



Universidad Nacional Autónoma de México

Facultad de Química

**“Diseño computacional del sitio de unión de
la proteína periplásmica LAO”**

Tesis

Que para obtener el grado de:

“Maestro en Ciencias Bioquímicas”

Presenta:

Daniel Adriano Silva Manzano

***Tutor:* Dr. Alejandro Sosa Peinado**

México, D. F.

2007



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.



PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS BIOQUÍMICAS

ENTIDADES ACADÉMICAS FACULTAD DE QUÍMICA
E INSTITUTO DE FISIOLÓGÍA CELULAR

Of. No. PMDCB/472/2007

LIBB. DANIEL ADRIANO SILVA MANZANO
Alumno del Programa en Ciencias Bioquímicas
P r e s e n t e

Los miembros del Subcomité Académico, en reunión ordinaria del día 24 de mayo del presente año, conocieron su solicitud de asignación de JURADO DE EXAMEN para optar por el grado de MAESTRO EN CIENCIAS (BIOQUÍMICA), con la tesis titulada "Diseño computacional del sitio de unión de la proteína periplásmica LAO", dirigida por el Dr. Alejandro Sosa Peinado.

De su análisis se acordó nombrar el siguiente jurado:

PRESIDENTE	Dr. Juan Pablo Pardo Vázquez
VOCAL	Dr. Arturo Rojo Domínguez
SECRETARIO	Dr. Gabriel del Río Guerra
SUPLENTE	Dr. Daniel Alejandro Fernández Velasco
SUPLENTE	Dr. Enrique García Hernández

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 30 de mayo de 2007.
EL COORDINADOR DE LA ENTIDAD ACADÉMICA


DR. ROGELIO RODRÍGUEZ SOTRES

C.c.p. Archivo



A mis padres...

Intentando no caer en la tentación de escribir agradecimientos interminables solamente mencionaré siete de estos:

A mis padres: Yolanda Manzano y Daniel Silva.

A mis abuelas: Rosalba Águila y Raquel Medina.

A mi novia Cinthia Liza.

A mi tutor: Alejandro Sosa Peinado.

A mis revisores de tesis.

A mis amigos, incluyendo a los miembros del LFeIP

A la DGSCA, especialmente a la Dra. Genevieve Lucet Lagriffoul, y al M.C. José Luis Gordillo Ruiz.

A todas estas personas agradezco profundamente por sus distintas contribuciones a mi formación académica y humana.

Abreviaturas y siglas	6
Introducción	7
Transportadores ABC	8
La proteína periplásmica que une L-lisina, L-arginina y L-ornitina: LAO	11
Diseño de proteínas	14
ROSETTA en el diseño de secuencias	15
Docking	20
Electrostática	21
Dinámica Molecular	25
Hipótesis.....	34
Objetivos.....	35
General	35
Particulares.....	35
Metodología	36
Estructuras proteicas	36
Docking.....	37
Diseño computacional.....	38
Cálculos electrostáticos.....	39
Simulaciones de dinámica molecular.....	40
Visualización.....	42
Recursos computacionales	42
Resultados y discusión	43
Evaluación de las interacciones PBP-ligando fisiológico por ROSETTA	43
Diseño de LAO para la unión de aminoácidos aromáticos.....	47
Diseño LAO-Fenilalanina.....	48
Diseño LAO-Tirosina	51
Diseño LAO-triptofano.....	52
Diseño LAO-serotonina.....	54
Correlación energía electrostática y lig_sum.....	55
Dinámica molecular de LAO como un método para analizar diseños.....	57
Estabilidad de la estructura secundaria de LAO a diferentes temperaturas	57
Ángulos diedros para determinar el grado de apertura de LAO	58
MD de LAO sin ligando	61
MD de LAO con ligandos fisiológicos	62
<i>MD LAO-arginina</i>	<i>63</i>
<i>MD LAO-histidina</i>	<i>66</i>
<i>MD LAO-lisina</i>	<i>68</i>
<i>LAO, los ligandos fisiológicos y sus propiedades dinámicas.....</i>	<i>71</i>
MD de LAO con ligandos no-fisiológicos.....	71
<i>MD LAO-tirosina</i>	<i>71</i>
<i>MD LAO-fenilalanina.....</i>	<i>74</i>
<i>MD LAO-triptofano</i>	<i>76</i>
Propiedades dinámicas de los ligandos no-fisiológicos vs fisiológicos	78
Análisis de los diseños mediante MD.....	81
Conclusiones	90
Perspectivas	93
Bibliografía.....	95

Abreviaturas y siglas

EA.- Estado abierto de la proteína LAO.

EC.- Estado Cerrado de la proteína LAO.

HSH.- Histidina con carga formal de +1e.

LAO.- Proteína Periplásmica de *Salmonella typhimurium* que une L-lisina, L-arginina y L-ornitina.

MD.- Dinámica molecular, de las siglas en inglés “Molecular Dynamics”.

PDB.- Protein Data Bank

PBP.- Proteína Periplásmica de Unión (bacteriana)., de las siglas en inglés “Periplasmic Binding Protein”.

r.m.s.d.- Desviación media de los cuadrados.

ΔG^{elec} .- Cambio en la energía electrostática.

$\Delta G^{\text{coulomb}}$.- Cambio en la energía de Coulomb.

Introducción

El diseño de proteínas con funciones específicas depende de nuestro entendimiento de la relación entre secuencia, estructura y estabilidad. Dado que las proteínas son moléculas complejas y hay muchos factores que intervienen en su estructura y plegamiento, el diseño de nuevas secuencias se enfrenta a grandes retos.

El aumento en el poder de cómputo ha permitido que hoy en día se realicen estudios *in silico* que hace pocos años se consideraban prohibitivos. Ahora es posible realizar cálculos electrostáticos de complejos macromoleculares o inclusive simulaciones de dinámica molecular (MD) que reflejen el comportamiento dinámico de una proteína solvatada en tiempos de simulación equivalentes al orden de nanosegundos. Por otra parte, durante la última década se han hecho grandes avances en el desarrollo de métodos computacionales que permiten generar secuencias *de novo* a partir de la estructura del esqueleto de una proteína (backbone). Estos avances han abierto el camino para el diseño de proteínas. Al respecto, el programa ROSETTA, creado en el laboratorio del doctor D. Baker (University of Washington, USA), ha tenido un gran éxito en el diseño *de novo* de proteínas (1).

Aunque las proteínas que la naturaleza ha generado y seleccionado desarrollan su papel *in vivo* de manera muy eficiente, se sabe que en términos de estabilidad y velocidad de plegamiento no se encuentran totalmente optimizadas (2-4), lo cual da lugar al refinamiento de dichas propiedades.

La “proteína que une L-lisina, L-arginina y L-ornitina” de *Salmonella typhimurium* (LAO) es una proteína de especificidad múltiple que pertenece a la familia de las “proteínas periplásmicas de unión” de bacterias (PBPs) (5). Estas proteínas forman parte del “sistema de transporte periplásmico” bacteriano, cuya función es justamente transportar una amplia gama de substratos del periplasma al interior de las bacterias (6, 7).

Dada la afinidad nanomolar de LAO por diversos ligandos, su tamaño relativamente pequeño (238 a.a.) y ya que se conocen las estructuras tridimensionales correspondientes a la proteína interactuando con cada uno de sus ligandos, LAO representa un modelo viable para generar diseños del sitio de unión mediante técnicas *in silico*, así como un buen candidato para estudiar las bases de la afinidad múltiple a ligandos.

Transportadores ABC

Los sistemas bacterianos de transporte periplásmico están formados por al menos dos componentes (8):

- 1) Una proteína periplásmica de unión (PBP) que une al ligando,
- 2) Un complejo membranal constituido por dos proteínas integrales de membrana y dos proteínas que hidrolizan ATP.

En estos sistemas, la PBP (que generalmente tiene una alta afinidad por el sustrato) una vez que ha unido al ligando interactúa con el complejo de membrana, desencadenando la activación del mecanismo de translocación del sustrato con el costo energético de una molécula de ATP o GTP (**Figura 1**). Al complejo membranal responsable de este transporte se le conoce como transportador ABC (ATP Binding Cassette) (8).

En resumen, las PBPs tienen dos funciones: unir el sustrato e interactuar con el complejo de membrana. En la literatura se ha sugerido que éstas funciones se encuentran localizadas en distintas regiones de la proteína, ya que en experimentos de mutagénesis dirigida, se han encontrado regiones de la proteína que afectan la capacidad de interactuar con el complejo de membrana sin modificar la capacidad de unir el ligando y viceversa (9-11).

Es importante mencionar que aunque en la familia de las PBPs no se ha encontrado conservación en la estructura primaria, existe una alta conservación en la estructura terciaria, la cual consiste en dos dominios globulares conectados por una “región bisagra” (12, 13).

Las PBPs pueden adoptar al menos dos estados conformacionales: una forma libre de ligando, llamada “estado abierto”; y una forma unida al ligando, llamada “estado cerrado”. El cambio conformacional entre un estado y otro, está dado principalmente por un movimiento en la bisagra que conecta a los dominios (14, 15).

Se han identificado dos clases de PBPs que difieren en la topología de los dominios (16):

clase I) Seis hebras- β en cada dominio,

clase II) Cinco hebras- β en cada dominio.

En ambos casos las hebras- β se encuentran rodeadas por hélices- α .

Se conoce la estructura cristalográfica de más de 100 PBPs (17) y una gran cantidad de ligandos que esta familia puede unir. Los ligandos abarcan un intervalo amplio de formas y tamaños que van desde pequeñas moléculas (ej. iones), hasta moléculas relativamente grandes (ej. disacáridos ó vitaminas).

El sitio de unión de las PBPs presenta dos características particulares:

- 1) Se encuentra en la interfase entre los dos dominios.
- 2) Presenta un gran cambio conformacional al unir el ligando.

En la literatura se ha sugerido que estas características generan propiedades fisicoquímicas que se consideran favorables como son (17):

- a) Posicionamiento del ligando en sitios que excluyen al solvente (y por lo tanto en regiones de una constante dieléctrica baja).

- b) Un gran cambio conformacional permite el desarrollo de diferentes funciones en cada estado conformacional.

En la ingeniería de proteínas se ha utilizado a las PBPs para diseñar biosensores, elementos de control alostérico, receptores biológicamente activos e inclusive para el diseño de actividades enzimáticas *de novo* (17, 18).

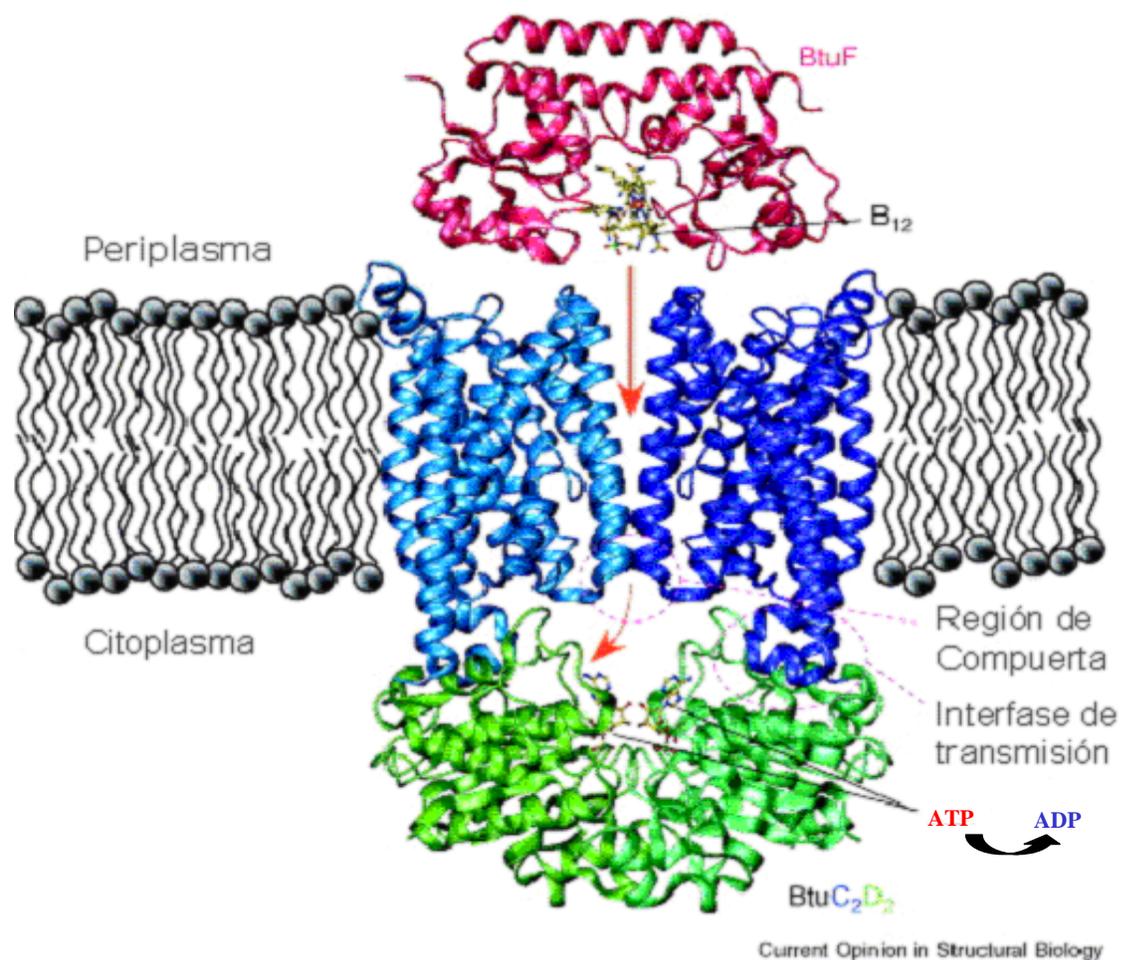


Figura 1. Ejemplificación del mecanismo de acción del transportador ABC de la vitamina B₁₂ de *Escherichia coli*. El transporte ocurre cuando la PBP (BtuF) interactúa con el transportador de membrana (complejo BtuC₂), evento que da lugar a la hidrólisis de una molécula de ATP por el complejo asociado “BtuD₂”, esta hidrólisis aporta la energía necesaria para que el ligando sea translocado al interior de la célula (viajando a través del complejo “BtuC₂”) y finalmente la PBP es liberada del transportador de membrana. *Figura modificada de la referencia 3.*

La proteína periplásmica que une L-lisina, L-arginina y L-ornitina: LAO

LAO es una PBP de 238 aminoácidos con un puente disulfuro entre los residuos 38 y 45, una masa molecular de 26 kDa y una absorbancia de 0.85 ($A_{280}^{0.1\%}$) (5). Su estructura tridimensional sin ligando se encuentra resuelta por cristalografía de rayos-X a 2.7 Å de resolución (10) (Figura 2).

