found in the Supporting Information Tables S1 and S2. Our analysis suggests that when a successful open to

closed transition took place the H-bond formation followed the next general steps:

1. Generation of ligand interactions with lobule I (residues 11, 70, 72, and 77), and specifically for the ligand L-lys also an interaction with the lobule II (residue 121).
2. Strengthening/or/formation of the interaction with the residue 121.
3. Decrease of the open-state interaction signature (residue pair 78–188) and formation of the closed-state interaction signature (residue pair 15–143). Further strengthen of the previously achieved interactions.

Table III

Ligand Stacking with LAO-BP Residues Tyr₁₄ and Phe₅₂

		L-arg closed	L-his-H closed	L-his-ND1 closed	L-his-ND2 closed
Stacking	Lig <-> Tyr ₁₄	0.59	0.65	0.19	0.14
	Lig <-> Phe ₅₂	0.80	0.45	0.17	0.15

The table resumes the average fraction (P) of stacking interaction found in simulations started in the apo-open, apo-closed, and holo-closed systems. In the analysis, the possible values were 0 (absence of interaction) or 1 (fully established interaction); see “Methods” for further details on the stacking interaction measurements. The shaded squares denote histidine tautomers with neutral charge. The data was analyzed over the last 10 ns of the MD simulations.

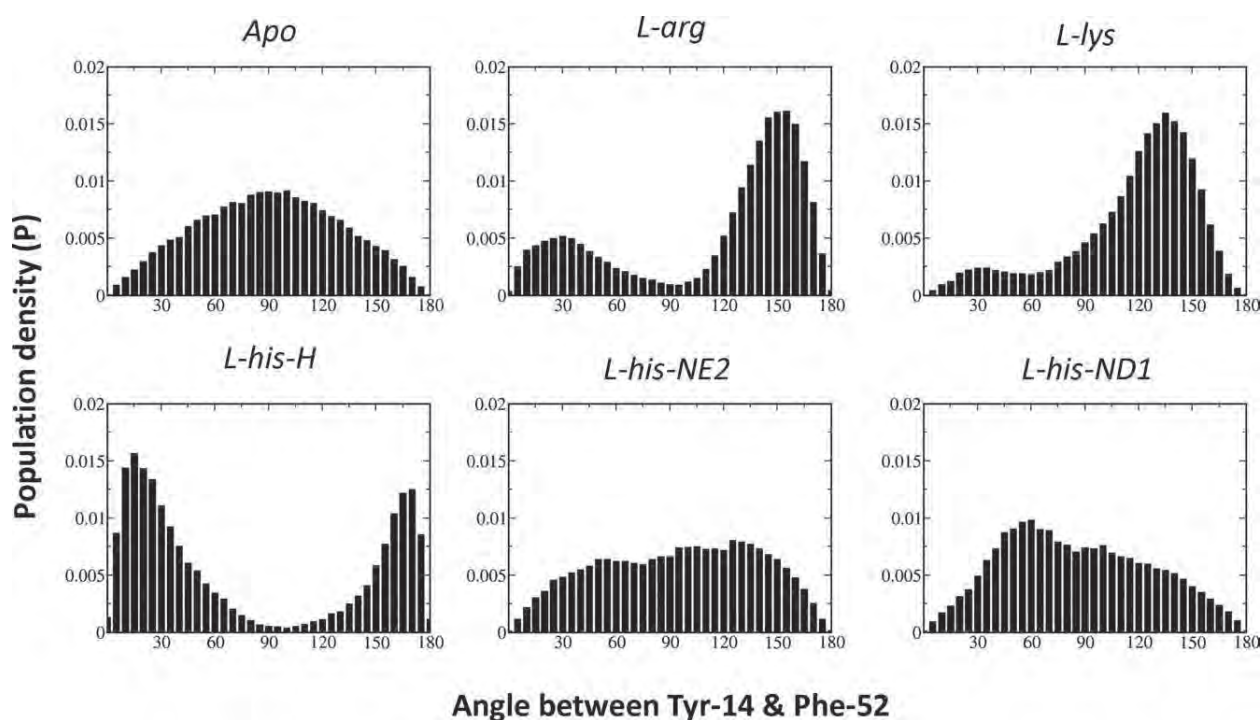


Figure 6

Histograms of the angle between the planes of the Tyr-14 and Phe-52 side chains, for simulations started in the closed state. The plots show that in the absence of ligand (Apo), the distribution of the angles is distributed around 90° . In the presence of any of the cationic ligands (L-arg, L-lys, or L-his-H), the angle distribution show well-defined populations, this evidence a correlation between the presence of the ligand and the generation of a coupling between the Tyr-14 and Phe-52. In the presence of neutral tautomers of L-histidine (L-his-ND1, L-his-NE2), the angles distributions resembles the ones found in the “Apo” state. The data corresponds to the last 10 ns of each MD simulation.

- Open-state signature interaction is lost and further enhancement of the previously achieved interactions occurs.

MD of LAO-BP in presence of noncharged L-histidine ligands

We tested the effects of uncharged histidine tautomers as ligands on the closed state of LAO-BP. The ligands used were L-his-ND1 (proton bound to N δ) and L-his-NE2 (proton bound to N ϵ). Twenty-three replicas of MD studied each condition.

For each ligand, we observed several events of destabilization of the closed conformation, and some of these were followed by the loss of the protein–ligand interaction. In the case of the L-his-NE2, we observed eight events of destabilization of the open state and three of them were accompanied by the lost of the ligand. In the case of the L-his-ND1, the effect was more dramatic since we observed 10 events of destabilization of the closed state and four of them were accompanied by the loops of the ligand. The Figure 8 shows the overall conformational space on the last 10 ns for all the simulations of each condition. It is observed that the overall closed conformation

has been destabilized and driven toward the open state. Our analysis of the stacking interaction (Figs. 5 and 6 and Table III) shows that the effect of either of the L-his tautomers as ligands is similar to that found in apo MDs.

DISCUSSION

Structural and biochemical data show that the ligand-induced open to close transition is fundamental for PBPs function.^{4,5,7,11,28} NMR and MD studies had revealed this transition results from two movements: opening and twisting.^{1,3,29} Most theoretical work on PBPs conformational changes has been analyzed by two methods: protein RMSD changes and the matrix overlap of principal components.^{1,2} Since RMSD is a global measurement of all the structural changes taking place, then the functionally relevant movements can be hidden by random fluctuations^{20,21} or combined movements (Supporting Information Fig. S1). In PCA, the eigenvectors order depends on the magnitude and speed of the movements in the analyzed trajectory; therefore, it cannot be guaranteed that the first or second eigenvector will correspond to the same conformational change in different trajectories. In this work, we present a method based on dihedral

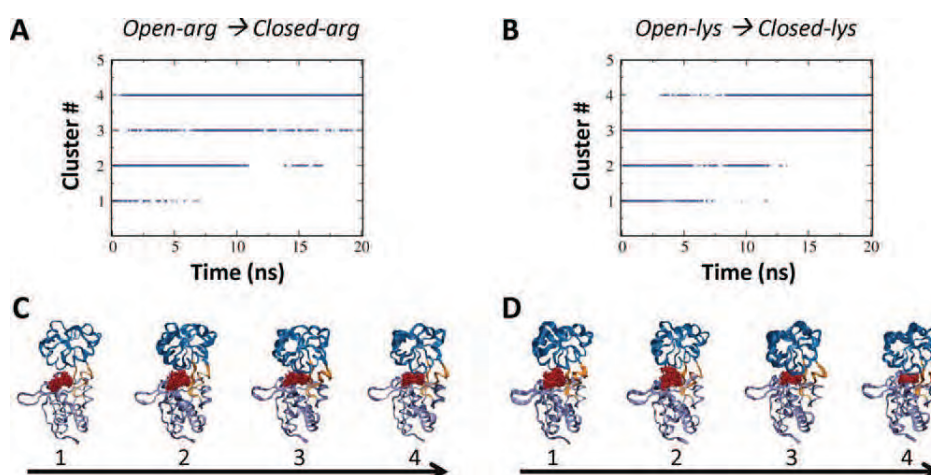


Figure 7

Ligand-induced transitions from the open to the closed state. The panels show the Time plot of the structural clusters identified in the transitions from the open to the closed state: (A) When the ligand was L-arg, and (B) ligand L-lys. The panels C and D) show the overall structure present in each cluster (Protein is represented in blue and purple ribbons and the ligand is represented with red sticks) when: (C) the ligand was Arginine and (D) the ligand was Lysine. Each protein representation shows 100 structures randomly picked from each cluster. The clusters were ordered with the objective of progressively displaying the transition from the open to the closed state (i.e., in the cluster 1 the protein is open, the clusters 2 and 3 are intermediates, and to the cluster 4 belongs the fully closed state).

angles to measure the main movements present in LAO-BP. This methodology has two advantages: is accurate and allows the independent measurement of the opening and twisting movements. This method could be easily extended to other molecules with hinge-mediated conformational changes.

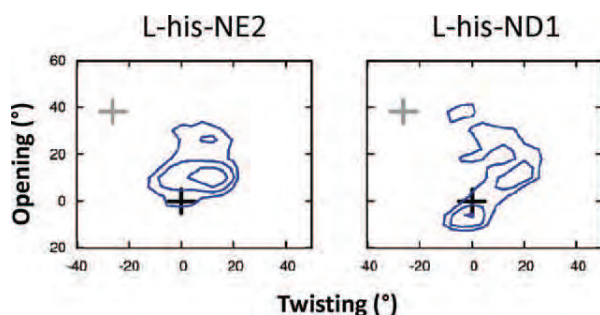


Figure 8

Conformational space visited by MDs of closed-LAO-BP in presence of neutral histidine tautomers as ligands. The contours show the overall conformational space (opening vs. twisting dihedral angles) for 23 MD replicas of each neutral histidine tautomer: L-his-NE2 or L-his-ND1. The plot reveals that in a clear difference to the cationic ligand L-his-H, the neutral tautomers fail to maintain the closed conformation of the protein. Each contour is proportional to the population density. For ease of comparison plotted values are on a negative logarithmic scale ($-\log P$) such every contour imply greater difference, for each axis bin size is 4° . The contour cut off is $-\log(P) = 5$ and each inner level is -0.5 . Gray and black crosses show the values for the crystallographic open or closed states, respectively. The data corresponds to the last 10 ns of each MD simulation. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

In agreement to PBP experimental data, we found that in the absence of ligand, the protein shifts its conformation to the open state ensemble. While this might be a consequence of the force field used, it has been reported that at least for small proteins different force fields (OPLS-AA, AMBER03, and GROMOS95 53a6) yield essentially the same results.³⁰

We found differences between the MD reached open conformational ensemble and the open crystallographic structure of LAO-BP.¹⁶ This prompts us to suggest that the open crystallographic structure belong to a low populated conformation of the native ensemble.

When the twisting motion of LAO-BP was taken into account, the open ensemble exhibited two different conformations: open and open twisted (Table I). Furthermore, each conformation presented a defined pattern of H-bonds (Table II). Interestingly, an analysis of the pathways followed by the simulations that reached this open-twisted state ruled out the possibility of this state as an obligate intermediate in the mechanism from the closed to the open state.