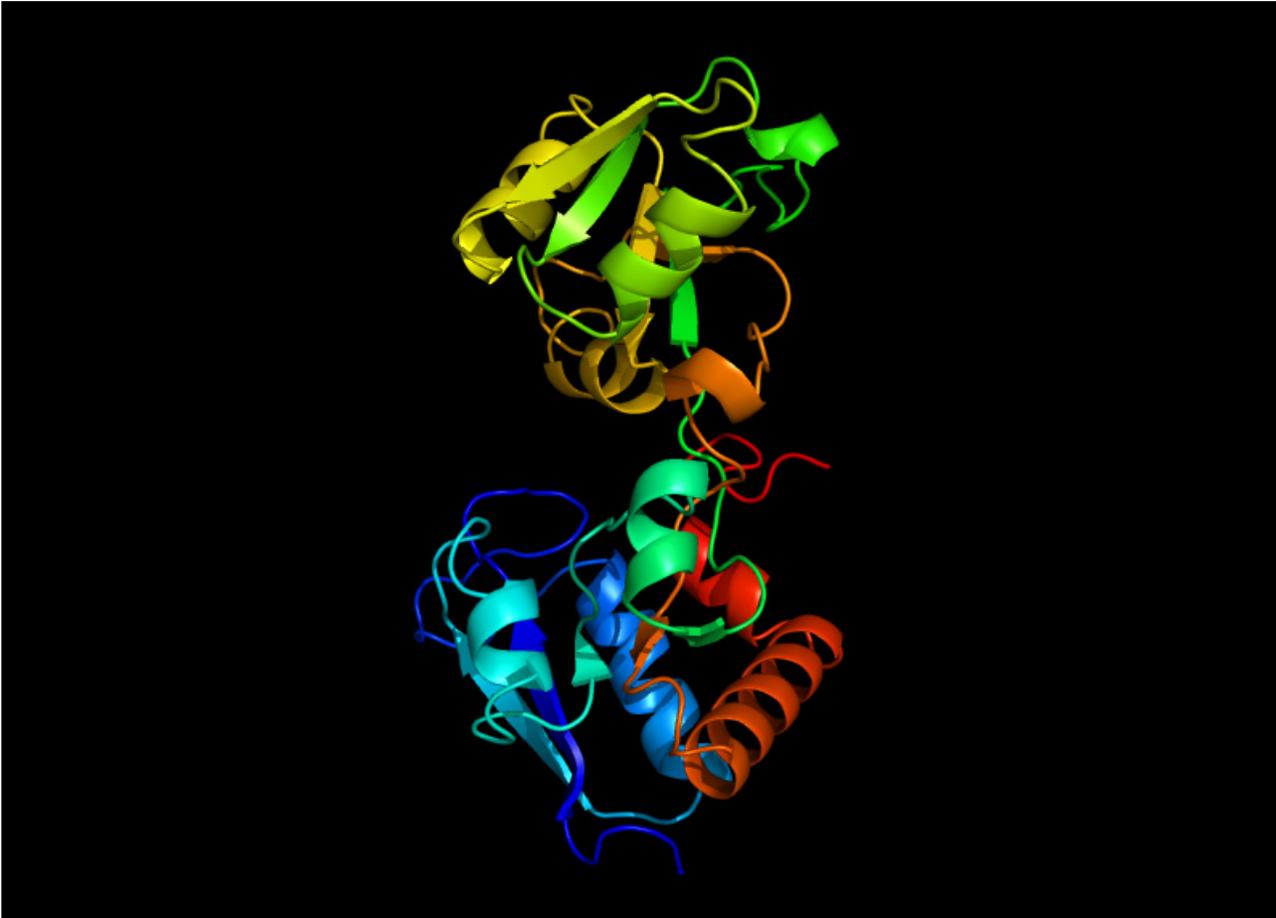


Figura 2. LAO en el estado abierto. Representación: “cartoons” (PDB: 2LAO). En la imagen se aprecia el lóbulo I (dominio inferior de la proteína), el lóbulo II (dominio superior de la proteína) y la región de bisagra que une a los dominios. El extremo amino se encuentra en color azul oscuro y el extremo carboxilo en color rojo.

LAO pertenece a la clase II de las PBP ya que tiene cinco hebras- β y cuatro hélices- α en cada lóbulo, cuatro de estas hebras- β son paralelas y una es antiparalela (en cada lóbulo). El lóbulo I contiene los residuos del extremo NH_2 -terminal 1-88 y a los residuos 195-238 del extremo COOH -terminal, mientras que el lóbulo II contiene los residuos 93 a 185. El COOH -terminal se localiza cerca de la bisagra y el NH_2 -terminal se encuentra en un extremo del lóbulo I. LAO

comparte el transportador de membrana llamado HisQ/M/P₂ con otra proteína periplásmica llamada HisJ (10).

Dentro de la familia de las PBPs LAO es considerada pequeña, ya que el tamaño promedio de los integrantes de esta familia es de 350 a.a. Se ha sugerido que esta diferencia de tamaño se ve reflejada tanto en la topología de la bisagra (que se encuentra formada por sólo dos segmentos polipeptídicos discontinuos en lugar de tres como la mayoría de las PBP), como en el hecho de que LAO carece del “subdominio COOH-terminal” que frecuentemente se encuentra en otras PBPs.

Se conocen cuatro ligandos fisiológicos (todos de alta afinidad) para la LAO: L-arginina, L-lisina, L-ornitina y L-histidina, con constantes de disociación (K_d) de 14, 15, 30 y 500 nM respectivamente (5). En estudios de competencia de ligando se han probado compuestos similares a estos aminoácidos como: ácido L-arginínico, L-glutamina y los enantiómeros-D de los ligandos arginina, lisina y ornitina. Todos han mostrado ser substratos de baja afinidad comparados con los ligandos fisiológicos ($K_d \approx$ “orden mM”) (5). Estos datos indican que en LAO el reconocimiento del ligando está determinado, además de por interacciones polares, por la forma y el tamaño. Los ligandos de LAO comparten la característica de tener carga positiva, sugiriendo una importante contribución electrostática para la unión.

En estudios de fluorimetría se han encontrado diferencias en los espectros de absorción en UV entre la proteína en el estado abierto (sin ligando) y el estado cerrado (con ligando). También se han encontrado diferencias en el espectro de absorción de LAO en presencia de los distintos ligandos. Se ha propuesto que estas diferencias pueden ser explicadas por los distintos estados conformacionales que adopta la proteína al unir sus ligandos (5).

Una ventaja de LAO como modelo teórico es que se conocen las estructuras tridimensionales de la proteína unida a cada uno de los cuatro ligandos de alta afinidad (**Figura 3**) (15). Globalmente

la conformación de la proteína LAO es la misma para las cuatro formas ligadas. Al comparar las estructuras tridimensionales de LAO unida a los distintos ligandos se encuentra que prácticamente no hay diferencias entre ellas (el r.m.s.d. de los C α entre los residuos 5-235 es menor a 0.18 Å) (19). La superposición de las cuatro estructuras con ligando muestra que los residuos involucrados directamente en la unión no difieren en la localización espacial de sus cadenas laterales (Asp-11, Tyr-14, Phe-52, Asp-30, Ser-69, Ser-70, Ser-72, Arg-77, Leu-117, Thr-121 y Asp 161) con la excepción del Asp-11 (19).

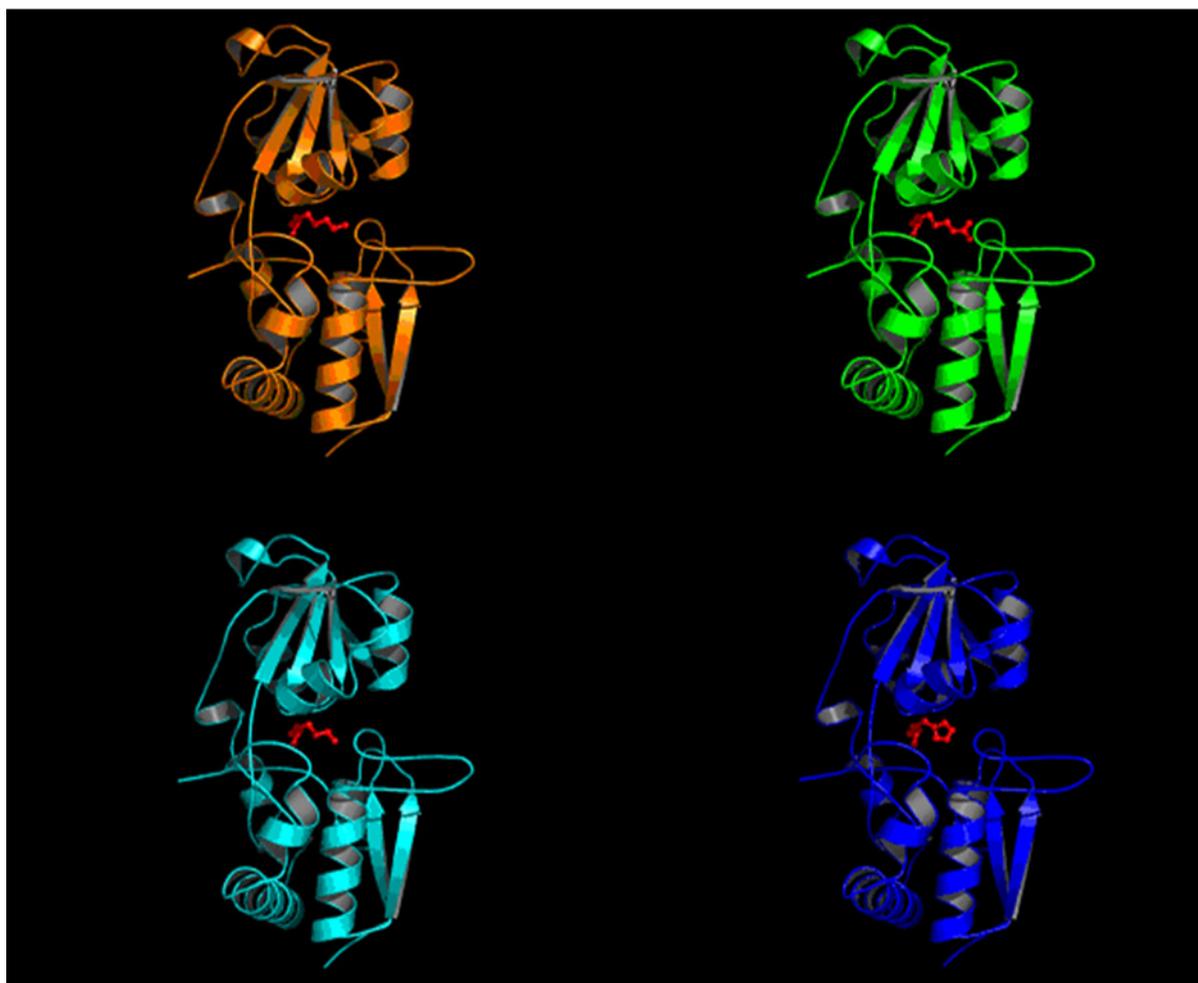


Figura 3. *La proteína LAO en el estado cerrado con sus distintos ligandos.* Representación de listones. En la parte superior izquierda el ligando L-lisina; en la parte superior derecha el ligando L-arginina; en la parte inferior izquierda el ligando L-ornitina; y en la parte inferior derecha el ligando L-histidina.

LAO, en comparación con otras PBP (ej. proteína que une galactosa [GBP], proteína que une arabinosa [ABP] o la proteína que une leu, ile y val [LIV-BP]), tiene una afinidad muy alta por sus ligandos (nanomolar vs milimolar). Esta característica se ha atribuido a las múltiples interacciones iónicas que forma la proteína con sus ligandos (**19**). Pese a los estudios que se han hecho en LAO, aún desconocemos el proceso a través del cual el ligando es reconocido y estabilizado en el sitio de unión de la proteína.

En las diferentes estructuras tridimensionales de LAO se encuentran conservadas dos moléculas de agua en el sitio de unión (Wat 401 y 402) (**19**). También existe una tercera molécula de agua (Wat 403), que participa directamente en la unión de la histidina y se ha sugerido que tiene un papel de neutralización de cargas parciales (**19**). En resumen, la “cavidad” de unión de LAO es lo suficientemente grande para alojar el volumen de cualquiera de los cuatro ligandos mas 2 ó 3 moléculas de agua (Wat 401, 402 y 403).

Diseño de proteínas

El diseño de proteínas por computadora utilizando algoritmos de búsqueda de secuencias y funciones de energía se conoce como “diseño racional de proteínas”, dada su capacidad de generar resultados en términos de unas pocas secuencias que presuntamente contienen las características deseadas (**20**). Para diseñar proteínas que adopten una conformación determinada se requiere evaluar la capacidad de una enorme cantidad de secuencias para adquirir la estructura y función deseadas. Los algoritmos de búsqueda necesariamente hacen un compromiso entre la velocidad de la búsqueda y su precisión (ya que inclusive para un número pequeño de aminoácidos es imposible analizar todo el espacio combinatorio). Se sabe que existen importantes dependencias entre la elección del algoritmo de búsqueda, la forma en que el espacio muestral se representa y las funciones de evaluación utilizadas.

ROSETTA en el diseño de secuencias

Una estrategia para el diseño de proteínas fue descrita por Kuhlman y Baker (21) e implementada en el programa ROSETTA (22, 23). Este algoritmo, al igual que otros basados en métodos de Monte-Carlo, prueba diferentes combinaciones de aminoácidos en las posiciones deseadas de una estructura protéica. La diferencia con otros algoritmos de diseño basados en Monte-Carlo se encuentra principalmente en:

- Las combinaciones de rotámeros que usa ROSETTA, para hacer búsquedas, provienen de bases de datos “jerarquizados” obtenidos de estructuras tridimensionales conocidas (Protein Data Bank, PDB (24)). La idea detrás de esta jerarquización es que con el uso de unos pocos rotámeros que se encuentran con frecuencia en la naturaleza es posible construir diseños exitosos de proteínas. Aunque pareciera que alimentar al algoritmo con estas bases de datos no difiere mucho de otros protocolos propuestos, el simple hecho de trabajar con rotámeros jerarquizados hace que el problema combinatorio sea mucho más pequeño que si se trabajara con todos los rotámeros posibles. De manera práctica, los diseños preliminares se pueden realizar con los rotámeros de la primera jerarquía (es decir los que aparecen con mayor frecuencia). Si se requiere refinar posiciones específicas en el diseño, se trabaja con una base de rotámeros más amplia, pero basándose en el diseño preliminar, lo que permite saber de antemano qué aminoácido o tipo de aminoácido (ej. polar, aromático, etc.) se utilizará en cada posición del refinamiento.
- La mayor bondad del programa ROSETTA no se encuentra sólo en su función de búsqueda, ya que además cuenta con una función semiempírica de evaluación de energía que ha demostrado ser excelente para propósitos de diseño, y probablemente sea la parte

clave del éxito que ROSETTA ha mostrado tanto en predicciones de estructuras terciarias como en el diseño de proteínas.

Con esta aproximación teórica, los autores, partiendo sólo de estructuras (tridimensionales) del esqueleto de proteínas conocidas, han diseñado secuencias de aminoácidos que experimentalmente han mostrado plegarse en estructuras estables (25). La capacidad de ROSETTA en el diseño de proteínas se puso a prueba nuevamente en otro trabajo de los mismos autores, para diseñar una proteína *de novo* con una estructura terciaria que no se ha observado nunca en la naturaleza. Este diseño fue completamente exitoso y dio lugar a la primer proteína de este tipo llamada Top7 (1).

La función de energía utilizada para evaluar los diseños en el algoritmo de ROSETTA es la suma lineal de 11 términos que se encuentran sujetos independientemente a un factor de ponderación (W) (23), el resultado de esta suma es expresado por el programa ROSETTA en un parámetro llamado “bk_tot”:

$$E_{\text{protein}} = W_{\text{rot}}E_{\text{rot}} + W_{\text{aa}|\phi,\psi}E_{\text{aa}|\phi,\psi} + W_{\text{rama}}E_{\text{rama}} + W_{\text{atr}}E_{\text{atr}} + W_{\text{solv}}E_{\text{solv}} + W_{\text{pair}}E_{\text{pair}} + W_{\text{bb_hbond}}E_{\text{bb_hbond}} + W_{\text{sc_hbond}}E_{\text{sc_hbond}} + W_{\text{sc_bb_hbond}}E_{\text{sc_bb_hbond}} + W_{\text{pair}}E_{\text{pair}} - E_{\text{ref}}$$

A continuación se explica la naturaleza de cada término y de los factores de peso:

En la función de energía de ROSETTA, para calcular la energía de solvatación (E_{solv}) y la energía de Lennard-Jones (E_{atr} y E_{rep}), los distintos átomos de los 20 aminoácidos se separan en tipos. Un potencial estándar de Lennard-Jones se utiliza con una modificación que hace que después de una distancia de corte el potencial se extrapole linealmente. Las energías favorables se agrupan en el término E_{atr} y las energías desfavorables en E_{rep} :

$$E_{atr} = \sum_i^{natom} \sum_{j>i}^{natom} \left[\left(\frac{r_{ij}}{d_{ij}} \right)^{12} - 2 \left(\frac{r_{ij}}{d_{ij}} \right)^6 \right] e_{ij} \quad \text{si } \frac{r_{ij}}{d_{ij}} < 1.12$$

$$E_{rep} = \sum_i^{natom} \sum_{j>i}^{natom} \left[\left(\frac{r_{ij}}{d_{ij}} \right)^{12} - 2 \left(\frac{r_{ij}}{d_{ij}} \right)^6 \right] e_{ij} \quad \text{si } 1.33 > \frac{r_{ij}}{d_{ij}} > 1.12 + \sum_i^{noatom} \sum_{j>i}^{natom} y_{intercept} - d_{ij} * slope$$

$$\text{si } \frac{r_{ij}}{d_{ij}} > 1.33 \quad slope = -12e_{if} (1.33^{13} - 1.33^7) * \left(\frac{1}{r_{ij}} \right)$$

$$y_{intercept} = slope * \left(\frac{r_{ij}}{1.33} \right) + e_{if} (1.33^{12} - 2(1.33)^6)$$

$$r_{ij} = r_i + r_j$$

$$e_{ij} = \sqrt{e_i + e_j}$$

Donde: “d” es la distancia de los elementos no-unidos, “r_{ij}” es el diámetro de Lennard-Jones, y “e” es la profundidad del pozo de dispersión.

Para evaluar la energía de solvatación (**E_{solv}**) se usa un modelo de solvatación implícita desarrollado por Lazaridis y Karplus (26):

$$E_{solv} = \sum_i^{natom} \sum_{j>i}^{natom} \left\{ \frac{-2\Delta G_i^{free}}{4\pi\sqrt{\pi}\lambda_i r_{ij}^2} \exp(-d_{ij}^2) V_j - \frac{-2\Delta G_j^{free}}{4\pi\sqrt{\pi}\lambda_j r_{ij}^2} \exp(-d_{ij}^2) V_i \right\}$$

Donde d_{ij} y r_{ij} representan a las mismas variables que en E_{atr}. El término ΔG_{free} está relacionado con la energía de solvatación del átomo completo, λ_i es una medida de correlación y V es el volumen atómico. Los valores de los parámetros se han tomado de Lazaridis y Karplus, con excepción de algunos de los valores de ΔG_{free} que se han perturbado para reproducir mejor las frecuencias relativas de los aminoácidos en el núcleo proteico.

La energía de cada rotámero (**E_{rot}**) se calcula con la siguiente ecuación:

$$E_{rot} = \sum_i^{nres} -\ln(P(rot(i) | phi(i), psi(i)))$$

donde E_{rot} representa la energía interna de un rotámero, derivada de estadísticas del “Protein Data Bank” (PDB), en las que se observa la probabilidad de un rotámero/aminoácido en particular para un ángulo phi o psi dado (27).

$E_{aa|phi,psi}$ se refiere a las preferencias de los aminoácidos por regiones particulares del espacio phi o psi. Para determinar este componente, los autores utilizaron un grupo no-redundante de archivos PDB para determinar las probabilidades de observar cada uno de los 20 aminoácidos en un espacio phi, psi ($P(aa|phi,psi)$). La energía se calcula como el $(-\log)$ de las probabilidades.

E_{rama} es el potencial de torsión dependiente de phi y psi de los aminoácidos. Para este término se ha determinado la frecuencia de los pares phi-psi.

E_{pair} está dado por la probabilidad de encontrar a dos aminoácidos a una cierta distancia en el espacio. Los datos se derivaron de la base de datos PDB y el término se evalúa para aminoácidos polares considerando dos ambientes: aminoácidos atrapados en un ambiente no polar y aminoácidos expuestos. La ecuación utilizada es:

$$E_{pair} = \sum_i^{nres} \sum_{j>i}^{nres} \frac{P(aa_i, aa_{jj} | d_{ij}, env_i, env_j)}{P(aa_i, aa_{jj} | d_{ij}, env_i)P(aa_j | d_{ij}, env_j)}$$

E_{bb_hbond} , E_{sc_hbond} , $E_{bb_sc_hbon}$. En cada una de estas contribuciones energéticas se suma un término de puentes de hidrógeno dependiente de la orientación: esqueleto $\leftarrow \rightarrow$ esqueleto; cadena lateral $\leftarrow \rightarrow$ esqueleto; y cadena lateral $\leftarrow \rightarrow$ cadena lateral, respectivamente, que están evaluados a partir de los ángulos y las distancias observadas en las proteínas nativas.

E_{ref} es la energía de referencia con respecto al estado desplegado de la proteína y está dada por:

$$E_{ref} = \sum_i^{nres} W_{ref}(aa(i))$$

Para calcular esta función, a cada aminoácido se le asigna una energía de referencia calculada empíricamente.

Funciones de peso W. Los pesos para los términos y las 20 energías de referencia se determinaron por la maximización del producto: $\exp(-E(aa_{obs}))/\exp(-E(aa_i))$ sobre un grupo de referencia de 30 proteínas, usando un método de optimización de gradiente-conjugado, donde $E(aa_{obs})$ es la energía del aminoácido nativo en una posición dada y la función de partición en el denominador es sobre los 20 aminoácidos en cada posición. En este proceso un residuo fue cambiado a la vez y los otros residuos se mantuvieron fijos en la conformación nativa. Subsecuentemente los parámetros se refinaron basándose en los cálculos del diseño de las proteínas del grupo de referencia.

En una versión reciente del programa ROSETTA (ROSETTALIGAND) (28), los autores han generado una nueva función de energía para la evaluación de las interacciones “proteína \leftrightarrow ligando”. Partiendo de la idea de que la función de energía para evaluar ligandos es correcta, los autores han propuesto que, “utilizando esta nueva función de energía en conjunto con la misma filosofía de diseño ya existente en ROSETTA”, en teoría, ROSETTA sería capaz de diseñar (*de novo*) sitios proteicos de unión a ligandos. En la práctica, esta nueva función de energía para pequeños ligandos es usada al mismo tiempo que la función de energía para el diseño de proteínas, de manera que al mismo tiempo que ROSETTA diseña el sitio de unión, se asegura de no afectar la estabilidad de la estructura proteica. Esta nueva función de energía para ligandos (lig_sum) está integrada por la suma de cinco componentes (28):

- 1) Un potencial de Lennard Jones 12-6 estándar para modelar interacciones de atracción con pozos de dispersión y radios de van der Waals obtenidos del campo de fuerza CHARMM27.

- 2) Un término de repulsión que conecta la amplitud y la pendiente con el potencial $12-6$ a $E = 0$, y que después es lineal hasta que dos átomos se encuentran a 0\AA de distancia.
- 3) Un término de solvatación similar al de Lazaridis–Karplus con un modelo de solvatación implícita.
- 4) Un potencial explícito de puentes de hidrógeno.
- 5) Un modelo de Coulomb con una constante dieléctrica dependiente de la distancia y que usa las cargas parciales del campo de fuerza CHARMM27.

En ésta función, los factores de peso de cada término fueron calculados por los autores con base en grupos de entrenamiento.