In the presence of ligands we found: (i) different average closed conformations for each ligand and (ii) specific patterns of protein–ligand hydrogen bonds interactions for each complex (Table II). Our analysis also reveals that the conformational spaces of the different closed states far from being small and sharp basins in the conformational space are wide regions in which the protein can transit from tightly closed conformations ($\sim 0^\circ$) to moderately open conformations ($\sim 25^\circ$). To compare whether in the presence of each ligand, the conformational

Table IV
Statistical Comparison of MD Simulations Started in Different Conditions

		Open		Closed	
Opening dihedral		Empty (Apo)	Empty (Apo)	L-arg	L-lys
Closed	Empty (Apo)	0.022	—	—	—
	L-arg	3.9^{-14}	7.8^{-12}	—	—
	L-lys	1.9^{-15}	1.4^{-13}	0.894	—
	L-his-H	3.7^{-10}	7.9^{-08}	0.253	0.260
Twisting dihedral		Open		Closed	
		Empty (Apo)	Empty (Apo)	L-arg	L-lys
Closed	Empty (Apo)	0.079	—	—	—
	L-arg	0.258	0.241	—	—
	L-lys	0.011	0.940	0.0318	—
	L-his-H	0.552	0.141	0.521	0.014

The table shows the results of a *t*-test employed to compare the Opening (Upper Table) and Twisting (Bottom Table) dihedral angles distributions (Opening and Twisting) for the different initial conditions studied by MD. The test was two-sided and hence the alternative hypothesis was that “the true difference in means is not equal to 0.” The variance was estimated separately for each group and the Welch modification to the degrees of freedom was used. The input dataset was 23 observations (averages of the last 10 ns for each of the 23 independent MD replicas) for each starting conformation simulated by MD. The shaded fields denote when the difference within the datasets is statistically relevant within a 95% significance level.

spaces were statistically different we performed a *t*-test. The results show that the conformational spaces are statistically different between any of the closed states (in the presence of ligands) and the open state (Table IV). Furthermore, the analysis also shows that the degree of twisting is significantly different when the ligand is L-lys compared with any of the other ligands (Table IV); however, we did not find a statistical significance in the difference between the conformation generated by L-histidine and L-arginine (Table IV). An analysis of the hydrogen bonds and stacking interactions revealed further evidence to support that each protein–ligand interaction presents structural differences. We believe that a plausible interpretation of these results is that all the protein–ligand ensembles in the closed state belong to a bigger conformational space that represents the overall closed state (therefore the conformational spaces present some overlap). Thus, the effect of the interaction with each of the different ligands is to displace the equilibrium to a particular region of the closed state. We stress that this high flexibility of the closed state may be a characteristic related to the multiple binding affinity. This variability could also explain why the crystal structures in the presence of the different ligands are essentially identical, since a high flexibility of the closed state combined with similar crystallization conditions could easily drive the protein to similar conformations. In addition, the volume calculated for unit cell in our MD simulation was 419.6 nm³, whereas the unit cell of the crystallographic structures 2LAO and 1LAF were 286 and 261 nm³ respectively, therefore, in the crystal the protein has less solvent space and the induced contacts may hinder

some of the flexibly or reduce the structural differences compared with the conditions in our MD simulations. Further studies are required to clarify this hypothesis. Nevertheless, our results are in agreement to previous experimental work in the homologous PBP HisJ¹¹ in which Wolf *et al.* found evidence of different protein conformations in the presence of L-arginine or L-lysine.¹¹ Our results provide a potential explanation for their results: the L-lysine ligand promotes a closed-twisted state that is different to the conformation observed with the ligand L-arginine.

Using simulations started on the open state in the presence of ligands, we found several spontaneous transitions to the closed state. A combination of a K-centers clustering and the results from previously identified hydrogen bonds allowed us to describe a general mechanism of hydrogen bonds formation in the transition from the open-holo complex to the successful closed complex. The progressive formation of the hydrogen bonds suggests that an induced fit mechanism may be involved in the open state to closed state transition. However, additional improvements to the data mining could reveal further insights into mechanism of the open to close transition.

The case for the ligand L-his is of special interest. Experimental evidence suggests that for the this ligand the protonation state plays an important role on LAO-BP affinity.⁹ Our simulations concur, indicating that the positive charged L-histidine ligand generates a stable complex with the protein, whereas neutral charged histidine tautomers are unable to do so. Our analysis of the histidine tautomers points that a stable stacking of the cationic L-histidine side chain to residues Tyr-14 and Phe-52 can be one of the main players into explain how the slight difference of one proton could drive such dramatic selectivity. As discussed, the stacking interaction was found to be present for the closed state in presence of any of the cationic ligands, but the stacking was found to be disrupted when the ligand was any of the uncharged tautomers (Figs. 5 and 6). We found the ligand stacking to be extremely interesting, since, instead of a pure stabilizing interaction, the evolution had selected an “unlikely” interaction between aromatic residues and cationic ligands. It can be hypothesized that in the binding site of LAO-BP, the cation– π stacking could provide at least two advantages: (i) Is extremely selective, since only a ligand that properly fits in the binding site can reach the correct position to “stack” and thereby to stabilize both: the position of the ligand and the interaction with the aromatic residues 14 and 52, (ii) Since the hydrophobic nature of these residues, when there is not ligand and hence no staking, the side chains of these residues are highly mobile and then a destabilizing contribution to the closed state. We suggest that these two particular amino acids (Tyr-14 and Phe-52) could be excellent targets for directed mutagenesis studies in order to further

investigate the molecular mechanism of multiple ligand binding.

CONCLUSIONS

Although the crystallographic structures of LAO-BP suggest that the binding site structure is essentially the same for any ligand (L-arg, L-lys or L-his), our results indicate that the protein adopts slightly different conformational assembles in order to properly fit each ligand. This fine tune results from a combination of opening and twisting movements.

Recently, the relevance of cation- π interactions between aromatic amino acids and cationic amines has been recognized, in both protein-protein and protein-ligand interactions. A few examples are the ligand binding in the glycine and betaine binding proteins and different drug receptors.^{31–33} Our work shows that this kind of interaction has an important role on the LAO-BP recognition of the ligand side chain, and most probably similar mechanisms should be present in the soluble part of receptors that bind amino acids through LAO-BP like domains. On the basis of these results, we suggest that the multiple affinity of LAO-BP for cationic L-amino acids could be given by a ligand (side chain) selectivity mediated by the cation- π interaction and an induced-fit mechanism that displaces the equilibrium of the closed state to the most complementary binding site for each ligand. Therefore, the conformational dynamics of LAO-BP provides the specific H-bonds and other important interactions to bind each ligand. The observation of specific conformations in each protein-ligand interaction is in agreement to the experimental findings of Wolf and coworkers for the homologous PBP HisJ. Together, these results can provide valuable data to direct future experiments in order to gain a deeper understanding of the multiple ligand recognition present in LAO-BP and other PBPs and could aid in drug designs targeting soluble domains of proteins receptors¹² in which LAO-BP like structural motifs play a role in the ligand recognition.

ACKNOWLEDGMENTS

Computational resources were provided by the supercomputing facility “Kan-Balam,” DGSCA, Universidad Nacional Autónoma de México. The authors thank Dr. Luis A. Tellez and Professors Alejandro Fernández and Nina Pastor for the critical reading and corrections to this manuscript. The authors also thank to Dr. Gregory Bowman for the discussions and helpful suggestions for the data analysis. Computing resources were kindly provided by “DGSCA-UNAM Kan-Balam supercomputing facility” and “UAM-Iztapalapa Aitzaloe supercomputing facility.” D-A.S. is a recipient of the fellowship for Ph.D. studies from CONACYT, México. A.S-P. is a recipient of the fellowship from DGAPA, UNAM.

REFERENCES

- Pang A, Arinaminpathy Y, Sansom MS, Biggin PC. Comparative molecular dynamics-similar folds and similar motions? *Proteins* 2005;61:809–822.
- Hall BA, Kaye SL, Pang A, Perera R, Biggin PC. Characterization of protein conformational states by normal-mode frequencies. *J Am Chem Soc* 2007;129:11394–11401.
- Stockner T, Vogel H, Tieleman D. A salt-bridge motif involved in ligand binding and large-scale domain motions of the maltose-binding protein. *Biophys J* 2005;89:3362–3371.
- Dwyer MA, Hellinga HW. Periplasmic binding proteins: a versatile superfamily for protein engineering. *Curr Opin Struct Biol* 2004;14:495–504.
- Higgins CF, Ames GF. Two periplasmic transport proteins which interact with a common membrane receptor show extensive homology: complete nucleotide sequences. *Proc Natl Acad Sci USA* 1981;78:6038–6042.
- Locher KP, Lee AT, Rees DC. The *E. coli* BtuCD structure: a framework for ABC transporter architecture and mechanism. *Science* 2002;296:1091–1098.
- Oh BH, Ames GF, Kim SH. Structural basis for multiple ligand specificity of the periplasmic lysine-, arginine-, ornithine-binding protein. *J Biol Chem* 1994;269:26323–26330.
- Oh BH, Kang CH, De Bondt H, Kim SH, Nikaido K, Joshi AK, Ames GF. The bacterial periplasmic histidine-binding protein. structure/function analysis of the ligand-binding site and comparison with related proteins. *J Biol Chem* 1994;269:4135–4143.
- Nikaido K, Ames GF. Purification and characterization of the periplasmic lysine-, arginine-, ornithine-binding protein (LAO) from *Salmonella typhimurium*. *J Biol Chem* 1992;267:20706–20712.
- Kang CH, Shin WC, Yamagata Y, Gokcen S, Ames GF, Kim SH. Crystal structure of the lysine-, arginine-, ornithine-binding protein (LAO) from *Salmonella typhimurium* at 2.7-Å resolution. *J Biol Chem* 1991;266:23893–23899.
- Wolf A, Lee KC, Kirsch JF, Ames GF. Ligand-dependent conformational plasticity of the periplasmic histidine-binding protein HisJ. Involvement in transport specificity. *J Biol Chem* 1996;271:21243–21250.
- Cao J, Huang S, Qian J, Huang J, Jin L, Su Z, Yang J, Liu J. Evolution of the class C GPCR Venus flytrap modules involved positive selected functional divergence. *BMC Evol Biol* 2009;9:67.
- Ahier A, Rondard P, Gouignard N, Khayath N, Huang S, Trollet J, Donoghue DJ, Gauthier M, Pin JP, Dissous C. A new family of receptor tyrosine kinases with a venus flytrap binding domain in insects and other invertebrates activated by aminoacids. *PLoS One* 2009;4:e5651.
- Cheung J, Le-Khac M, Hendrickson WA. Crystal structure of a histidine kinase sensor domain with similarity to periplasmic binding proteins. *Proteins* 2009;77:235–241.
- Loeffler HH, Kitao A. Collective dynamics of periplasmic glutamine binding protein upon domain closure. *Biophys J* 2009;97:2541–2549.
- Oh BH, Pandit J, Kang CH, Nikaido K, Gokcen S, Ames GF, Kim SH. Three-dimensional structures of the periplasmic lysine/arginine/ornithine-binding protein with and without a ligand. *J Biol Chem* 1993;268:11348–11355.
- Rodriguez R, China G, Lopez N, Pons T, Vriend G. Homology modeling, model and software evaluation: three related resources. *Bioinformatics* 1998;14:523–528.
- Lindahl E, Hess B, van der Spoel D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J Mol Model* 2001;7:306–317.
- Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. *J Comput Chem* 2005;26:1701–1718.