Como se puede ver, la función de ROSETTA para ligandos es aditiva por pares y es mucho más simple que la función utilizada para el diseño de proteínas, por lo que se puede calcular muy rápidamente.

Docking

Localizar el sitio de unión de un ligando en una estructura proteica es un problema complejo que ha sido abordado por el método computacional conocido como Docking (anclaje molecular simulado) (29). El problema del Docking se compone esencialmente de dos partes: 1) buscar las conformaciones espaciales y con características fisicoquímicas favorables para la interacción entre un receptor y su ligando, y 2) estimar la energética de la interacción. Ambas partes se encuentran relacionadas en los algoritmos de Docking, ya que estos métodos dependen de una función que evalúa las conformaciones espaciales, al mismo tiempo que calcula la energética de la interacción, y de esta manera determina cuales conformaciones tienen mayor probabilidad de representar un complejo biológicamente significativo.

Se han desarrollado diversos algoritmos de búsqueda por Docking que comprenden desde los algoritmos de fuerza bruta (intentar explorar “todas” las combinatorias de conformaciones y posibles sitios de unión), hasta otros menos demandantes computacionalmente, como algoritmos genéticos y de Monte-Carlo. También existen distintos métodos de evaluación de los complejos (“scoring functions”) y se clasifican principalmente en: 1) basados en campos de fuerza explícitos, y 2) semi-empíricos. Los primeros estiman la energía de unión a partir de campos de fuerza como CHARMM o AMBER, mientras que las funciones semi-empíricas se basan en la suma ponderada de diversos tipos de interacciones como: la formación de puentes de hidrógeno, efectos de solvatación, etc. Los pesos relativos se derivan mediante ajustes de datos de “entrenamiento” con afinidades experimentales conocidas.

Para el caso del diseño del sitio de unión de una proteína, probablemente la función de evaluación más relevante es la de generar una predicción de la afinidad de unión. Aunque en los últimos años se han realizado avances en la precisión de la predicción por docking, los métodos aún no son tan confiables como se desearía (30).

Electrostática

Identificar las interacciones responsables de la unión específica de ligandos es un requisito esencial para el diseño de proteínas. La unión de ligandos está dada por una gran cantidad de factores, entre los que se encuentran: la geometría del sitio de unión, interacciones polares, interacciones hidrofóbicas y energías de solvatación.

El proceso de unión ligando-proteína puede ser descrito en 3 etapas consecutivas:

- Difusión del ligando hasta la superficie proteica.
- Difusión del ligando a través de la proteína hasta el sitio de unión.
- Rearreglo del ligando y del sitio de unión hacia la orientación adecuada.

Aunque la fuerza motriz para este proceso es generalmente el efecto hidrofóbico, las interacciones electrostáticas también son importantes, tanto para los residuos polares como para los apolares (31). En el caso de la PBP se puede definir una cuarta etapa asociada al cambio conformacional de la proteína.

La electrostática juega un papel crítico en la estructura de las proteínas y su función. No obstante, para propósitos de diseño no existe un consenso acerca de cómo se deben modelar estas interacciones. Un problema es que tratar de calcular estas interacciones con la ecuación de Poisson-Boltzmann y al mismo tiempo utilizarla dentro de una estrategia de diseño combinatorio no es factible, debido a la complejidad misma que se requiere para llegar a la convergencia con esta ecuación. Como alternativa, aproximaciones simplificadas, como reducir las constantes dieléctricas dependientes de la distancia, se han usado en algunas estrategias de diseño (32, 33). Se piensa que una función efectiva para evaluar estas interacciones se puede obtener a través de las interacciones de Coulomb (34), y una práctica frecuente es imponer un punto de corte para la contribución máxima de las interacciones electrostáticas que tiene como resultado un menor tiempo de cálculo (35).

En general, las contribuciones energéticas de las interacciones electrostáticas en los procesos de plegamiento y asociación resultan de un balance entre dos efectos separados: 1) una contribución desestabilizante al proceso de asociación debido a la pérdida de interacciones favorables de solvatación entre los residuos cargados o polares y la fase acuosa altamente dieléctrica, y 2) un cambio en las interacciones de los pares de Coulomb como resultado del plegamiento o de la asociación de dos cuerpos. El signo y la magnitud de estos efectos dependen de las posiciones de las cargas que interactúan. Los cálculos basados en electrostática continua sugieren que, para los pares aislados de los grupos polares y los pares ionizados que están interactuando, las interacciones de Coulomb generalmente no son lo suficientemente intensas para compensar el

efecto desfavorable de la desolvatación (36). Entonces, se espera que la contribución de las interacciones electrostáticas a un proceso de ensamblaje sea desfavorable. De lo anterior se deduce que se puede esperar una gran contribución de este efecto desfavorable mientras haya más aminoácidos cargados y grupos polares en la superficie de unión. Esto da lugar a la pregunta ¿de dónde viene la energía que lleva a la formación de los complejos? En ciertos estudios se ha encontrado que, en efecto, las interacciones electrostáticas tienden a oponerse a la formación de los complejos, mientras que la fuerza que promueve la unión proviene principalmente de las interacciones no polares (36). Sin embargo, Lee y Tidor han hecho observaciones que indican que la interfase de la proteína Barstar está electrostáticamente optimizada para interactuar con la Barnasa, aunque esto no implica que todas las interfases proteicas estén optimizadas de la misma manera (37).

Los cálculos electrostáticos generalmente suponen casos hipotéticos donde no existe cambio conformacional alguno después de la unión, es decir, los cálculos se llevan a cabo suponiendo que la estructura de los monómeros, o del monómero y el ligando, son iguales tanto en el complejo como en la forma libre (36). Aunque sin duda esta observación es una gran simplificación lo que sucede *in vivo*, ofrece una aproximación efectiva para la solución del problema.

La energía electrostática total de un sistema macromolecular se define como el trabajo necesario para ensamblar todas sus cargas y se calcula como (36):

$$\Delta G_{elec}^{todo} = \frac{1}{2} \sum_{i=1}^N q_i \phi(r_i)$$

En esta ecuación q_i es la carga del cuerpo i , $\phi(r_i)$ es el valor del potencial electrostático en la posición del cuerpo i , y la suma se lleva a cabo sobre todas las cargas atómicas parciales de un sistema molecular. La expresión dada en la ecuación anterior corresponde a la energía libre de un

sistema en relación a una estructura de forma idéntica pero sin cargas en él. Esta expresión puede utilizarse para definir la contribución electrostática de los procesos de ensamblaje macromolecular. Para una asociación proteína-proteína (o proteína-ligando) (36):

$$\Delta G_{elec}^{todo} = \frac{1}{2} \sum_{i \in A:B} q_i \phi^{complejo}(r_i) - \frac{1}{2} \sum_{i \in A} q_i \phi^{libre}(r_i) - \frac{1}{2} \sum_{i \in B} q_i \phi^{libre}(r_i)$$

La suma en cada uno de los tres términos de esta ecuación se realiza por separado para las cargas parciales en las estructuras: A, B y el complejo A:B. Con la aproximación de un sistema macromolecular en una conformación fija y en un estado de protonación dado, la ecuación describe todos los efectos electrostáticos que contribuyen a la unión del complejo, incluyendo las interacciones intramoleculares de Coulomb y las interacciones electrostáticas del solvente modificado tras la unión.

Cuando la forma lineal de la “ecuación de Poisson-Boltzmann” (PBE) se usa para calcular los potenciales electrostáticos, el potencial en cualquier punto puede expresarse como la suma de las contribuciones de las cargas individuales (i.e. $s_j(r_i) = \sum_{j=1}^N q_j / r_{ij}$, donde $s_j(r_i)$ es el potencial electrostático en la posición i debido a la carga q_j).

Existen pocas soluciones analíticas de la PBE para las geometrías de las biomoléculas. Sin embargo, esta ecuación se resuelve numéricamente por distintos métodos computacionales, los cuales transforman la solución continua en un conjunto dimensional-finito de funciones básicas. En el caso de la forma lineal de la PBE, las ecuaciones resultantes se pueden arreglar en la forma de matriz-vector y ser resueltas directamente. Pero las ecuaciones no lineales obtenidas de la ecuación completa de PBE requieren de técnicas más especializadas como los métodos de Newton (38).

Dinámica Molecular

El término “modelado molecular” indica el proceso general de describir “sistemas químicos complejos” en términos de un “modelo atómico realista”, con el objetivo de entender o predecir propiedades macroscópicas basándose en un conocimiento detallado a escala microscópica (atómica).

La dinámica molecular (MD) es un método que fue introducido alrededor en 1957 por Alder y Wainwright (Alder and Wainwright, (39, 40)) para estudiar las interacciones de esferas sólidas. El siguiente avance crucial se dio en 1964, cuando Rahman realizó la primera dinámica molecular utilizando un potencial realista para argón líquido (41). Después, en 1974, Stillinger y Rahman realizaron la primera MD de un “sistema realista” compuesto por 216 moléculas de agua en una simulación equivalente a 8 ps (42). Posteriormente, en 1977, McCammon et al. reportaron la primer simulación de MD para una proteína: “el inhibidor pancreático de tripsina bobina (BPTI)”, con una integración temporal de 10 ps (43). En los siguientes 30 años los algoritmos y técnicas computacionales que se utilizan para realizar dinámicas moleculares han mejorado considerablemente. Esto, aunado al creciente poder de cómputo, ha permitido que hoy en día se realicen simulaciones de proteínas, pequeños complejos proteicos e inclusive complejos de membrana en solventes explícitos, con integraciones temporales de nanosegundos.

Las simulaciones de dinámica molecular generan información de nivel microscópico de las propiedades de un sistema, incluyendo las posiciones de los átomos y la velocidad con que estos se mueven. La conversión de esta información en propiedades macroscópicas como la presión, la energía, la capacidad calorífica, etc, requiere del empleo de la ‘mecánica estadística’ (44).

Las propiedades macroscópicas se pueden dividir en:

- a) Propiedades de equilibrio (ej. constantes de unión, la energía potencial promedio)

- b) Propiedades dinámicas o de no-equilibrio (ej. viscosidad de un líquido, cinéticas de reacción, procesos de difusión)

La elección de la técnica para calcular las propiedades dinámicas de un sistema depende de la pregunta que se desea responder y de la capacidad del método para dar resultados útiles. Idealmente, la ecuación de Schrödinger describe las propiedades de sistemas moleculares con una gran certeza, pero cualquier cosa más compleja que un sistema en equilibrio con pocos átomos no puede ser manejado a nivel *ab-initio*. Por lo tanto, se requieren aproximaciones al problema. Mientras más complejo es el sistema y más largo el tiempo de integración que se quiere estudiar, se requieren aproximaciones más reduccionistas, y en algún momento la complejidad es tal que la aproximación *ab-initio* debe ser aumentada o remplazada por una parametrización empírica del modelo usado.

Las propiedades macroscópicas son promedios de un ensamble estadísticamente representativo de un sistema molecular (ya sea en equilibrio o no). Para el modelado molecular esto tiene dos consecuencias importantes (44):

- 1) Conocer una sola estructura, aunque ésta estructura corresponda al mínimo de energía global, no es suficiente para calcular propiedades macroscópicas, por lo que es necesario generar un ensamble representativo a una temperatura dada para poder calcular estas propiedades. Sin embargo, esto no es suficiente para calcular las propiedades termodinámicas en el equilibrio basadas en términos de energía libre, como son: equilibrios de fase, constantes de unión, solubilidades, estabilidades de conformaciones moleculares, entre otras. Para calcular las energías libres y los potenciales termodinámicos se requiere una extensión de las técnicas de simulación molecular.

- 2) Mientras que las simulaciones de dinámica molecular (en principio) proveen detalles atómicos de las estructuras y de sus movimientos, frecuentemente estos detalles no son relevantes para las propiedades macroscópicas de interés, de manera que se abre un camino para simplificar la descripción de las interacciones y realizar promedios sobre detalles relevantes. La mecánica estadística provee un andamio teórico para tales simplificaciones, y existen una gran variedad de métodos que van desde considerar grupos de átomos como una unidad, hasta dinámicas mesoscópicas que describen densidades en lugar de átomos; flujos que responden a gradientes termodinámicos en lugar de velocidades ó aceleraciones que responden a fuerzas.

Para generar un ensamble representativo del equilibrio existen al menos dos métodos:

- 1) Simulaciones de Monte-Carlo (MC)
- 2) Simulaciones de Dinámica Molecular (MD)

Mientras que para generar ensambles que no se encuentren al equilibrio sólo se pueden utilizar las dinámicas moleculares. Por otra parte, a pesar de que en las simulaciones de Monte-Carlo no se requiere el cálculo de fuerzas y, por lo tanto, son más sencillas que las MD (en términos de cálculos), en general con las simulaciones de Monte-Carlo no son capaces de obtener mejores resultados que con las MD en el mismo tiempo de cómputo.