20. Berendsen H, Postma J, Van Gunsteren W, Hermans J. Interaction models for water in relation to protein hydration. In: Pullman B, editor. Intermolecular forces. The Netherlands: Reidel, Dordrecht; 1981.331–342 p.
21. Oostenbrink C, Soares T, van der Vegt N, van Gunsteren W. Validation of the 53A6 GROMOS force field. *Eur Biophys J* 2005;34: 273–284.
22. Berendsen H, Postma J, Van Gunsteren W, DiNola A, Haak J. Molecular dynamics with coupling to an external bath. *J Chem Phys* 1984;81:3684–3690.
23. Miyamoto S, Kollman P. SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J Comput Chem* 1992;13:952–962.
24. Darden T, York D, Pedersen L. Particle mesh Ewald: an $N \log(N)$ method for Ewald sums in large systems. *J Chem Phys* 1993;98: 10089–10092.
25. Essmann U, Perera L, Berkowitz M, Darden T, Lee H, Pedersen L. A smooth particle mesh Ewald method. *J Chem Phys* 1995;103: 8577–8593.
26. Amadei A, Linssen A, Berendsen H. Essential dynamics of proteins. *Proteins* 1993;17:412–425.
27. Pang A, Arinaminpathy Y, Sansom MS, Biggin PC. Interdomain dynamics and ligand binding: molecular dynamics simulations of glutamine binding protein. *FEBS Lett* 2003;550:168–174.
28. Felder CB, Graul RC, Lee AY, Merkle HP, Sadee W. The Venus fly-trap of periplasmic binding proteins: an ancient protein module present in multiple drug receptors. *AAPS Pharm Sci* 1999;1:7–26.
29. Tang C, Schwieters CD, Clore GM. Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature* 2007;449:1078–1082.
30. Marklund EG, Larsson DSD, van der Spoel D, Patriksson A, Coleman C. Structural stability of electrosprayed proteins: temperature and hydration effects. *Phys Chem Chem Phys* 2009;11:8069–8078.
31. Dougherty DA. Cation- π interactions involving aromatic amino acids. *J Nutr* 2007;137(6Suppl 1):1504S–1508S; discussion1516S–1517S.
32. Horn C, Sohn-Bösser L, Breed J, Welte W, Schmitt L, Bremer E. Molecular determinants for substrate specificity of the ligand-binding protein OpuAC from *Bacillus subtilis* for the compatible solutes glycine betaine and proline betaine. *J Mol Biol* 2006;357:592–606.
33. Gallivan JP, Dougherty DA. Cation- π interactions in structural biology. *Proc Natl Acad Sci USA* 1999;96:9459–9464.

A Role for Both Conformational Selection and Induced Fit in Ligand Binding by the LAO Protein

Daniel-Adriano Silva^{1,2}, Gregory R. Bowman³, Alejandro Sosa-Peinado², Xuhui Huang^{1*}

1 Department of Chemistry, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, **2** Department of Biochemistry, Medicine School, Universidad Nacional Autónoma de México, México D.F., México, **3** Department of Chemistry, Stanford University, Stanford, California, United States of America

Abstract

Molecular recognition is determined by the structure and dynamics of both a protein and its ligand, but it is difficult to directly assess the role of each of these players. In this study, we use Markov State Models (MSMs) built from atomistic simulations to elucidate the mechanism by which the Lysine-, Arginine-, Ornithine-binding (LAO) protein binds to its ligand. We show that our model can predict the bound state, binding free energy, and association rate with reasonable accuracy and then use the model to dissect the binding mechanism. In the past, this binding event has often been assumed to occur via an induced fit mechanism because the protein's binding site is completely closed in the bound state, making it impossible for the ligand to enter the binding site after the protein has adopted the closed conformation. More complex mechanisms have also been hypothesized, but these have remained controversial. Here, we are able to directly observe roles for both the conformational selection and induced fit mechanisms in LAO binding. First, the LAO protein tends to form a partially closed encounter complex via conformational selection (that is, the apo protein can sample this state), though the induced fit mechanism can also play a role here. Then, interactions with the ligand can induce a transition to the bound state. Based on these results, we propose that MSMs built from atomistic simulations may be a powerful way of dissecting ligand-binding mechanisms and may eventually facilitate a deeper understanding of allostery as well as the prediction of new protein-ligand interactions, an important step in drug discovery.

Citation: Silva D-A, Bowman GR, Sosa-Peinado A, Huang X (2011) A Role for Both Conformational Selection and Induced Fit in Ligand Binding by the LAO Protein. *PLoS Comput Biol* 7(5): e1002054. doi:10.1371/journal.pcbi.1002054

Editor: Ruth Nussinov, National Cancer Institute, United States of America and Tel Aviv University, Israel

Received: December 7, 2010; **Accepted:** March 31, 2011; **Published:** May 26, 2011

Copyright: © 2011 Silva et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by Hong Kong Research Grants Council DAG09/10.SC03, HKUST6/CRF/10, and University Grants Council RPC10SC03 to XH. This work was also supported by grant 53633 from CONACYT, México. GRB was funded by the Berry Foundation. Computing resources were provided by the Aitzalaoa supercomputing facility, Universidad Autónoma Metropolitana, Mexico, and CCB computer cluster in HKUST. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: xuhuihuang@ust.hk

Introduction

Molecular recognition plays important roles in many biological processes. For example, enzymes must recognize their substrates and drugs must be designed to have specific binding partners. Unfortunately, our understanding of how ligand binding occurs remains incomplete. In particular, the role that protein dynamics play in protein-ligand binding is unclear.

Two popular models for protein-ligand binding are the induced fit and conformational selection mechanisms. Both attempt to explain how a protein could transition from an unbound conformation to a bound conformation in complex with a ligand. In the induced fit model—introduced by Koshland [1]—the ligand first binds to the protein in its unbound conformation and this binding event induces the protein to transition to the bound state. Such models have been applied to many protein-protein and protein-DNA/RNA binding systems [2,3,4]. The conformational selection (or population shift) model [5,6,7,8,9,10,11,12] is a popular alternative to the induced fit mechanism. In this model, the intrinsic dynamics of the protein lead it to constantly transition between a stable unbound conformation and a less stable bound conformation. The ligand can then bind directly to the bound conformation, thereby stabilizing the bound state and increasing its population relative to the unbound state. The conformational

selection model has recently gained popularity in antibody or small ligand binding systems [10,11,12]. Some docking studies have also tried to exploit conformational selection by generating an ensemble of protein structures and docking small molecules against each of them in the hopes of identifying a transiently populated bound conformation that will be stabilized by the ligand [13].

Many recent studies have attempted to determine whether a variety of systems under different conditions can be best described by the induced fit or conformational selection model [14,15,16,17,18,19,20,21]. For example, Okazaki *et al.* [20] have found that strong and long range protein-ligand interactions favor the induced fit model, while weak and short range interactions favor the conformational selection model. Based on an analytic model, Zhou has suggested that the determining factor in ligand binding is the timescale for transitioning between the unbound and bound states with and without the ligand [18]. He found that conformational selection dominates when transitioning between the unbound and bound states is slow, while the induced fit mechanism dominates when this transition is fast. Many studies have proposed that conformational selection and induced fit are not mutually exclusive; instead, a blend of these two models may best describe most realistic systems [15,17,18,20,22,23]. For example, Zagrovic and coworkers [17] have suggested that

Author Summary

Protein-ligand interactions are crucial to chemistry, biology and medicine. Many studies have been conducted to probe the mechanism of protein-ligand binding, leading to the development of the induced fit and conformational selection models. Unfortunately, experimentally probing the atomistic details of protein-ligand binding mechanisms is challenging. Computer simulations have the potential to provide a detailed picture of molecular recognition events. In this study, we construct kinetic network models from atomistic simulations to elucidate the mechanism by which the LAO protein binds to its ligand. Because the LAO protein completely encompasses its substrate in the bound state, it has generally been assumed that it operates via an induced fit mechanism. We find that both the conformational selection and induced fit mechanisms play important roles in LAO binding. Furthermore, we have identified a number of parallel pathways for binding, all of which pass through a single gatekeeper state, which we refer to as the encounter complex state because the protein is partially closed and only weakly interacting with its substrate.

conformational selection and induced fit play equal roles in ubiquitin binding based on their analysis of NMR structures. However, in many cases, it is still difficult to dissect the chemical details of binding mechanisms. While it is clear that the bound and unbound states of a protein and their respective interactions with a ligand molecule are of great importance [15,18,20,21], it may also be important to take other conformational states into account. Protein dynamics are ultimately determined by their underlying free energy landscapes, whose ruggedness frequently gives rise to numerous metastable regions—sets of rapidly mixing conformations that tend to persist for extended periods of time.

In this work, we use Markov state models (MSMs) to map out the relevant conformational states in LAO binding and describe mechanistic details of this process. MSMs are a kinetic network model and a powerful approach to automatically identifying metastable states and calculating their equilibrium thermodynamics and kinetics [24,25,26,27]. MSMs focus on metastable regions of phase space, while there also exist other kinetic network models to study transition state [28]. MSMs partition conformational space into a number of metastable states; such that intra-state transitions are fast but inter-state transitions are slow. This separation of timescales ensures an MSM is Markovian (i.e. that the probability of transitioning from state i to state j depends only on the identity of i and not any previously visited state) and allows MSMs built from short simulations to model long timescale events. Many recent studies have demonstrated how MSMs can provide insight into drastic conformational changes like protein and RNA folding [26,29,30,31,32]. Here we demonstrate that MSMs built with a hierarchical clustering algorithm [30] can capture the mechanism by which the Lysine-, Arginine-, Ornithine-binding (LAO) protein, one of Periplasmic Binding Proteins (PBPs), binds to arginine. The LAO protein has a high binding affinity and undergoes large-scale domain rearrangements from an open to a closed state upon ligand binding [33,34,35,36] (see Fig. 1), making it a valuable model system for probing the coupling between protein conformational changes and binding.

Many have assumed that PBP binding occurs via the induced fit mechanism because the ligand is completely encapsulated by the protein in the bound state (see Fig. S1). Experimental studies of many PBPs support the induced fit mechanism, where the closure

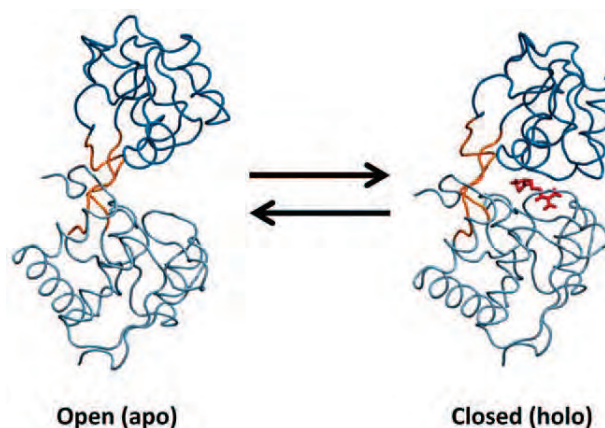


Figure 1. The open (PDB ID: 2LAO) and closed (PDB ID: 1LAF) states of the Lysine-, Arginine-, Ornithine-binding (LAO) Protein. The ligand, Arginine, is shown in red.
doi:10.1371/journal.pcbi.1002054.g001

of the domains is triggered by the binding of the ligand [33,34,36,37,38,39,40]. However, a few experimental studies indicate that some PBPs (including GGBP [41] and ChoX [42]) are able to reach the closed conformation in the absence of the ligand. This has been suggested as a sign of the conformational selection mechanism [38]. Furthermore, recent NMR studies with paramagnetic relaxation enhancement (PRE) of maltose-binding protein (MBP) identified a minor (~5%) un-liganded partially closed form. This partially closed state is in equilibrium with the open state and, therefore, is available for the binding of the ligand, which may further facilitate the transition to the bound state. This work suggests a more complex binding mechanism where both conformational selection and induced fit play significant roles [43], but since the ligand was not present during the experiments, it is unclear exactly what roles the two mechanisms may play.