Cuando se realizan simulaciones de MD frecuentemente se desea hacer una extrapolación a las propiedades macroscópicas del sistema. La mecánica estadística provee de relaciones matemáticas que relacionan las propiedades macroscópicas con la distribución y movimiento de los átomos de un sistema. Este sistema puede ser desde uno simple, como una colección de moléculas de solvente, hasta un sistema complejo como una o varias proteínas solvatadas. Con el objetivo de relacionar el comportamiento microscópico de un sistema con el macroscópico, se requiere hacer relaciones dependientes con el tiempo.

El estado termodinámico de un sistema se define usualmente por un número limitado de parámetros, por ejemplo: la temperatura (T), la presión (P) y el número de partículas (N).

El estado mecánico o estado microscópico se encuentra definido por las posiciones atómicas (p) y el momento (q). Estas dos “propiedades” también pueden ser consideradas como coordenadas en un espacio multidimensional llamado “espacio de fase”.

Las simulaciones de dinámica molecular se basan en la segunda ley de Newton (ecuación del movimiento): $F = ma$, donde “F” es la fuerza aplicada a una partícula, “m” es la masa y “a” es la aceleración. Conociendo la fuerza en cada átomo en el sistema es posible determinar la aceleración de cada átomo. La integración de la ecuación del movimiento da lugar a una trayectoria que describe la posición, velocidad y aceleración de las partículas del sistema a lo largo del tiempo. El método es determinista, es decir, una vez que se conocen las velocidades y posiciones de cada átomo, en principio, el estado del sistema se puede predecir en cualquier momento en el pasado o en el futuro, aunque puede ser muy costoso computacionalmente. Hoy en día se realizan simulaciones de dinámica molecular de proteínas solvatadas con tiempos de integración de nanosegundos, e inclusive existen casos en los que se han reportado tiempos de integración en la escala de los milisegundos.

Si tomamos la ecuación del movimiento de Newton:

$$F_i = m_i a_i = m \frac{d^2 x}{dt^2}$$

F es la fuerza aplicada a una partícula “i”, “m” es la masa de la partícula “i” y “a” es la aceleración de la partícula “i”. La fuerza también se puede expresar como el gradiente de la energía potencial:

$$F_i = -\nabla_i V = -\frac{dV}{dr_i}$$

combinando estas dos ecuaciones tenemos

$$-\frac{dV}{dr_i} = m_i \frac{d^2 r_i}{dt^2}$$

donde “V” es la energía potencial del sistema y “r” es la posición. Se observa que la ecuación del movimiento puede relacionar la derivada de la energía potencial con respecto a la posición, con los cambios en las posiciones de los átomos en función del tiempo.

$$F = ma = \frac{dV}{dr} = m \frac{d^2 x}{dt^2}$$

Si se toma un caso sencillo en el que la aceleración es constante:

$$a = \frac{dv}{dt}$$

Se obtiene una expresión para la velocidad tras la integración:

$$v = at + v_0$$

y dado que

$$v = \frac{dx}{dt}$$

se puede integrar nuevamente para obtener:

$$x = vt + x_0$$

Si se combina esta ecuación con la expresión para la velocidad obtenemos la siguiente relación que nos da el valor de “x” en el tiempo “t” como una función de la aceleración, la posición inicial “x₀” y la velocidad inicial “v₀”:

$$x = at^2 + v_0t + x_0$$

La aceleración está dada por la derivada de la energía potencial con respecto a la posición:

$$a = -\frac{1}{m} \frac{dV}{dr}$$

Por lo tanto, para calcular una trayectoria, sólo se necesita conocer: las posiciones iniciales de los átomos, la distribución inicial de las velocidades y la aceleración (esta última está determinada por el gradiente de la función de energía potencial). Como se mencionó, las ecuaciones del movimiento son deterministas, por lo que las velocidades y posiciones en el tiempo cero determinan las posiciones y velocidades en cualquier otro momento. Las posiciones iniciales pueden ser obtenidas de estructuras experimentales (ej. estructuras determinadas por cristalografía de rayos X o aquellas obtenidas mediante espectroscopía de resonancia magnética nuclear (NMR), por sus siglas en inglés de Nuclear Magnetic Resonance).

La distribución inicial de las velocidades de los átomos se determina usualmente de una distribución aleatoria de una magnitud “igual” a la temperatura requerida y corregida de manera que no exista momento en el sistema. Por ejemplo:

$$p = \sum_{i=1}^N m_i v_i = 0$$

En este tipo de estudios, las velocidades (v_i) frecuentemente se eligen aleatoriamente de una distribución de Maxwell-Boltzman o de una distribución Gaussiana a una temperatura dada, la

cual define la probabilidad de que el átomo “i” tenga la velocidad “v” en la dirección “x” a la temperatura “T”:

$$p(v_{ix}) = \left(\frac{m_i}{2\pi k_B T} \right)^{\frac{1}{2}} \exp \left[-\frac{1}{2} \frac{m_i v_{ix}^2}{k_B T} \right]$$

Por otra parte, la energía potencial es una función de las posiciones atómicas (en tres dimensiones “3N”) de todos los átomos en el sistema. No existe una solución analítica a las ecuaciones de movimiento, y por lo tanto, deben ser resueltas de manera numérica. Se han generado numerosos algoritmos para integrar las ecuaciones de movimiento (Verlet, Leap-frog, Velocity verlet, Beemans, etc), la elección del algoritmo de integración es una parte muy importante de una simulación de dinámica molecular e idealmente debería cumplir los siguientes criterios:

- El algoritmo debe conservar la energía y el momento
- Ser eficiente computacionalmente
- Permitir pasos largos para la integración

Se puede decir que las simulaciones de dinámica molecular resuelven la ecuación de Newton para el movimiento de un sistema de N átomos que interactúan:

$$m_i \frac{d^2 r_i}{dt^2} = F_i, i = 1 \dots N$$

Las fuerzas son las derivadas negativas de una función de potencial V (r_1, r_2, \dots, r_N):

$$F_i = -\frac{dV}{dr_i}$$

Estas ecuaciones se resuelven simultáneamente para pasos temporales pequeños. El sistema se sigue por algún tiempo, teniendo cuidado de que la temperatura y la presión permanezcan en los

valores requeridos y las coordenadas resultantes se escriben en intervalos regulares a un archivo de salida. Estas coordenadas en función del tiempo, representan la trayectoria del sistema. Después de algunos cambios iniciales el sistema usualmente alcanzará un estado de equilibrio, y promediando sobre una trayectoria al equilibrio se pueden extraer propiedades macroscópicas del sistema.

Es importante considerar que las MD tienen limitaciones, por lo que siempre se deben revisar propiedades experimentales conocidas para valorar la precisión de la simulación (44). A continuación se mencionan algunas de las limitaciones de las MD:

- Las simulaciones utilizan mecánica clásica. Usar la ecuación de Newton implica usar mecánica clásica para describir los movimientos de los átomos. Esto generalmente no representa problema para la mayor parte de los átomos a temperaturas biológicas, pero existen excepciones, por ejemplo, algunas veces los movimientos de los protones son esencialmente cuánticos y esto no puede ser tratado correctamente por la mecánica clásica. Además de realizar simulaciones de dinámica cuántica, este problema se puede tratar de dos maneras: a) realizar simulaciones utilizando osciladores armónicos para los enlaces, ó b) tratar a los enlaces como restricciones en la ecuación de movimiento. La lógica detrás de esto es que un oscilador cuántico en su estado basal se parece más a un enlace restringido que a un oscilador clásico. Una razón práctica para utilizar estas restricciones es que, al remover las frecuencias más rápidas del sistema se pueden utilizar pasos de tiempo más largos para la integración.
- Los electrones se encuentran en el estado basal. En las dinámicas moleculares se utiliza un campo de fuerza que solamente es una función de las posiciones de los átomos. Esto significa que los movimientos electrónicos no se consideran, sino que se supone que “los

electrones ajustan su dinámica de manera instantánea cuando una posición atómica cambia” (Aproximación de Born-Oppenheimer) y permanecen en el estado basal. Por este motivo, en una MD no se consideran los estados electrónicos excitados ni los procesos de transferencia de electrones, y por lo tanto las reacciones químicas no se pueden simular.

- Los campos de fuerza son aproximaciones. Los campos de fuerza proveen las fuerzas y en realidad no son parte del método de la simulación, de manera que, de ser necesario, sus parámetros pueden ser modificados por el usuario.
- Las interacciones de largo-alcance tienen un punto de corte. Se utiliza un punto de corte para las interacciones de Lennard-Jones y algunas veces para las interacciones de Coulomb. Para sistemas grandes los errores generados por esto son pequeños. Sin embargo, para sistemas pequeños, con mucha correlación espacial interna y condiciones periódicas en los límites, los puntos de corte pueden provocar que aumente la correlación interna y por tanto el error.

Introducción

El diseño de proteínas con funciones específicas depende de nuestro entendimiento de la relación entre secuencia, estructura y estabilidad. Dado que las proteínas son moléculas complejas y hay muchos factores que intervienen en su estructura y plegamiento, el diseño de nuevas secuencias se enfrenta a grandes retos.

El aumento en el poder de cómputo ha permitido que hoy en día se realicen estudios *in silico* que hace pocos años se consideraban prohibitivos. Ahora es posible realizar cálculos electrostáticos de complejos macromoleculares o inclusive simulaciones de dinámica molecular (MD) que reflejen el comportamiento dinámico de una proteína solvatada en tiempos de simulación equivalentes al orden de nanosegundos. Por otra parte, durante la última década se han hecho grandes avances en el desarrollo de métodos computacionales que permiten generar secuencias *de novo* a partir de la estructura del esqueleto de una proteína (backbone). Estos avances han abierto el camino para el diseño de proteínas. Al respecto, el programa ROSETTA, creado en el laboratorio del doctor D. Baker (University of Washington, USA), ha tenido un gran éxito en el diseño *de novo* de proteínas (1).

Aunque las proteínas que la naturaleza ha generado y seleccionado desarrollan su papel *in vivo* de manera muy eficiente, se sabe que en términos de estabilidad y velocidad de plegamiento no se encuentran totalmente optimizadas (2-4), lo cual da lugar al refinamiento de dichas propiedades.

La “proteína que une L-lisina, L-arginina y L-ornitina” de *Salmonella typhimurium* (LAO) es una proteína de especificidad múltiple que pertenece a la familia de las “proteínas periplásmicas de unión” de bacterias (PBPs) (5). Estas proteínas forman parte del “sistema de transporte periplásmico” bacteriano, cuya función es justamente transportar una amplia gama de substratos del periplasma al interior de las bacterias (6, 7).

Dada la afinidad nanomolar de LAO por diversos ligandos, su tamaño relativamente pequeño (238 a.a.) y ya que se conocen las estructuras tridimensionales correspondientes a la proteína interactuando con cada uno de sus ligandos, LAO representa un modelo viable para generar diseños del sitio de unión mediante técnicas *in silico*, así como un buen candidato para estudiar las bases de la afinidad múltiple a ligandos.

Transportadores ABC

Los sistemas bacterianos de transporte periplásmico están formados por al menos dos componentes (8):

- 1) Una proteína periplásmica de unión (PBP) que une al ligando,
- 2) Un complejo membranal constituido por dos proteínas integrales de membrana y dos proteínas que hidrolizan ATP.

En estos sistemas, la PBP (que generalmente tiene una alta afinidad por el sustrato) una vez que ha unido al ligando interactúa con el complejo de membrana, desencadenando la activación del mecanismo de translocación del sustrato con el costo energético de una molécula de ATP o GTP (**Figura 1**). Al complejo membranal responsable de este transporte se le conoce como transportador ABC (ATP Binding Cassette) (8).

En resumen, las PBPs tienen dos funciones: unir el sustrato e interactuar con el complejo de membrana. En la literatura se ha sugerido que éstas funciones se encuentran localizadas en distintas regiones de la proteína, ya que en experimentos de mutagénesis dirigida, se han encontrado regiones de la proteína que afectan la capacidad de interactuar con el complejo de membrana sin modificar la capacidad de unir el ligando y viceversa (9-11).

Es importante mencionar que aunque en la familia de las PBPs no se ha encontrado conservación en la estructura primaria, existe una alta conservación en la estructura terciaria, la cual consiste en dos dominios globulares conectados por una “región bisagra” (12, 13).

Las PBPs pueden adoptar al menos dos estados conformacionales: una forma libre de ligando, llamada “estado abierto”; y una forma unida al ligando, llamada “estado cerrado”. El cambio conformacional entre un estado y otro, está dado principalmente por un movimiento en la bisagra que conecta a los dominios (14, 15).

Se han identificado dos clases de PBPs que difieren en la topología de los dominios (16):

clase I) Seis hebras- β en cada dominio,

clase II) Cinco hebras- β en cada dominio.

En ambos casos las hebras- β se encuentran rodeadas por hélices- α .

Se conoce la estructura cristalográfica de más de 100 PBPs (17) y una gran cantidad de ligandos que esta familia puede unir. Los ligandos abarcan un intervalo amplio de formas y tamaños que van desde pequeñas moléculas (ej. iones), hasta moléculas relativamente grandes (ej. disacáridos ó vitaminas).