With our MSM, we can directly monitor the mechanism of LAO binding and assess the role of both conformational selection and induced fit. Our model suggests that three dominant states need to be considered to adequately describe LAO binding and that both conformational selection and induced fit play important roles in the transitions between these states. The third dominant state in our model—besides the open and closed states—is only partially closed and weakly bound to the ligand; therefore, we refer to it as the encounter complex state. The ligand can induce the protein to transition from the open state to the encounter complex; however, the ligand-free protein can also transition to the encounter complex state, indicating an important role for conformational selection. In contrast, on our dataset the ligand-free protein never sampled the closed state, this suggests that the closed state in the absence of the ligand may represent a very high free energy state and that once the ligand reaches the binding site an induced fit mechanism is responsible for transitions from the encounter complex to the closed state.

Results

Model validation by *ab initio* prediction of the bound state and binding kinetics

Before drawing system-specific conclusions from a simulation study, it is important to first test the model against existing experimental data. MSMs built using the Super-level-set Hierarchical Clustering (SHC) algorithm [30] greatly facilitate this task by decomposing a system's conformational space into its

constituent metastable regions and describing the thermodynamics and kinetics of each. For instance, one can easily extract representative conformations from each state, determine the equilibrium probability of each state, or calculate the rates of transitioning between sets of states and compare to experimental results. In this study, we describe protein conformations by the opening and twisting angles between their two domains [44] and the location of the ligand because these degrees of freedom describe the slowest dynamics of the system (see Fig. S2). We then construct a 54-state MSM using SHC (See Methods for details of MSM construction). The dominant conformational states in our model are displayed in the Fig. S3.

Fig. 2 demonstrates that our model is capable of *ab initio* prediction of the bound state. As described in the Methods section, no knowledge of the bound state was included at any stage of our simulations or model construction. Based on the high binding affinity measured in experiments ($K_d \sim 14$ nM) [45] we postulated that the bound state should be the most populated state in our model. Indeed, representative conformations from our most populated state (having an equilibrium population of 74.9%) agree well with the crystal structure of the bound state, with an RMSD to the crystal structure of the binding site as little as 1.2 Å,

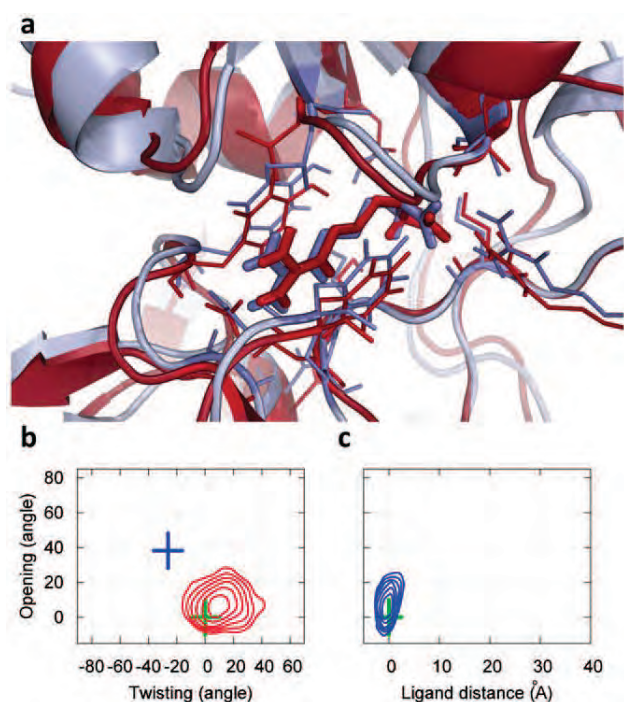


Figure 2. The bound state of our MSM for LAO binding (which is also the most populated state, having an equilibrium population of 74.9%). (a) A snapshot from our simulations (red) achieves a 1.2 Å RMSD to the X-ray bound state (blue, PDB ID: 1LAF). The RMSD is computed from the protein C α atoms that are within 8 Å to the center of mass of the ligand in the X-ray bound state (Residues 9–15, 17–19, 30, 50–53, 55–56, 67–74, 77, 83, 88, 90–92, 117–124, 141–143, 159–162, 164, 190–191 and 194–196). If all-protein C α atoms are included the RMSD is 1.8 Å. (b) Free energy plot of the protein opening angle versus twisting angle. The bin size is (5°, 5°), and the interval between two adjacent contour levels is 0.5 kT. The green and blue crosses correspond to X-ray structures of the bound and apo conformations respectively. (c) Free energy plot of the protein opening angle versus the distance between the ligand and the binding site. The bin size is (1.5 Å, 5°), and interval between contour levels is 0.5 kT. doi:10.1371/journal.pcbi.1002054.g002

as shown in Fig. 2A. Moreover, Figs. 2B and 2C show that the crystal structure of the bound state lies within the minimum of the most populated free energy basin. These figures also show that our model's bound state covers a relatively large region of phase space, suggesting that it is flexible, possibly to accommodate favorable interactions with all four of LAO's binding partners (L-lysine, L-arginine, L-ornithine and L-histidine). The structural properties of the remaining states are also consistent with experiments (see the Text S1 for more details). For example, many of the unbound states also contain partially closed protein conformations, consistent with NMR experiments on another PBP protein: MBP [43].

Our model is also in reasonable agreement with the experimentally measured binding free energy and association rates. For example, from the MFPT from all unbound states to the bound state, our model predicts an association timescale of 0.258 ± 0.045 μ s (see Methods for calculation details). Since rates are proportional to the exponential of the free energy barrier, an 8-fold difference in rates roughly corresponds to a 2 kT difference in the height of the free energy barrier. Therefore, our result is in reasonable agreement with the experimental value of ~ 2.0 μ s found in the highly homologous HisJ protein [46] (see Methods for similarity between LAO and HisJ protein). We also estimate a binding free energy of -8.46 kcal/mol using the algorithm introduced by van Gunsteren and co-workers [47], which is also in reasonable agreement with the experimental value of -9.95 kcal/mol for the LAO protein (see Methods for calculation details). Together, this agreement between theory and experiment suggests that our model is a sufficiently good reflection of reality to make hypotheses about details of the binding mechanism.

Arriving at these conclusions with a single long simulation would have been quite difficult due to the slow timescales involved. For example, transitioning from the bound state to an unbound state takes 2.15 ± 0.51 μ s on average. Therefore, observing enough transitions to gather statistics on the binding and unbinding rates in a single simulation would require that it be tens of μ s long. Such simulations are now possible [31,48] but are still challenging to perform. Moreover, scaling the long simulation approach to millisecond timescales is still infeasible. MSMs built from many μ s timescale simulations, however, have already proven capable of capturing events in a 10 millisecond timescale [49] and can likely scale to even slower processes.

Insights into the mechanism of LAO binding

In addition to predicting experimental parameters, MSMs are also useful for mapping out conformational transitions like protein-ligand binding. For example, Figs. 3 and 4 show the 10 highest flux pathways from any of the unbound states in our model to the bound state. All ten pathways pass through an obligatory, gatekeeper state (state 11) that we refer to as the encounter complex state because the protein is partially closed and only weakly interacting with the ligand (see Figs. 3, 4 and State 11 in Fig. S3). In the encounter complex (see Fig. 5) the two lobes of the LAO protein are structurally very similar to those in both the apo and bound X-ray structures (with RMSD less than 2 Å, see Table S1). Therefore the conformational change between crystal structures and the encounter complex could be achieved through a rigid body rotation. We also found that in the encounter complex the ligand was stacked between the lobe I Tyr14 and Phe52 and protrudes upward to interact with the lobe II Thr121. These contacts are also observed in the X-ray bound structure. (see Fig. S4). To further support our conclusion that the encounter complex state is an obligatory step in ligand binding, we have calculated that the average timescale for transitioning from the unbound states to the encounter complex state is 0.190 ± 0.037

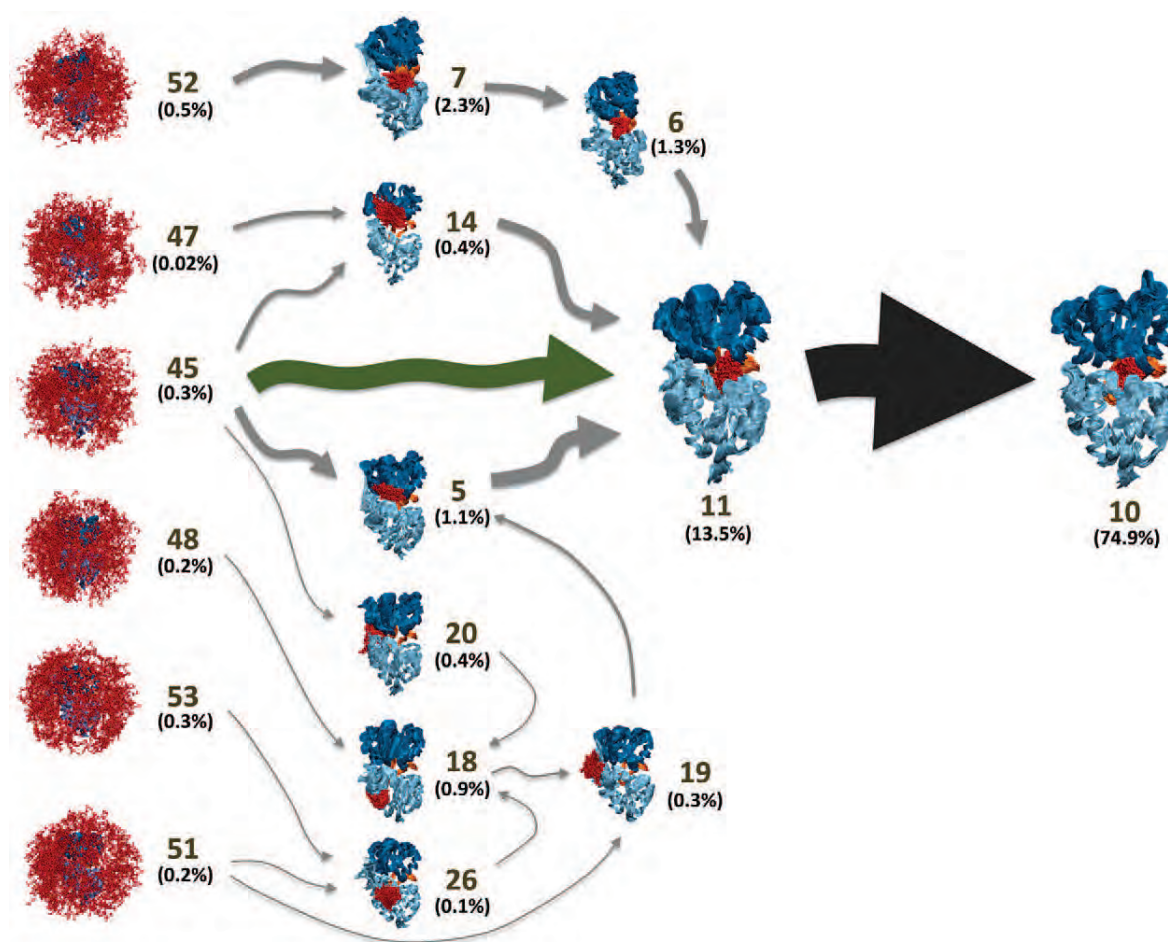


Figure 3. Superposition of the 10 highest flux pathways from the unbound states to the bound state. The flux was calculated using a greedy backtracking algorithm [31,66] applied to a 54-state MSM generated with the SHC algorithm [30]. These pathways account for 35% of the total flux from unbound states to the bound state. The arrow sizes are proportional to the interstate flux. State numbers and their equilibrium population calculated from MSM are also shown. The conformational selection and induced pathways from the unbound states to the encounter complex state is shown in green and grey arrows respectively.
doi:10.1371/journal.pcbi.1002054.g003

while the average timescale for transitioning from the unbound states to the bound state is $0.258 \pm 0.045 \mu\text{s}$. The average timescale for transitioning from the encounter complex state to the bound state is $0.090 \pm 0.015 \mu\text{s}$ (see Methods for calculation details). Thus, the unbound protein will typically transition to the encounter complex before reaching the bound state.