El sitio de unión de las PBPs presenta dos características particulares:

- 1) Se encuentra en la interfase entre los dos dominios.
- 2) Presenta un gran cambio conformacional al unir el ligando.

En la literatura se ha sugerido que estas características generan propiedades fisicoquímicas que se consideran favorables como son (17):

- a) Posicionamiento del ligando en sitios que excluyen al solvente (y por lo tanto en regiones de una constante dieléctrica baja).

- b) Un gran cambio conformacional permite el desarrollo de diferentes funciones en cada estado conformacional.

En la ingeniería de proteínas se ha utilizado a las PBP para diseñar biosensores, elementos de control alostérico, receptores biológicamente activos e inclusive para el diseño de actividades enzimáticas *de novo* (17, 18).

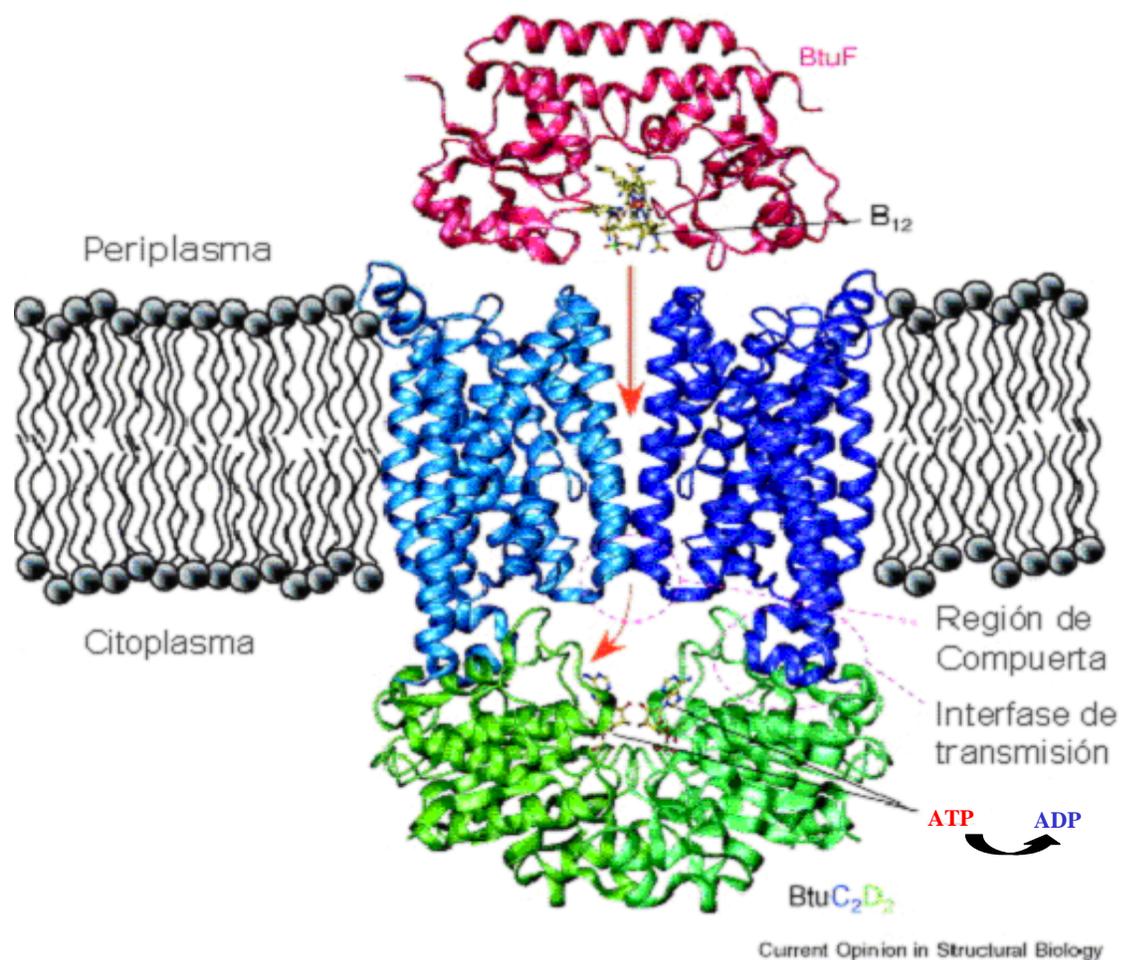


Figura 1. Ejemplificación del mecanismo de acción del transportador ABC de la vitamina B₁₂ de *Escherichia coli*. El transporte ocurre cuando la PBP (BtuF) interactúa con el transportador de membrana (complejo BtuC₂), evento que da lugar a la hidrólisis de una molécula de ATP por el complejo asociado “BtuD₂”, esta hidrólisis aporta la energía necesaria para que el ligando sea translocado al interior de la célula (viajando a través del complejo “BtuC₂”) y finalmente la PBP es liberada del transportador de membrana. *Figura modificada de la referencia 3.*

La proteína periplásmica que une L-lisina, L-arginina y L-ornitina: LAO

LAO es una PBP de 238 aminoácidos con un puente disulfuro entre los residuos 38 y 45, una masa molecular de 26 kDa y una absorbancia de 0.85 ($A_{280}^{0.1\%}$) (5). Su estructura tridimensional sin ligando se encuentra resuelta por cristalografía de rayos-X a 2.7 Å de resolución (10) (Figura 2).

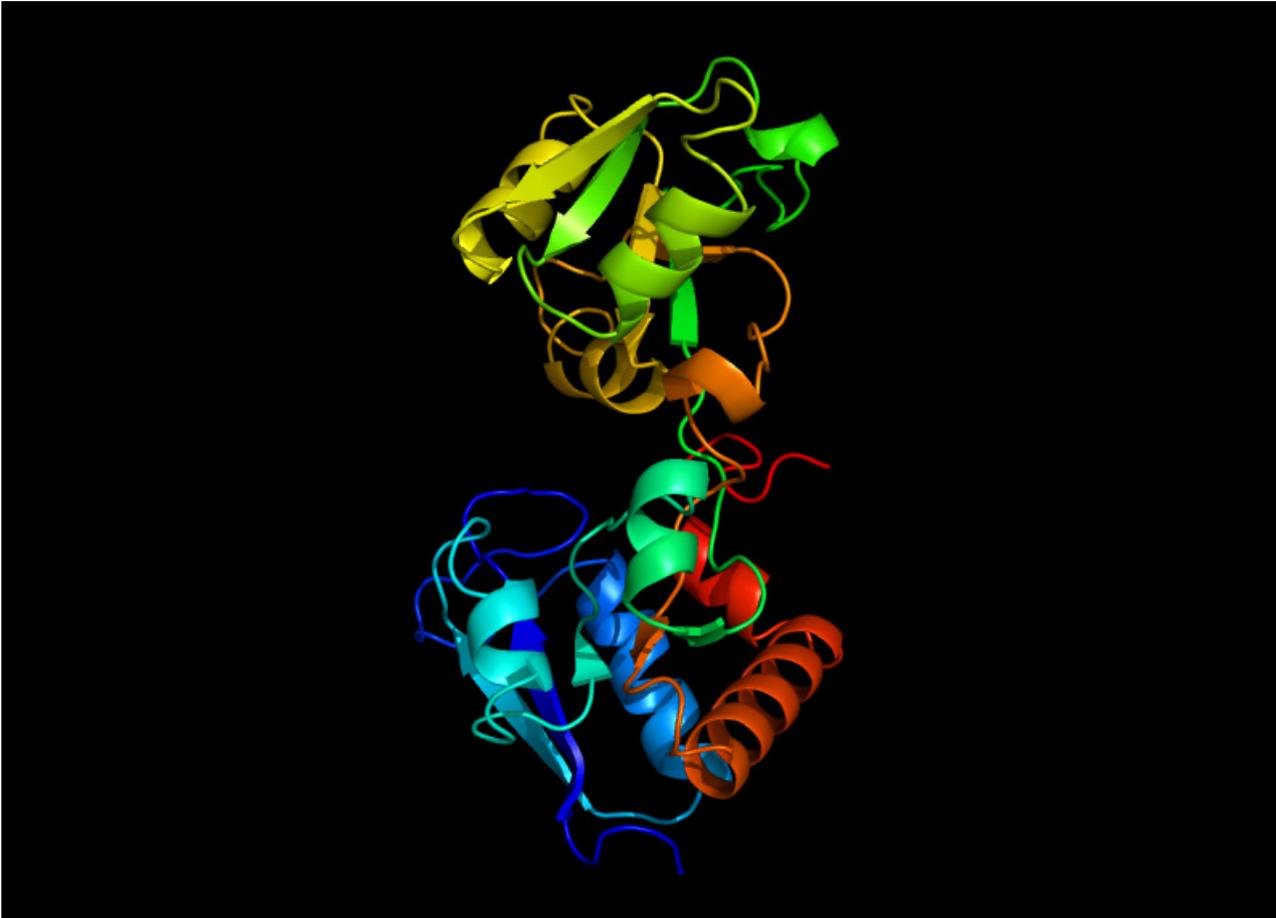


Figura 2. LAO en el estado abierto. Representación: “cartoons” (PDB: 2LAO). En la imagen se aprecia el lóbulo I (dominio inferior de la proteína), el lóbulo II (dominio superior de la proteína) y la región de bisagra que une a los dominios. El extremo amino se encuentra en color azul oscuro y el extremo carboxilo en color rojo.

LAO pertenece a la clase II de las PBP ya que tiene cinco hebras- β y cuatro hélices- α en cada lóbulo, cuatro de estas hebras- β son paralelas y una es antiparalela (en cada lóbulo). El lóbulo I contiene los residuos del extremo NH_2 -terminal 1-88 y a los residuos 195-238 del extremo COOH -terminal, mientras que el lóbulo II contiene los residuos 93 a 185. El COOH -terminal se localiza cerca de la bisagra y el NH_2 -terminal se encuentra en un extremo del lóbulo I. LAO

comparte el transportador de membrana llamado HisQ/M/P₂ con otra proteína periplásmica llamada HisJ (10).

Dentro de la familia de las PBPs LAO es considerada pequeña, ya que el tamaño promedio de los integrantes de esta familia es de 350 a.a. Se ha sugerido que esta diferencia de tamaño se ve reflejada tanto en la topología de la bisagra (que se encuentra formada por sólo dos segmentos polipeptídicos discontinuos en lugar de tres como la mayoría de las PBP), como en el hecho de que LAO carece del “subdominio COOH-terminal” que frecuentemente se encuentra en otras PBPs.

Se conocen cuatro ligandos fisiológicos (todos de alta afinidad) para la LAO: L-arginina, L-lisina, L-ornitina y L-histidina, con constantes de disociación (K_d) de 14, 15, 30 y 500 nM respectivamente (5). En estudios de competencia de ligando se han probado compuestos similares a estos aminoácidos como: ácido L-arginínico, L-glutamina y los enantiómeros-D de los ligandos arginina, lisina y ornitina. Todos han mostrado ser substratos de baja afinidad comparados con los ligandos fisiológicos ($K_d \approx$ “orden mM”) (5). Estos datos indican que en LAO el reconocimiento del ligando está determinado, además de por interacciones polares, por la forma y el tamaño. Los ligandos de LAO comparten la característica de tener carga positiva, sugiriendo una importante contribución electrostática para la unión.

En estudios de fluorimetría se han encontrado diferencias en los espectros de absorción en UV entre la proteína en el estado abierto (sin ligando) y el estado cerrado (con ligando). También se han encontrado diferencias en el espectro de absorción de LAO en presencia de los distintos ligandos. Se ha propuesto que estas diferencias pueden ser explicadas por los distintos estados conformacionales que adopta la proteína al unir sus ligandos (5).

Una ventaja de LAO como modelo teórico es que se conocen las estructuras tridimensionales de la proteína unida a cada uno de los cuatro ligandos de alta afinidad (**Figura 3**) (15). Globalmente

la conformación de la proteína LAO es la misma para las cuatro formas ligadas. Al comparar las estructuras tridimensionales de LAO unida a los distintos ligandos se encuentra que prácticamente no hay diferencias entre ellas (el r.m.s.d. de los C α entre los residuos 5-235 es menor a 0.18 Å) (19). La superposición de las cuatro estructuras con ligando muestra que los residuos involucrados directamente en la unión no difieren en la localización espacial de sus cadenas laterales (Asp-11, Tyr-14, Phe-52, Asp-30, Ser-69, Ser-70, Ser-72, Arg-77, Leu-117, Thr-121 y Asp 161) con la excepción del Asp-11 (19).