The top ten paths from the unbound states to the encounter complex can be divided into two sets, one that is best described by conformational selection and one that is better described by the induced fit mechanism. For example, the pathway from state 45 directly to 11 operates through conformational selection (see green arrow in Fig. 4): in the unbound state 45 the protein and ligand are not interacting but the protein conformations are very similar to those in the encounter complex. Since the protein adopts similar conformations in these two states, the ligand can always bind to a pre-existing encounter-complex-like (state 11 like) protein conformation (the conformational selection mechanism). The binding kinetics of this conformational selection pathway is quite rapid, having a mean first passage time for transitioning from the unbound state 45 to the encounter complex state 11 of $0.220 \pm 0.054 \mu\text{s}$, and this pathway accounts for $\sim 45\%$ of the flux of the top ten pathways from unbound states to the bound state.

The second group of pathways to the encounter complex, which together account for $\sim 55\%$ of the flux may be better described by the induced fit mechanism. In general, these pathways start off in conformations that are much more open or twisted than the encounter complex. Next, the system transitions to one or more intermediate states where the ligand is interacting with the protein at (or near) its binding site, though the protein is still quite open or twisted. Finally, the protein-ligand interactions induce a transition to the encounter complex state. For example, the pathway starting from state 47, passing through state 14, and ending at state 11 falls into this category (see Fig. 4).

Transitions from the encounter complex to the bound state are best described by the induced fit mechanism. When the system enters the encounter complex state, the protein is generally in a relatively open conformation (opening angle within 20° to 70° , see Fig. 6). However, when the system leaves the encounter complex state to enter the bound state, the protein is mostly in a more closed conformation (opening angle smaller than 30° , see Fig. 6). Thus, it appears interactions with the ligand induce the protein to close. Furthermore, our model predicts that the encounter complex-to-bound transition ($0.090 \pm 0.015 \mu\text{s}$) is much faster than the encounter complex-to-unbound transition ($1.927 \pm 0.499 \mu\text{s}$), so

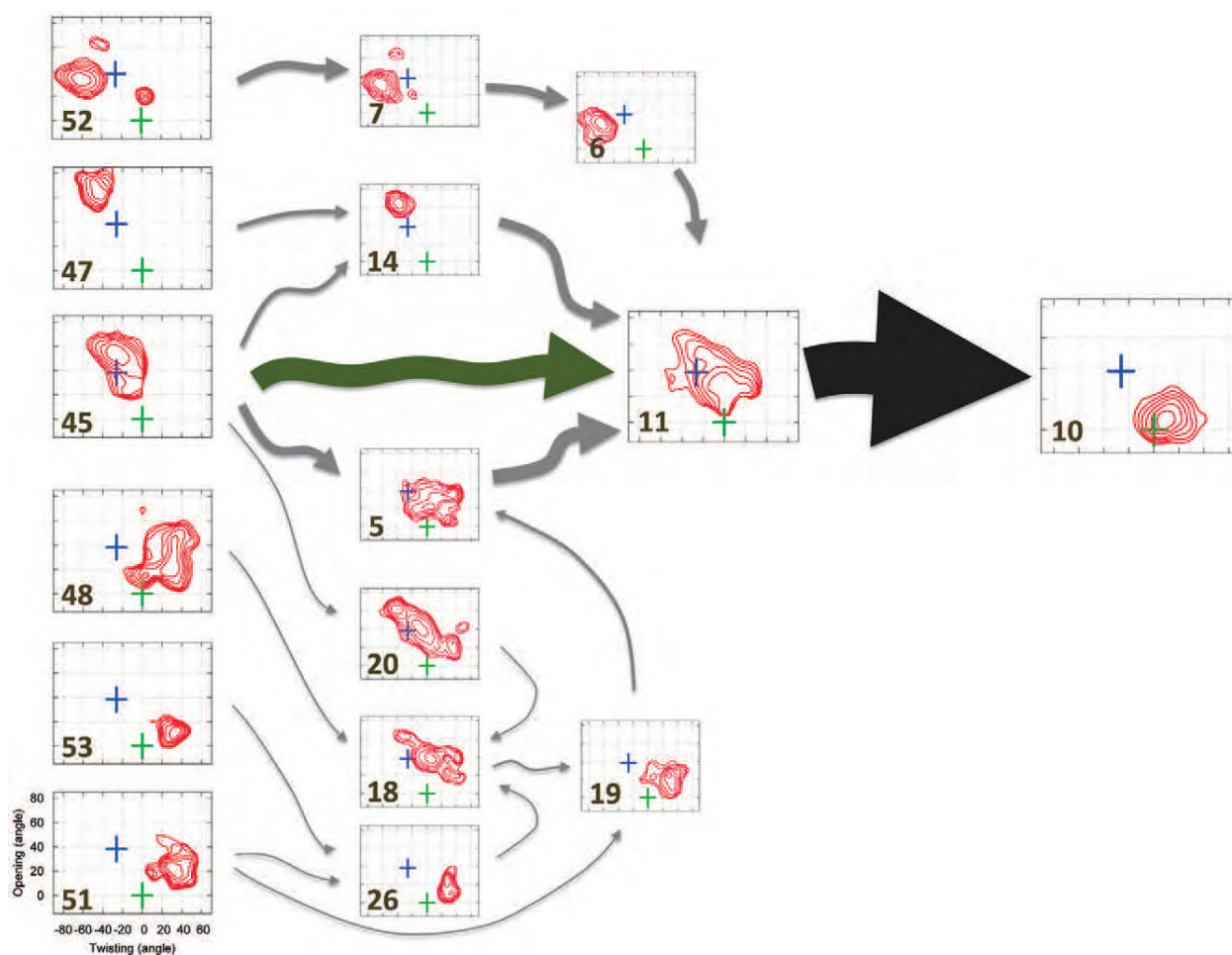


Figure 4. Superposition of the 10 highest flux pathways from the unbound states to the bound state as in Fig. 2 but with representative structures replaced with free energy plots of the protein opening angle versus twisting angle. The green and blue crosses correspond to X-ray structures of the bound and apo conformations respectively. The bin size is (5° , 5°), and the interval between contour levels is 0.5 KT (same as Fig. 2b). The conformational selection and induced pathways from the unbound states to the encounter complex state is shown in green and grey arrows respectively. doi:10.1371/journal.pcbi.1002054.g004

the encounter complex is not likely to diffuse back to the unbound state instead of converting into the bound state. In addition, the protein never samples fully-closed conformations in the absence of the ligand in our simulations. Together, these observations indicate that the induced fit mechanism should dominate transitions from the encounter complex to the bound state.

Mis-bound states

We have also identified a number of metastable mis-bound states, like states 4 and 20 in Fig. S3. In these states, the ligand interacts with the protein outside of the binding site. For example, in state 20 (population $\sim 0.4\%$) the ligand is bound to the hinge region between the two domains of the protein. Transitioning from a mis-bound state to the bound state generally requires passing through an unbound state (see Fig. S5). Therefore, these states are mostly off pathway and likely slow down the overall binding kinetics.

Discussion

The LAO protein is a member of the PBP family, which is responsible for transporting low molecular weight ligands from the

outer to the inner membrane in the ABC transport mechanism of Gram-negative bacteria [34,50]. Crystal structures for this system have shown that the binding site is completely closed-off in the bound state [33,37], making it impossible for the ligand to enter the binding site after the protein has adopted the closed conformation (see Fig. S1). Therefore, it has often been proposed that LAO binding occurs through an induced fit mechanism [51,52,53].

Our observations of LAO binding indicate that, like protein folding, ligand binding is a multi-state process with parallel pathways. All top 10 pathways pass through a gate-keeper state that we refer to as the encounter complex state because the protein is partially closed and only weakly interacting with the ligand. The system can reach this state through either the induced fit mechanism or conformational selection. Rather than being a transient state, this encounter complex is quite metastable. Once in the encounter complex state, the ligand is able to quickly induce protein conformational changes that lead to a transition to the fully closed, bound state.

Other systems may operate through a similar mechanism such as other PBPs. Indeed, our model is consistent with Tang

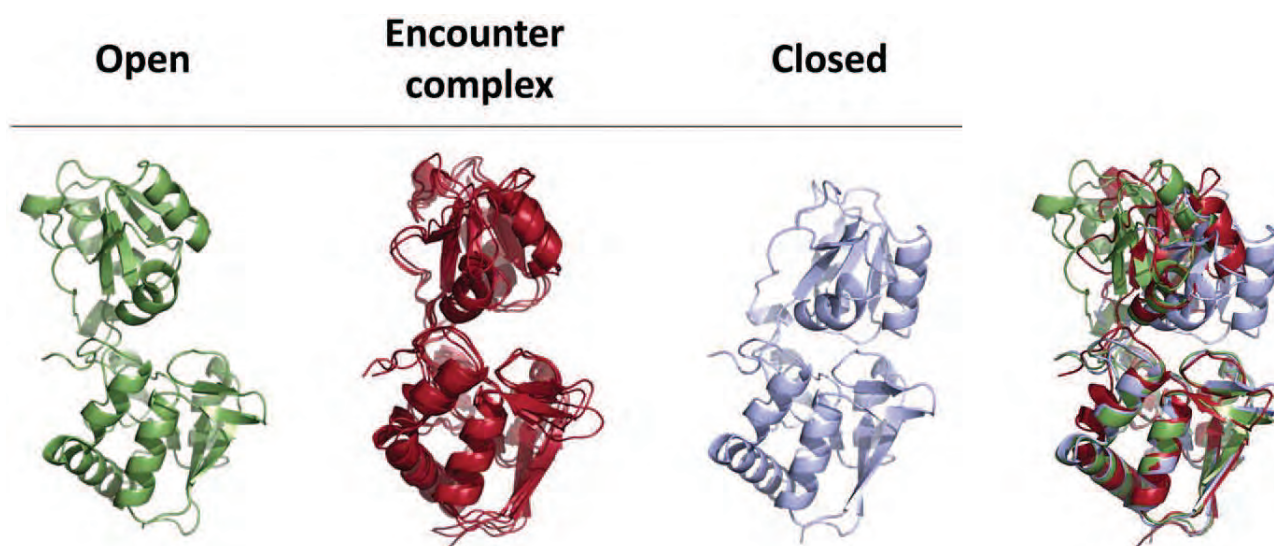


Figure 5. Structural comparisons between encounter complex and X-ray apo (or open) and bound (or closed) structures for the LAO protein. The X-ray apo (PDB ID: 2LAO) and bound structure (PDB ID: 1LAF) are shown in green and light blue respectively. Three representative conformations from the encounter complex state are superimposed and shown in red. These three conformations are representative of 10,000 randomly selected conformations from the encounter complex state (i.e. they have the smallest protein C α RMSD to all the rest of the randomly selected conformations and are, therefore, the most central/typical of the state). In the right panel, the open, closed X-ray structures are overlaid with one of the representative conformations from the encounter complex state.
doi:10.1371/journal.pcbi.1002054.g005

et. al.[43]'s findings for MBP. Specifically, they discovered a partially closed state in equilibrium with the open state for the apo protein. Thus, this state is available for the binding of the ligand to form the encounter complex through the conformational selection mechanism. Next, the binding could facilitate the transition to the bound state via the induced fit mechanism. Their model was proposed mainly based on NMR experiments in the absence of the ligand, but our simulations directly observed this interplay at

atomic resolution. Furthermore, our results suggest that transitions from the open to the partially closed state occur via a combination of conformational selection and induced fit mechanisms. For several PBPs including MBP and HisJ, the rate constant measured by the stopped flow experiments is proportional to the ligand concentration, indicating a simple two-state binding mechanism [46]. However, as discussed by Tang *et. al.* [43], these stopped flow measurements may not be able to capture the intermediate encounter complex state because the overall binding timescale is extremely rapid, e.g. a few hundred nanoseconds for LAO.