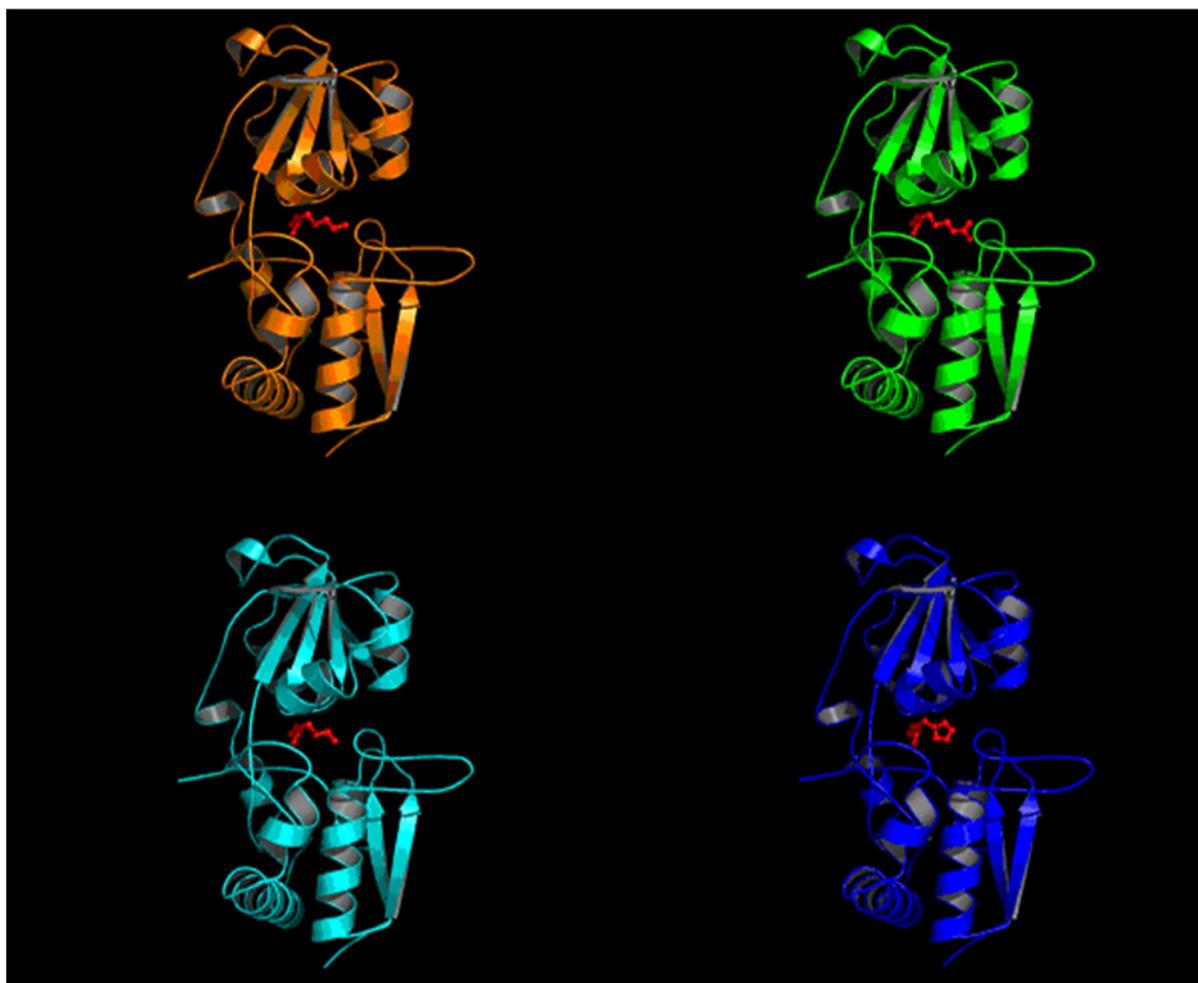


Figura 3. *La proteína LAO en el estado cerrado con sus distintos ligandos.* Representación de listones. En la parte superior izquierda el ligando L-lisina; en la parte superior derecha el ligando L-arginina; en la parte inferior izquierda el ligando L-ornitina; y en la parte inferior derecha el ligando L-histidina.

LAO, en comparación con otras PBP (ej. proteína que une galactosa [GBP], proteína que une arabinosa [ABP] o la proteína que une leu, ile y val [LIV-BP]), tiene una afinidad muy alta por sus ligandos (nanomolar vs milimolar). Esta característica se ha atribuido a las múltiples interacciones iónicas que forma la proteína con sus ligandos (**19**). Pese a los estudios que se han hecho en LAO, aún desconocemos el proceso a través del cual el ligando es reconocido y estabilizado en el sitio de unión de la proteína.

En las diferentes estructuras tridimensionales de LAO se encuentran conservadas dos moléculas de agua en el sitio de unión (Wat 401 y 402) (**19**). También existe una tercera molécula de agua (Wat 403), que participa directamente en la unión de la histidina y se ha sugerido que tiene un papel de neutralización de cargas parciales (**19**). En resumen, la “cavidad” de unión de LAO es lo suficientemente grande para alojar el volumen de cualquiera de los cuatro ligandos mas 2 ó 3 moléculas de agua (Wat 401, 402 y 403).

Diseño de proteínas

El diseño de proteínas por computadora utilizando algoritmos de búsqueda de secuencias y funciones de energía se conoce como “diseño racional de proteínas”, dada su capacidad de generar resultados en términos de unas pocas secuencias que presuntamente contienen las características deseadas (**20**). Para diseñar proteínas que adopten una conformación determinada se requiere evaluar la capacidad de una enorme cantidad de secuencias para adquirir la estructura y función deseadas. Los algoritmos de búsqueda necesariamente hacen un compromiso entre la velocidad de la búsqueda y su precisión (ya que inclusive para un número pequeño de aminoácidos es imposible analizar todo el espacio combinatorio). Se sabe que existen importantes dependencias entre la elección del algoritmo de búsqueda, la forma en que el espacio muestral se representa y las funciones de evaluación utilizadas.

ROSETTA en el diseño de secuencias

Una estrategia para el diseño de proteínas fue descrita por Kuhlman y Baker (21) e implementada en el programa ROSETTA (22, 23). Este algoritmo, al igual que otros basados en métodos de Monte-Carlo, prueba diferentes combinaciones de aminoácidos en las posiciones deseadas de una estructura protéica. La diferencia con otros algoritmos de diseño basados en Monte-Carlo se encuentra principalmente en:

- Las combinaciones de rotámeros que usa ROSETTA, para hacer búsquedas, provienen de bases de datos “jerarquizados” obtenidos de estructuras tridimensionales conocidas (Protein Data Bank, PDB (24)). La idea detrás de esta jerarquización es que con el uso de unos pocos rotámeros que se encuentran con frecuencia en la naturaleza es posible construir diseños exitosos de proteínas. Aunque pareciera que alimentar al algoritmo con estas bases de datos no difiere mucho de otros protocolos propuestos, el simple hecho de trabajar con rotámeros jerarquizados hace que el problema combinatorio sea mucho más pequeño que si se trabajara con todos los rotámeros posibles. De manera práctica, los diseños preliminares se pueden realizar con los rotámeros de la primera jerarquía (es decir los que aparecen con mayor frecuencia). Si se requiere refinar posiciones específicas en el diseño, se trabaja con una base de rotámeros más amplia, pero basándose en el diseño preliminar, lo que permite saber de antemano qué aminoácido o tipo de aminoácido (ej. polar, aromático, etc.) se utilizará en cada posición del refinamiento.
- La mayor bondad del programa ROSETTA no se encuentra sólo en su función de búsqueda, ya que además cuenta con una función semiempírica de evaluación de energía que ha demostrado ser excelente para propósitos de diseño, y probablemente sea la parte

clave del éxito que ROSETTA ha mostrado tanto en predicciones de estructuras terciarias como en el diseño de proteínas.

Con esta aproximación teórica, los autores, partiendo sólo de estructuras (tridimensionales) del esqueleto de proteínas conocidas, han diseñado secuencias de aminoácidos que experimentalmente han mostrado plegarse en estructuras estables (25). La capacidad de ROSETTA en el diseño de proteínas se puso a prueba nuevamente en otro trabajo de los mismos autores, para diseñar una proteína *de novo* con una estructura terciaria que no se ha observado nunca en la naturaleza. Este diseño fue completamente exitoso y dio lugar a la primer proteína de este tipo llamada Top7 (1).

La función de energía utilizada para evaluar los diseños en el algoritmo de ROSETTA es la suma lineal de 11 términos que se encuentran sujetos independientemente a un factor de ponderación (W) (23), el resultado de esta suma es expresado por el programa ROSETTA en un parámetro llamado “bk_tot”:

$$E_{\text{protein}} = W_{\text{rot}}E_{\text{rot}} + W_{\text{aa}|\phi,\psi}E_{\text{aa}|\phi,\psi} + W_{\text{rama}}E_{\text{rama}} + W_{\text{atr}}E_{\text{atr}} + W_{\text{solv}}E_{\text{solv}} + W_{\text{pair}}E_{\text{pair}} + W_{\text{bb_hbond}}E_{\text{bb_hbond}} + W_{\text{sc_hbond}}E_{\text{sc_hbond}} + W_{\text{sc_bb_hbond}}E_{\text{sc_bb_hbond}} + W_{\text{pair}}E_{\text{pair}} - E_{\text{ref}}$$

A continuación se explica la naturaleza de cada término y de los factores de peso:

En la función de energía de ROSETTA, para calcular la energía de solvatación (E_{solv}) y la energía de Lennard-Jones (E_{atr} y E_{rep}), los distintos átomos de los 20 aminoácidos se separan en tipos. Un potencial estándar de Lennard-Jones se utiliza con una modificación que hace que después de una distancia de corte el potencial se extrapole linealmente. Las energías favorables se agrupan en el término E_{atr} y las energías desfavorables en E_{rep} :

$$E_{atr} = \sum_i^{natom} \sum_{j>i}^{natom} \left[\left(\frac{r_{ij}}{d_{ij}} \right)^{12} - 2 \left(\frac{r_{ij}}{d_{ij}} \right)^6 \right] e_{ij} \quad \text{si } \frac{r_{ij}}{d_{ij}} < 1.12$$

$$E_{rep} = \sum_i^{natom} \sum_{j>i}^{natom} \left[\left(\frac{r_{ij}}{d_{ij}} \right)^{12} - 2 \left(\frac{r_{ij}}{d_{ij}} \right)^6 \right] e_{ij} \quad \text{si } 1.33 > \frac{r_{ij}}{d_{ij}} > 1.12 + \sum_i^{noatom} \sum_{j>i}^{natom} y_{intercept} - d_{ij} * slope$$

$$\text{si } \frac{r_{ij}}{d_{ij}} > 1.33 \quad slope = -12e_{if} (1.33^{13} - 1.33^7) * \left(\frac{1}{r_{ij}} \right)$$

$$y_{intercept} = slope * \left(\frac{r_{ij}}{1.33} \right) + e_{if} (1.33^{12} - 2(1.33)^6)$$

$$r_{ij} = r_i + r_j$$

$$e_{ij} = \sqrt{e_i + e_j}$$

Donde: “d” es la distancia de los elementos no-unidos, “r_{ij}” es el diámetro de Lennard-Jones, y “e” es la profundidad del pozo de dispersión.

Para evaluar la energía de solvatación (**E_{solv}**) se usa un modelo de solvatación implícita desarrollado por Lazaridis y Karplus (26):

$$E_{solv} = \sum_i^{natom} \sum_{j>i}^{natom} \left\{ \frac{-2\Delta G_i^{free}}{4\pi\sqrt{\pi}\lambda_i r_{ij}^2} \exp(-d_{ij}^2) V_j - \frac{-2\Delta G_j^{free}}{4\pi\sqrt{\pi}\lambda_j r_{ij}^2} \exp(-d_{ij}^2) V_i \right\}$$

Donde d_{ij} y r_{ij} representan a las mismas variables que en E_{atr}. El término ΔG_{free} está relacionado con la energía de solvatación del átomo completo, λ_i es una medida de correlación y V es el volumen atómico. Los valores de los parámetros se han tomado de Lazaridis y Karplus, con excepción de algunos de los valores de ΔG_{free} que se han perturbado para reproducir mejor las frecuencias relativas de los aminoácidos en el núcleo proteico.

La energía de cada rotámero (**E_{rot}**) se calcula con la siguiente ecuación:

$$E_{rot} = \sum_i^{nres} -\ln(P(rot(i) | phi(i), psi(i)))$$

donde E_{rot} representa la energía interna de un rotámero, derivada de estadísticas del “Protein Data Bank” (PDB), en las que se observa la probabilidad de un rotámero/aminoácido en particular para un ángulo phi o psi dado (27).

$E_{aa|phi,psi}$ se refiere a las preferencias de los aminoácidos por regiones particulares del espacio phi o psi. Para determinar este componente, los autores utilizaron un grupo no-redundante de archivos PDB para determinar las probabilidades de observar cada uno de los 20 aminoácidos en un espacio phi, psi ($P(aa|phi,psi)$). La energía se calcula como el $(-\log)$ de las probabilidades.

E_{rama} es el potencial de torsión dependiente de phi y psi de los aminoácidos. Para este término se ha determinado la frecuencia de los pares phi-psi.