More broadly, there may also exist other proteins with closed active sites that have metastable encounter complex states. Sullivan *et. al.* [16] suggested that in such encounter complex state substrate-enzyme interactions are almost identical to the active state, while the enzyme has not yet reached the active form. Furthermore, the enzyme must operate by an induced fit mechanism to reach the active form because of the closure of the enzyme would prevent the substrate from entering the active site. This model is consistent with our findings for the LAO protein. However, in order to reach this encounter complex state, we suggest that both induced fit and conformational selection may play important roles. In general, other proteins with closed active sites may also make use of both conformational selection and induced fit to reach the encounter complex. However, the relative contributions of these mechanisms may vary depending on factors like the relative strength of the protein-ligand interactions [20].

The ability to map out the details of LAO binding using MSMs is an important step towards a deeper understanding of protein-ligand interactions for this system. Future application of these methodologies to other systems could even lead to the identification of general principles of protein-ligand interactions and allostery. This knowledge may also greatly aid in computational drug design. For example, it may not always be possible to identify all the relevant states via other structural methods, like crystallography. Using MSMs, however, one can hope to identify the most important relevant states and design small-molecules to

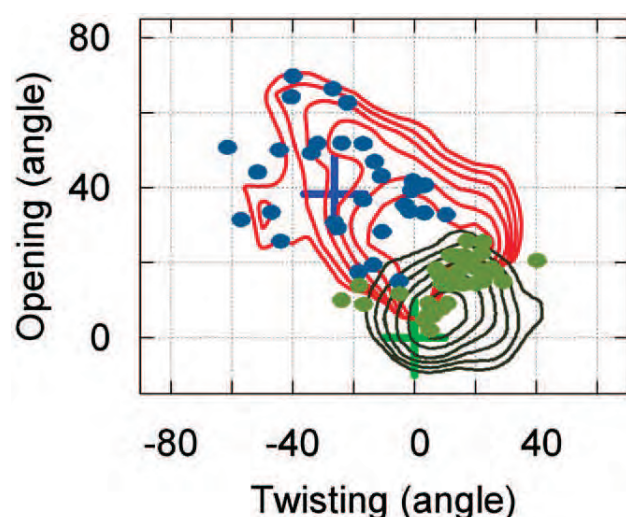


Figure 6. Overlay of free energy plots of the protein opening angle versus twisting angle for the encounter complex (red) and bound state (blue). Green dots correspond to where transitions from the encounter complex to the bound state occur. Blue dots correspond to where transitions into the encounter complex from other states occur. Only transitions without re-crossing are counted (minimum residence time in the final state after the transition is 6 ns).
doi:10.1371/journal.pcbi.1002054.g006

specifically stabilize one or more states over the others. Future work with improved force fields and greater sampling could also greatly enhance our understanding of protein-ligand interactions. However, we stress that the present work lays out the methodology that would be employed in such future research.

Conclusions

In this study, we demonstrate the power of MSMs for understanding protein-ligand interactions using the LAO protein as a model system. Our results indicate that LAO binding is a two-step process involving many states and parallel pathways. In the first step, the ligand binds to a partially closed protein to form an encounter complex. Both the conformational selection and induced fit mechanisms play significant roles in this step. In the second step, the system transits from the encounter complex state to the bound state via the induced fit mechanism. This two-step binding mechanism (see Fig. 7 for schematic diagram of the binding mechanism) may also be used by other systems, such as other PBP proteins, enzymes with closed active sites, and systems where the apo protein dynamics rarely visits the bound conformation. Future applications of MSMs with improved force fields, greater sampling, and to other protein-ligand interactions will reveal how general this mechanism is, aid in understanding allostery, and lay a foundation for improved drug design.

Methods

Simulation dataset

We have performed 65 molecular dynamics simulations, each 200 ns long, of the LAO protein from the organism *Salmonella typhimurium* and one of its ligands, L-arginine. Ten simulations were started from the open protein conformation (PDB ID: 2LAO [37]) with the ligand at more than 25 Å away from the binding site. The other simulations were initialized from conformations

randomly selected from the first ten simulations. We saved conformations every 20ps with a total of more than 650,000 conformations. Among these MD simulations, we have observed multiple binding events and unbinding events (see Fig. S6 and Text S2). The protein was solvated in a water box with 11,500 SPC waters [54] and 1 Na⁺ ion. All the simulations were performed using the GROMACS 4.0.5 simulation package [55] with the GROMOS96 force field [56]. The simulation system was minimized using a steepest descent algorithm, followed by a 250 ps MD simulation applying a position restraint potential to the protein heavy atoms. The simulations were performed under isothermal-isobaric conditions (NPT) with P = 1 bar and T = 318 K, using Berendsen thermostat and Berendsen barostat with coupling constants of 0.1 ps⁻¹ and 1 ps⁻¹ respectively [57]. A cutoff of 10 Å was used for both VDW and short-range electrostatic interactions. Long-range electrostatic interactions were treated with the Particle-Mesh Ewald (PME) method [58]. Nonbonded pair-lists were updated every 10 steps. Waters were constrained using the SETTLE algorithm [59] and all protein bonds were constrained using the LINCS algorithm [60]. Hydrogen atoms were treated as virtual interaction sites, enabling us to use an integration step size of 5 fs [61].

Markov State Model (MSM) state decomposition

We used MSMBuilder [27] and SHC [30] to construct the state decomposition for our MSM for LAO binding.

We first used the k-centers algorithm in MSMBuilder [27] to cluster our data into a large number of microstates. The objective of this clustering was to group together conformations that are so geometrically similar that one can reasonably assume (and later verify) that they are also kinetically similar.

Because we had to account for both the protein and ligand, we performed two independent clusterings; one based on the opening and twisting angle of the protein and one based on the relative position of the ligand (see Fig. S2). We then combined the two clusterings by treating them as independent sets. For example, M protein-based clusters and N ligand-based clusters would lead to a total of M×N clusters.

For the protein-based clustering, we created 50 clusters using the Euclidean distance to a vector containing the protein opening and twisting angles. The opening angle (see Fig. S2a) was defined as the angle between the normal vectors of the two planes defined by the center of masses of the following groups of C_α atoms:

- Plane-A: Residues 6–88 & 195–227; 162–168; and 121–127;
- Plane B: Residues 92–185; 162–168; and 121–127.

The twisting angle (see Fig. S2b) is the angle between the following two planes:

- Plane-A: Residues 6–88 & 195–227; 83–88 & 194–199; and 92–97 & 156–161;
- Plane-B: Residues 92–185; 83–88 & 194–199; and 92–97 & 156–161.

The strong correlation between opening and twisting angles of the protein and the two slowest eigenvectors from Principle Component Analysis (PCA) analysis of a 20 ns MD simulation started from the apo structure in the absence of the ligand demonstrates that they are a reasonable descriptor of the protein's conformation (see Figs. S2c and S2d). As a reference point, we note that the holo X-ray structure (PDB ID: 1LAF) [33] has both opening and twisting angles equal zero, while the apo X-ray

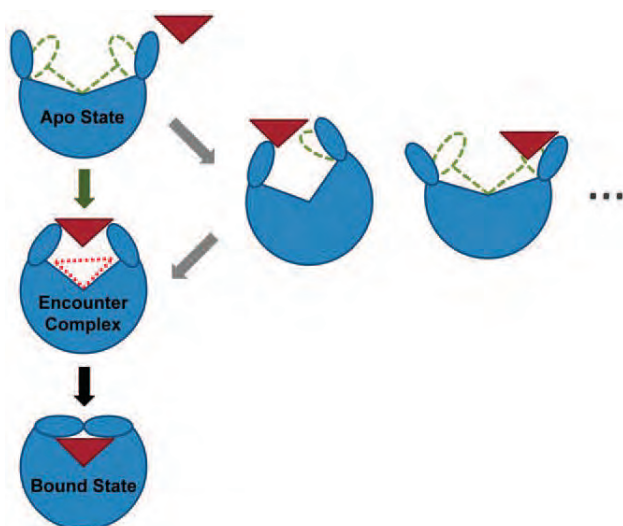


Figure 7. A schematic diagram describing the proposed two-step binding mechanism for proteins in steric occlusion of the direct binding of the ligands. The first step is the transition from the apo to the encounter complex state. In this step, multiple pathways exist where both conformational selection and the induced fit mechanisms play important roles. The second step is the transition from the encounter complex to the bound state, where the induced fit mechanism is adopted.

doi:10.1371/journal.pcbi.1002054.g007

structure (PDB ID: 2LAO) [37] has opening angle = 38.2° and twisting angle = -26.2°.

For the ligand-based clustering, we created 5000 clusters using the Euclidean distance between all heavy-atoms.

We then had to modify our clustering to account for the fact that the ligand dynamics fall into two different regimes (see Fig. S7): one where the ligand moves slowly due to interactions with the protein and one where the ligand is freely diffusing in solution. The existing clusters are adequate for describing the first regime. However, when the ligand is freely diffusing (more than 5 Å from the protein), the procedure outlined above results in a large number of clusters with poor statistics (less than ten transitions to other states). Better sampling of these states would be a waste of computational resources as there are analytical theories for diffusing molecules and a detailed MSM would provide little new insight. Instead, we chose to re-cluster these states using the same protein coordinates and the Euclidean distance between the ligand's center of mass (as opposed to the Euclidean distance between all ligand heavy-atoms). For this stage, we created 10 new protein clusters and 100 new ligand clusters.

After dropping empty clusters, this procedure yielded 3,730 microstates. Of these, 3,290 microstates came from the initial high resolution clustering and 440 came from the data that was reclustered at low resolution. To verify that the final microstate model is valid (Markovian) we plotted the implied timescales and found that they level off at a lag time between 2 and 6 ns (see Text S3 and Fig. S8a), implying that the model is Markovian for lag times in this range. Therefore, we can conclude that the microstates are sufficiently small to guarantee that conformations in the same state are kinetically similar.

We then lumped kinetically related microstates into macrostates using the SHC algorithm [30]. This is a powerful lumping method that efficiently generates more humanly comprehensible macrostate models (i.e. ones with fewer small macrostates arising from statistical error) than the PCCA algorithm currently implemented in MSMBuilder.

In SHC, one performs spectral clustering hierarchically using super level sets (or density levels) starting from the highest density level, thereby guaranteeing that highly populated meta-stable regions are identified before less populated ones. For SHC, we selected density levels $L_{\text{high}} = [0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95, 0.99]$ and $L_{\text{low}} = [0.4, 0.95]$, for the high and low-density regions respectively. The low and high resolution states were lumped separately because the states in each set have different sizes, so it is difficult to compare their densities. We then combined these two sets of macrostates to construct an MSM with 54 macrostates. Once again, we used the implied timescales test to verify that the model is Markovian and found that a 6 ns lag time yields Markovian behavior (see Fig. S8b).

Calculating transition matrices

To calculate a transition matrix using the above state decomposition we first counted the number of transitions between each pair of states at some observation interval (the lag time) to generate a transition count matrix, where the entry in row x and column y gives the number of transitions observed from state x to state y . In particular, we use a sliding window of the lag time on each 200 ns trajectory with a 20 ps interval between stored conformations (i.e. each trajectory contains 10,000 conformations) to count the transitions. Because we use a hard cutoff between states, simulations at the tops of barriers between states can quickly oscillate from one state to the other, leading to an over-estimate of the transition rate between states [62]. To mitigate the effect of these recrossing events, we only counted transitions from state x to

state y if the protein remained in state y for at least 300 ps before transitioning to a new state. To generate the transition probability matrix (where the entry in row x and column y gives the probability of transitioning from state x to state y in one lag time), we normalized each row of the transition count matrix.

Mean First Passage Time (MFPT) calculation

We followed the procedure in Ref [63] to compute the mean first passage time (**MFPT**) from initial state i to final state f , i.e. the average time taken to get from state i to state f . In particular, the **MFPT** (X_{ij}) given that a transition from state i to j was made first is the time it took to get from state i to j plus the **MFPT** from state j to f . **MFPT** (X_{ij}) can be defined as,

$$X_{ij} = \sum_j P_{ij}(t_{ij} + X_{jf}) \quad (1)$$

where $t_{ij} = 6$ ns is the lag time of the transition matrix \mathbf{T} . The boundary condition is:

$$X_{ff} = 0 \quad (2)$$

A set of linear equations defined by Equation (1) and (2) can be solved to obtain the **MFPT** X_{ij} . We used bootstrapping to put error bars on MFPTs. That is, one-hundred new data sets were created by randomly choosing trajectories 130 times with replacement. We then calculated the MFPTs for each data set and reported their means and standard deviations. In MFPT calculations, the encounter complex was considered to contain state 11 and state 5, because state 5 also has features of the encounter-complex, though it plays a significantly smaller role (refer to the Text S1 for details).

Calculation of the binding free energy and association rate

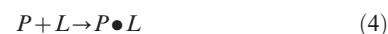
To compute the binding free energy ΔG from our simulations, we use the method introduced by van Gunsteren and co-workers [47]:

$$\Delta G = -k_b T \ln K_b = -k_b T \ln \left(\frac{\alpha_{\text{bound}}}{c_0 \alpha_{\text{free}}^2} \right) \quad (3)$$

where α_{bound} and α_{free} are the fractions of bound and free species respectively. c_0 is the overall concentration of ligand. In our simulations, $c_0 = 1/(N_A V_{\text{box}})$, where N_A is Avogadro's number, and V_{box} is the volume of the simulation box. In our system, $c_0 = 0.0049$ mol/L, $T = 318$ K. We consider all the unbound states as free species, so that $\alpha_{\text{free}} = 1.75\%$. Thus, the bound species $\alpha_{\text{bound}} = 1 - \alpha_{\text{free}} = 98.25\%$, which contains all the states where the protein and ligand are in close contact. From Eq. (3), we have:

$$\Delta G = -8.46 \text{Kcal/mol}$$

We can also derive the association rate. Given a system in equilibrium described by the protein \mathbf{P} , ligand \mathbf{L} and the protein-ligand complex $\mathbf{P}\bullet\mathbf{L}$, the protein-ligand binding reaction can be written as:



Rate equation can be written as

$$\frac{d[P]}{dt} = -k_{on}[P][L] + k_{off}[P\bullet L] \quad (5)$$

Where k_{on} and k_{off} are forward and backward rate constants respectively. Since the forward reaction (i.e. association) only depends on k_{on} , the rate equation for the forward reaction can be written as:

$$\frac{d[P]}{dt} = -k_{on}[P][L] \quad (6)$$

The total concentrations of protein and ligand in the system are constant

$$[P_{tot}] \stackrel{def}{=} [P] + [P\bullet L] \quad (7)$$

$$[L_{tot}] \stackrel{def}{=} [L] + [P\bullet L] \quad (8)$$

Thus only one concentration among $[P]$, $[L]$, and $[P\bullet L]$ is independent. If we choose $[P]$ as the independent concentration and then rate equation for forward reaction can be rewritten as:

$$\frac{d[P]}{dt} = -k_{on}[P][L] = -k_{on}([L_{tot}] - [P_{tot}] + [P])[P] \quad (9)$$

If we consider the condition $[P_{tot}] = [L_{tot}]$, which is the case for our simulations:

$$\frac{d[P]}{dt} = -k_{on}[P]^2 \quad (10)$$

We can solve Eq (10) with the initial condition $[P] = [P_0]$ at $t = 0$:

$$[P] = \frac{[P_0]}{k_{on}[P_0]t + 1} \quad (11)$$

We define the association timescale ($t_{1/2}$) as the time when half of the protein ($[P] = [P_0]/2$) has associated with the ligand:

$$t_{1/2} = \frac{1}{k_{on}[P_0]} \quad (12)$$

Since there is no experimental k_{on} rate constant available for the LAO protein, we choose for comparison the k_{on} from the Histidine binding (HisJ) protein. The LAO and HisJ proteins have considerable similarity both in structure and function. For example, both proteins are the same size (238 a.a.) and have a 70% sequence identity. In fact, if conservative mutations are taken into account the sequence identity increases to 83% [64]. These homologous proteins also bind to the same membrane receptor (HisQ/HisM/HisP) [37] and the same ligands (cationic L-amino acids) [65]. The RMSD between the X-ray crystal structures of holo LAO and HisJ bound to histidine (1LAG and 1HBP) is also quite small (C α RMSD as low as 0.62 Å). The binding affinities of these proteins to their ligands are also similar (all about a nanomolar, though the binding affinities are not exactly the same [Histidine binds to HisJ most strongly, but binds to LAO most

loosely]). The similarity between LAO and HisJ has also been discussed in detail in a previous study by Oh *et al.* [65]. Therefore, we think it is a reasonable assumption that the LAO and HisJ proteins have similar binding kinetics.

For the Histidine binding protein, $k_{on} = 1 \times 10^8 M^{-1} sec^{-1}$. In our simulation, the initial concentration of protein is $[P_0] = 0.0049 mol/L$, thus, the association timescale:

$$t_{1/2} = 2\mu s$$

The experimental association timescale is about eight times slower than that computed from our simulations. However, we note that the only available experimental k_{on} was measured at 293K [46], while our simulations were performed at a higher temperature (318K) with faster kinetics. Thus, the difference between the experimental and simulation rates will be smaller at the same temperature. For the binding free energy, the experimental measurement was at an even lower temperature (277K) [45]. Thus, the experimental binding free energy at the temperature we simulated should be closer to our calculated value.

Supporting Information

Figure S1 The X-ray bound structure of LAO (PDB ID: 1LAF) from four different viewing angles shows that the protein completely encloses the ligand. Thus, unbound ligands cannot enter the binding site when the protein is closed. (TIF)

Figure S2 (a) Opening and (b) twisting angles used to describe the motion of the protein. (c) The projection of conformations on the second eigenvector from Principle Component Analysis (PCA), and protein opening dihedral angle (see SI Sec 2 for detailed definition) as a function of time are shown in red and black respectively. The 20 ns simulation is started from protein in the closed state (PDB ID: 1LAF), but ligand was not included in the simulation. (d) Same as (c) except that the projection of conformations on the first eigenvector from PCA and protein twisting dihedral angle are plotted. In this system, the twisting and opening angles are correlated well with the first and second eigenvectors from PCA. (TIF)

Figure S3 Populations of top 23 most populated macrostates (macrostate-population > 0.1%) of the 54-state MSM. 1,000 conformations randomly selected from each macrostate are overlaid and shown from four different viewing angles. Free energy plots of the protein opening angle versus twisting angle (O, T) and the distance between the ligand and the binding site versus the opening angle (L, O) are displayed in red and blue respectively. The bin size is 5° for “O” and “T” and 1.5 Å for the ligand distance. The interval between two adjacent contour levels is 0.5 KT. The green and blue crosses correspond to the X-ray bound and apo structures respectively. For the free energy plots the x and y axis-scales are the same than those found in the Fig. 2 of the main text. (TIF)

Figure S4 The three representative conformations from the encounter complex state based on RMSD of the (a) protein C- α atoms and (b) binding site plus the ligand (heavy atoms of the residues within a cutoff of 8 Å from the ligand in the crystal structure with PDB ID: 1LAF) are compared to (c) the crystal bound structure of LAO (PDB ID: 1LAF). The protein is shown in

cartoon representation while the residues within a cutoff of 8 Å from the ligand in the crystal structure of 1LAF are shown in lines. In these conformations, we found the ligand side chain interacts with the lobe I Tyr14 and Phe52 through cation- π interactions. From five out of the six conformations, we also observed that the NH_3COO^- group of the ligand protrudes upward to interact with the lobe II residue Thr121. The same interactions are also observed in the X-ray bound structure.

(TIF)

Figure S5 Superposition of the 5 highest flux pathways from mis-bound states to the bound state. The flux was calculated using a greedy backtracking algorithm from a 53-state Markov State Model (MSM) generated with the SHC algorithm. The arrow size is proportional to the interstate flux. (a) Representative structures and equilibrium populations are shown for each state. (b) Free energy plots of protein opening versus twisting angle are shown. The macrostate number is inserted in each of the free energy plots.

(TIF)

Figure S6 The ligand distance to the binding site vs. simulation time was plotted for 12 MD simulations where binding (11 simulations) or unbinding (1 simulation, gray shaded) events were observed.

(TIF)

Figure S7 Ligand rotational autocorrelation functions for the unbound states (blue), the encounter complex state (green), and the bound state (red) are shown in the left panel. Autocorrelation functions generated from different trajectories are overlaid in the same figure. On the right panel, a schematic figure illustrates that the ligand rotates quickly when it is far away from the protein but the ligand rotation is restrained when it interacts with the protein. Thus, when constructing MSMs, we only consider the ligand center of mass motion when the ligand does not have strong interactions with the protein but we consider motion of all the

ligand heavy atoms when the ligand is strongly interacting with the protein.

(TIF)

Figure S8 Twenty slowest implied timescales as a function of lag time computed from (a) MSM containing 3730 microstates and (b) MSM containing 54 macrostates. Both plots level off at a lag time of ~ 4 ns. Thus we choose 6 ns to construct final MSMs.

(TIF)

Table S1 Averaged RMSD of the LAO protein Lobes I and II between three representative encounter complex conformations and the apo and bound X-ray structures. $C\alpha$ atoms of the Lobe I residues 6-88 & 195-227 or Lobe II residues 92-185 were included in the RMSD calculations. The structural alignment and RMSD calculation were performed separately for each Lobe.

(PDF)

Text S1 Structural features of the macrostates.

(PDF)

Text S2 Binding/unbinding transitions observed in the MD dataset.

(PDF)

Text S3 Implied timescales calculation.

(PDF)

Acknowledgments

We would like to thank Prof. Yuan Yao and Dr. Arturo Rojo Domínguez for helpful discussions.

Author Contributions

Conceived and designed the experiments: DAS ASP XH. Performed the experiments: DAS. Analyzed the data: DAS GRB XH. Contributed reagents/materials/analysis tools: ASP XH. Wrote the paper: DAS GRB XH.

References

- Koshland DE (1958) Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc Natl Acad Sci U S A* 44: 98–104.
- Bui JM, McCammon JA (2006) Protein complex formation by acetylcholinesterase and the neurotoxin fasciculin-2 appears to involve an induced-fit mechanism. *Proc Natl Acad Sci U S A* 103: 15451–15456.
- Levy Y, Onuchic JN, Wolynes PG (2007) Fly-casting in protein-DNA binding: frustration between protein folding and electrostatics facilitates target recognition. *J Am Chem Soc* 129: 738–739.
- Williamson JR (2000) Induced fit in RNA-protein recognition. *Nat Struct Biol* 7: 834–837.
- Kumar S, Ma B, Tsai CJ, Sinha N, Nussinov R (2000) Folding and binding cascades: dynamic landscapes and population shifts. *Protein Sci* 9: 10–19.
- Ma B, Kumar S, Tsai CJ, Nussinov R (1999) Folding funnels and binding mechanisms. *Protein Eng* 12: 713–720.
- Ma B, Shatsky M, Wolfson HJ, Nussinov R (2002) Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations. *Protein Sci* 11: 184–197.
- Tsai CJ, Kumar S, Ma B, Nussinov R (1999) Folding funnels, binding funnels, and protein function. *Protein Sci* 8: 1181–1190.
- Tsai CJ, Ma B, Nussinov R (1999) Folding and binding cascades: shifts in energy landscapes. *Proc Natl Acad Sci U S A* 96: 9970–9972.
- James LC, Roversi P, Tawfik DS (2003) Antibody multispecificity mediated by conformational diversity. *Science* 299: 1362–1367.
- Arora K, Brooks CL, 3rd (2007) Large-scale allosteric conformational transitions of adenylate kinase appear to involve a population-shift mechanism. *Proc Natl Acad Sci U S A* 104: 18496–18501.
- Bahar I, Chennubhotla C, Tobi D (2007) Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation. *Curr Opin Struct Biol* 17: 633–640.
- Wong S, Jacobson MP (2008) Conformational selection in silico: loop latching motions and ligand binding in enzymes. *Proteins* 71: 153–164.
- Csermely P, Palotai R, Nussinov R (2010) Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem Sci* 35: 539–546.
- Hammes GG, Chang YC, Oas TG (2009) Conformational selection or induced fit: a flux description of reaction mechanism. *Proc Natl Acad Sci U S A* 106: 13737–13741.
- Sullivan SM, Holyoak T (2008) Enzymes with lid-gated active sites must operate by an induced fit mechanism instead of conformational selection. *Proc Natl Acad Sci U S A* 105: 13829–13834.
- Wlodarski T, Zagrovic B (2009) Conformational selection and induced fit mechanism underlie specificity in noncovalent interactions with ubiquitin. *Proc Natl Acad Sci U S A* 106: 19346–19351.
- Zhou HX (2010) From induced fit to conformational selection: a continuum of binding mechanism controlled by the timescale of conformational transitions. *Biophys J* 98: L15–17.
- Swift RV, McCammon JA (2009) Substrate induced population shifts and stochastic gating in the PBCV-1 mRNA capping enzyme. *J Am Chem Soc* 131: 5126–5133.
- Okazaki K, Takada S (2008) Dynamic energy landscape view of coupled binding and protein conformational change: induced-fit versus population-shift mechanisms. *Proc Natl Acad Sci U S A* 105: 11182–11187.
- Weikl TR, von Deuster C (2009) Selected-fit versus induced-fit protein binding: kinetic differences and mutational analysis. *Proteins* 75: 104–110.
- Formanek MS, Ma L, Cui Q (2006) Reconciling the “old” and “new” views of protein allostery: a molecular simulation study of chemotaxis Y protein (CheY). *Proteins* 63: 846–867.
- Bakan A, Bahar I (2009) The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding. *Proc Natl Acad Sci U S A* 106: 14349–14354.
- Noe F, Fischer S (2008) Transition networks for modeling the kinetics of conformational change in macromolecules. *Curr Opin Struct Biol* 18: 154–162.
- Chodera JD, Singhal N, Pande VS, Dill KA, Swope WC (2007) Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *J Chem Phys* 126: 155101.
- Buchete NV, Hummer G (2008) Coarse master equations for peptide folding dynamics. *J Phys Chem B* 112: 6057–6069.

27. Bowman GR, Huang X, Pande VS (2009) Using generalized ensemble simulations and Markov state models to identify conformational states. *Methods* 49: 197–201.
28. Muff S, Callisch A (2009) Identification of the protein folding transition state from molecular dynamics trajectories. *J Chem Phys* 130: 125104.
29. Noe F, Schutte C, Vanden-Eijnden E, Reich L, Weikl TR (2009) Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proc Natl Acad Sci U S A* 106: 19011–19016.
30. Huang X, Yao Y, Bowman GR, Sun J, Guibas LJ, et al. (2010) Constructing multi-resolution markov state models (msms) to elucidate RNA hairpin folding mechanisms. *Pac Symp Biocomput*. pp 228–239.
31. Voelz VA, Bowman GR, Beauchamp K, Pande VS (2010) Molecular simulation of ab initio protein folding for a millisecond folder NTL9(1–39). *J Am Chem Soc* 132: 1526–1528.
32. Morcos F, Chatterjee S, McClendon CL, Brenner PR, Lopez-Rendon R, et al. (2010) Modeling conformational ensembles of slow functional motions in Pin1-WW. *PLoS Comput Biol* 6: e1001015.
33. Oh BH, Ames GF, Kim SH (1994) Structural basis for multiple ligand specificity of the periplasmic lysine-, arginine-, ornithine-binding protein. *J Biol Chem* 269: 26323–26330.
34. Ames GF (1986) Bacterial periplasmic transport systems: structure, mechanism, and evolution. *Annu Rev Biochem* 55: 397–425.
35. Pang A, Arinaminpathy Y, Sansom MS, Biggin PC (2005) Comparative molecular dynamics—similar folds and similar motions? *Proteins* 61: 809–822.
36. Stockner T, Vogel H, Tieleman D (2005) A salt-bridge motif involved in ligand binding and large-scale domain motions of the maltose-binding protein. *Biophys J* 89: 3362–3371.
37. Oh BH, Pandit J, Kang CH, Nikaido K, Gokcen S, et al. (1993) Three-dimensional structures of the periplasmic lysine/arginine/ornithine-binding protein with and without a ligand. *J Biol Chem* 268: 11348–11355.
38. Bernejo GA, Strub MP, Ho C, Tjandra N (2010) Ligand-free open-closed transitions of periplasmic binding proteins: the case of glutamine-binding protein. *Biochemistry* 49: 1893–1902.
39. Loeffler HH, Kitao A (2009) Collective dynamics of periplasmic glutamine binding protein upon domain closure. *Biophys J* 97: 2541–2549.
40. Wolf A, Lee KC, Kirsch JF, Ames GF (1996) Ligand-dependent conformational plasticity of the periplasmic histidine-binding protein HisJ. Involvement in transport specificity. *J Biol Chem* 271: 21243–21250.
41. Flocco MM, Mowbray SL (1994) The 1.9 Å x-ray structure of a closed unliganded form of the periplasmic glucose/galactose receptor from *Salmonella typhimurium*. *J Biol Chem* 269: 8931–8936.
42. Oswald C, Smits SH, Hoing M, Sohn-Bosser L, Dupont L, et al. (2008) Crystal structures of the choline/acetylcholine substrate-binding protein ChoX from *Sinorhizobium meliloti* in the liganded and unliganded-closed states. *J Biol Chem* 283: 32848–32859.
43. Tang C, Schwieters CD, Clore GM (2007) Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature* 449: 1078–1082.
44. Silva D-A, Domínguez-Ramírez L, Rojo-Domínguez A, Sosa-Peinado A (2011) Conformational dynamics of L-lysine, L-arginine, L-ornithine binding protein reveals ligand-dependent plasticity. *Proteins*: In press.
45. Nikaido K, Ames GF (1992) Purification and characterization of the periplasmic lysine-, arginine-, ornithine-binding protein (LAO) from *Salmonella typhimurium*. *J Biol Chem* 267: 20706–20712.
46. Miller DM, 3rd, Olson JS, Pflugrath JW, Quioco FA (1983) Rates of ligand binding to periplasmic proteins involved in bacterial transport and chemotaxis. *J Biol Chem* 258: 13665–13672.
47. Hunenberger PH, Granwehr JK, Aebischer JN, Ghoneim N, Haselbach E, et al. (1997) Experimental and theoretical approach to hydrogen-bonded diastereomeric interactions in a model complex. *J Am Chem Soc* 119: 7533–7544.
48. Maragakis P, Lindorff-Larsen K, Eastwood MP, Dror RO, Klepeis JL, et al. (2008) Microsecond molecular dynamics simulation shows effect of slow loop dynamics on backbone amide order parameters of proteins. *J Phys Chem B* 112: 6155–6158.
49. Bowman GR, Voelz VA, Pande VS (2011) Atomistic folding simulations of the five-helix bundle protein λ_{6-85} . *J Am Chem Soc* 133: 664–667.
50. Higgins CF (1992) ABC transporters: from microorganisms to man. *Annu Rev Cell Biol* 8: 67–113.
51. Quioco FA (1990) Atomic structures of periplasmic binding proteins and the high-affinity active transport systems in bacteria. *Philos Trans R Soc Lond B Biol Sci* 326: 341–351; discussion 351–342.
52. Sandak B, Wolfson HJ, Nussinov R (1998) Flexible docking allowing induced fit in proteins: insights from an open to closed conformational isomers. *Proteins* 32: 159–174.
53. Yang AY, Mancera RL (2008) Molecular dynamics simulations of ligand-induced backbone conformational changes in the binding site of the periplasmic lysine-, arginine-, ornithine-binding protein. *J Comput Aided Mol Des* 22: 799–814.
54. Berendsen H, Postma J, Van Gunsteren W, Hermans J (1981) Interaction models for water in relation to protein hydration. In Pullman B, ed. *Intermolecular Forces*. Dordrecht: Reidel. pp 331–342.
55. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J Chem Theory Comput* 4: 435–447.
56. Oostenbrink C, Soares T, van der Vegt N, van Gunsteren W (2005) Validation of the 53A6 GROMOS force field. *Eur Biophys J* 34: 273–284.
57. Berendsen H, Postma J, Van Gunsteren W, DiNola A, Haak J (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys* 81: 3684–3690.
58. Darden T, York D, Pedersen L (1995) A smooth particle mesh Ewald potential. *J Chem Phys* 103: 3014–3021.
59. Miyamoto S, Kollman P (1992) SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J Comp Chem* 13: 952–962.
60. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: a linear constraint solver for molecular simulations. *J Comput Chem* 18: 1463–1472.
61. Hess K, Berendsen H (1999) Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J Comp Chem* 20: 786–798.
62. Berezhkovskii A, Hummer G, Szabo A (2009) Reactive flux and folding pathways in network models of coarse-grained protein dynamics. *J Chem Phys* 130: 205102.
63. Singhal N, Snow CD, Pande VS (2004) Using path sampling to build better Markovian state models: predicting the folding rate and mechanism of a tryptophan zipper beta hairpin. *J Chem Phys* 121: 415–425.
64. Kang CH, Shin WC, Yamagata Y, Gokcen S, Ames GF, et al. (1991) Crystal structure of the lysine-, arginine-, ornithine-binding protein (LAO) from *Salmonella typhimurium* at 2.7-Å resolution. *J Biol Chem* 266: 23893–23899.
65. Oh BH, Kang CH, De Bondt H, Kim SH, Nikaido K, et al. (1994) The bacterial periplasmic histidine-binding protein. structure/function analysis of the ligand-binding site and comparison with related proteins. *J Biol Chem* 269: 4135–4143.
66. E W, Vanden-Eijnden E (2006) Towards a Theory of Transition Paths. *J Stat Phys* 123: 503–523.