E_{pair} está dado por la probabilidad de encontrar a dos aminoácidos a una cierta distancia en el espacio. Los datos se derivaron de la base de datos PDB y el término se evalúa para aminoácidos polares considerando dos ambientes: aminoácidos atrapados en un ambiente no polar y aminoácidos expuestos. La ecuación utilizada es:

$$E_{pair} = \sum_i^{nres} \sum_{j>i}^{nres} \frac{P(aa_i, aa_{jj} | d_{ij}, env_i, env_j)}{P(aa_i, aa_{jj} | d_{ij}, env_i)P(aa_j | d_{ij}, env_j)}$$

E_{bb_hbond} , E_{sc_hbond} , $E_{bb_sc_hbon}$. En cada una de estas contribuciones energéticas se suma un término de puentes de hidrógeno dependiente de la orientación: esqueleto $\leftarrow \rightarrow$ esqueleto; cadena lateral $\leftarrow \rightarrow$ esqueleto; y cadena lateral $\leftarrow \rightarrow$ cadena lateral, respectivamente, que están evaluados a partir de los ángulos y las distancias observadas en las proteínas nativas.

E_{ref} es la energía de referencia con respecto al estado desplegado de la proteína y está dada por:

$$E_{ref} = \sum_i^{nres} W_{ref}(aa(i))$$

Para calcular esta función, a cada aminoácido se le asigna una energía de referencia calculada empíricamente.

Funciones de peso W. Los pesos para los términos y las 20 energías de referencia se determinaron por la maximización del producto: $\exp(-E(aaobs))/\exp(-E(aai))$ sobre un grupo de referencia de 30 proteínas, usando un método de optimización de gradiente-conjugado, donde $E(aaobs)$ es la energía del aminoácido nativo en una posición dada y la función de partición en el denominador es sobre los 20 aminoácidos en cada posición. En este proceso un residuo fue cambiado a la vez y los otros residuos se mantuvieron fijos en la conformación nativa. Subsecuentemente los parámetros se refinaron basándose en los cálculos del diseño de las proteínas del grupo de referencia.

En una versión reciente del programa ROSETTA (ROSETTALIGAND) (28), los autores han generado una nueva función de energía para la evaluación de las interacciones “proteína \leftrightarrow ligando”. Partiendo de la idea de que la función de energía para evaluar ligandos es correcta, los autores han propuesto que, “utilizando esta nueva función de energía en conjunto con la misma filosofía de diseño ya existente en ROSETTA”, en teoría, ROSETTA sería capaz de diseñar (*de novo*) sitios proteicos de unión a ligandos. En la práctica, esta nueva función de energía para pequeños ligandos es usada al mismo tiempo que la función de energía para el diseño de proteínas, de manera que al mismo tiempo que ROSETTA diseña el sitio de unión, se asegura de no afectar la estabilidad de la estructura proteica. Esta nueva función de energía para ligandos (lig_sum) está integrada por la suma de cinco componentes (28):

- 1) Un potencial de Lennard Jones 12-6 estándar para modelar interacciones de atracción con pozos de dispersión y radios de van der Waals obtenidos del campo de fuerza CHARMM27.

- 2) Un término de repulsión que conecta la amplitud y la pendiente con el potencial $12-6$ a $E = 0$, y que después es lineal hasta que dos átomos se encuentran a 0\AA de distancia.
- 3) Un término de solvatación similar al de Lazaridis–Karplus con un modelo de solvatación implícita.
- 4) Un potencial explícito de puentes de hidrógeno.
- 5) Un modelo de Coulomb con una constante dieléctrica dependiente de la distancia y que usa las cargas parciales del campo de fuerza CHARMM27.

En ésta función, los factores de peso de cada término fueron calculados por los autores con base en grupos de entrenamiento.

Como se puede ver, la función de ROSETTA para ligandos es aditiva por pares y es mucho más simple que la función utilizada para el diseño de proteínas, por lo que se puede calcular muy rápidamente.

Docking

Localizar el sitio de unión de un ligando en una estructura proteica es un problema complejo que ha sido abordado por el método computacional conocido como Docking (anclaje molecular simulado) (29). El problema del Docking se compone esencialmente de dos partes: 1) buscar las conformaciones espaciales y con características fisicoquímicas favorables para la interacción entre un receptor y su ligando, y 2) estimar la energética de la interacción. Ambas partes se encuentran relacionadas en los algoritmos de Docking, ya que estos métodos dependen de una función que evalúa las conformaciones espaciales, al mismo tiempo que calcula la energética de la interacción, y de esta manera determina cuales conformaciones tienen mayor probabilidad de representar un complejo biológicamente significativo.

Se han desarrollado diversos algoritmos de búsqueda por Docking que comprenden desde los algoritmos de fuerza bruta (intentar explorar “todas” las combinatorias de conformaciones y posibles sitios de unión), hasta otros menos demandantes computacionalmente, como algoritmos genéticos y de Monte-Carlo. También existen distintos métodos de evaluación de los complejos (“scoring functions”) y se clasifican principalmente en: 1) basados en campos de fuerza explícitos, y 2) semi-empíricos. Los primeros estiman la energía de unión a partir de campos de fuerza como CHARMM o AMBER, mientras que las funciones semi-empíricas se basan en la suma ponderada de diversos tipos de interacciones como: la formación de puentes de hidrógeno, efectos de solvatación, etc. Los pesos relativos se derivan mediante ajustes de datos de “entrenamiento” con afinidades experimentales conocidas.

Para el caso del diseño del sitio de unión de una proteína, probablemente la función de evaluación más relevante es la de generar una predicción de la afinidad de unión. Aunque en los últimos años se han realizado avances en la precisión de la predicción por docking, los métodos aún no son tan confiables como se desearía (30).

Electrostática

Identificar las interacciones responsables de la unión específica de ligandos es un requisito esencial para el diseño de proteínas. La unión de ligandos está dada por una gran cantidad de factores, entre los que se encuentran: la geometría del sitio de unión, interacciones polares, interacciones hidrofóbicas y energías de solvatación.

El proceso de unión ligando-proteína puede ser descrito en 3 etapas consecutivas:

- Difusión del ligando hasta la superficie proteica.
- Difusión del ligando a través de la proteína hasta el sitio de unión.
- Rearreglo del ligando y del sitio de unión hacia la orientación adecuada.

Aunque la fuerza motriz para este proceso es generalmente el efecto hidrofóbico, las interacciones electrostáticas también son importantes, tanto para los residuos polares como para los apolares (31). En el caso de la PBP se puede definir una cuarta etapa asociada al cambio conformacional de la proteína.

La electrostática juega un papel crítico en la estructura de las proteínas y su función. No obstante, para propósitos de diseño no existe un consenso acerca de cómo se deben modelar estas interacciones. Un problema es que tratar de calcular estas interacciones con la ecuación de Poisson-Boltzmann y al mismo tiempo utilizarla dentro de una estrategia de diseño combinatorio no es factible, debido a la complejidad misma que se requiere para llegar a la convergencia con esta ecuación. Como alternativa, aproximaciones simplificadas, como reducir las constantes dieléctricas dependientes de la distancia, se han usado en algunas estrategias de diseño (32, 33). Se piensa que una función efectiva para evaluar estas interacciones se puede obtener a través de las interacciones de Coulomb (34), y una práctica frecuente es imponer un punto de corte para la contribución máxima de las interacciones electrostáticas que tiene como resultado un menor tiempo de cálculo (35).

En general, las contribuciones energéticas de las interacciones electrostáticas en los procesos de plegamiento y asociación resultan de un balance entre dos efectos separados: 1) una contribución desestabilizante al proceso de asociación debido a la pérdida de interacciones favorables de solvatación entre los residuos cargados o polares y la fase acuosa altamente dieléctrica, y 2) un cambio en las interacciones de los pares de Coulomb como resultado del plegamiento o de la asociación de dos cuerpos. El signo y la magnitud de estos efectos dependen de las posiciones de las cargas que interactúan. Los cálculos basados en electrostática continua sugieren que, para los pares aislados de los grupos polares y los pares ionizados que están interactuando, las interacciones de Coulomb generalmente no son lo suficientemente intensas para compensar el

efecto desfavorable de la desolvatación (36). Entonces, se espera que la contribución de las interacciones electrostáticas a un proceso de ensamblaje sea desfavorable. De lo anterior se deduce que se puede esperar una gran contribución de este efecto desfavorable mientras haya más aminoácidos cargados y grupos polares en la superficie de unión. Esto da lugar a la pregunta ¿de dónde viene la energía que lleva a la formación de los complejos? En ciertos estudios se ha encontrado que, en efecto, las interacciones electrostáticas tienden a oponerse a la formación de los complejos, mientras que la fuerza que promueve la unión proviene principalmente de las interacciones no polares (36). Sin embargo, Lee y Tidor han hecho observaciones que indican que la interfase de la proteína Barstar está electrostáticamente optimizada para interactuar con la Barnasa, aunque esto no implica que todas las interfases proteicas estén optimizadas de la misma manera (37).

Los cálculos electrostáticos generalmente suponen casos hipotéticos donde no existe cambio conformacional alguno después de la unión, es decir, los cálculos se llevan a cabo suponiendo que la estructura de los monómeros, o del monómero y el ligando, son iguales tanto en el complejo como en la forma libre (36). Aunque sin duda esta observación es una gran simplificación lo que sucede *in vivo*, ofrece una aproximación efectiva para la solución del problema.

La energía electrostática total de un sistema macromolecular se define como el trabajo necesario para ensamblar todas sus cargas y se calcula como (36):

$$\Delta G_{elec}^{todo} = \frac{1}{2} \sum_{i=1}^N q_i \phi(r_i)$$

En esta ecuación q_i es la carga del cuerpo i , $\phi(r_i)$ es el valor del potencial electrostático en la posición del cuerpo i , y la suma se lleva a cabo sobre todas las cargas atómicas parciales de un sistema molecular. La expresión dada en la ecuación anterior corresponde a la energía libre de un

sistema en relación a una estructura de forma idéntica pero sin cargas en él. Esta expresión puede utilizarse para definir la contribución electrostática de los procesos de ensamblaje macromolecular. Para una asociación proteína-proteína (o proteína-ligando) (36):

$$\Delta G_{elec}^{todo} = \frac{1}{2} \sum_{i \in A:B} q_i \phi^{complejo}(r_i) - \frac{1}{2} \sum_{i \in A} q_i \phi^{libre}(r_i) - \frac{1}{2} \sum_{i \in B} q_i \phi^{libre}(r_i)$$

La suma en cada uno de los tres términos de esta ecuación se realiza por separado para las cargas parciales en las estructuras: A, B y el complejo A:B. Con la aproximación de un sistema macromolecular en una conformación fija y en un estado de protonación dado, la ecuación describe todos los efectos electrostáticos que contribuyen a la unión del complejo, incluyendo las interacciones intramoleculares de Coulomb y las interacciones electrostáticas del solvente modificado tras la unión.

Cuando la forma lineal de la “ecuación de Poisson-Boltzmann” (PBE) se usa para calcular los potenciales electrostáticos, el potencial en cualquier punto puede expresarse como la suma de las contribuciones de las cargas individuales (i.e. $s_j(r_i) = \sum_{j=1}^n q_j / r_{ij}$, donde $s_j(r_i)$ es el potencial electrostático en la posición i debido a la carga q_j).

Existen pocas soluciones analíticas de la PBE para las geometrías de las biomoléculas. Sin embargo, esta ecuación se resuelve numéricamente por distintos métodos computacionales, los cuales transforman la solución continua en un conjunto dimensional-finito de funciones básicas. En el caso de la forma lineal de la PBE, las ecuaciones resultantes se pueden arreglar en la forma de matriz-vector y ser resueltas directamente. Pero las ecuaciones no lineales obtenidas de la ecuación completa de PBE requieren de técnicas más especializadas como los métodos de Newton (38).

Dinámica Molecular

El término “modelado molecular” indica el proceso general de describir “sistemas químicos complejos” en términos de un “modelo atómico realista”, con el objetivo de entender o predecir propiedades macroscópicas basándose en un conocimiento detallado a escala microscópica (atómica).

La dinámica molecular (MD) es un método que fue introducido alrededor en 1957 por Alder y Wainwright (Alder and Wainwright, (39, 40)) para estudiar las interacciones de esferas sólidas. El siguiente avance crucial se dio en 1964, cuando Rahman realizó la primera dinámica molecular utilizando un potencial realista para argón líquido (41). Después, en 1974, Stillinger y Rahman realizaron la primera MD de un “sistema realista” compuesto por 216 moléculas de agua en una simulación equivalente a 8 ps (42). Posteriormente, en 1977, McCammon et al. reportaron la primer simulación de MD para una proteína: “el inhibidor pancreático de tripsina bobina (BPTI)”, con una integración temporal de 10 ps (43). En los siguientes 30 años los algoritmos y técnicas computacionales que se utilizan para realizar dinámicas moleculares han mejorado considerablemente. Esto, aunado al creciente poder de cómputo, ha permitido que hoy en día se realicen simulaciones de proteínas, pequeños complejos proteicos e inclusive complejos de membrana en solventes explícitos, con integraciones temporales de nanosegundos.

Las simulaciones de dinámica molecular generan información de nivel microscópico de las propiedades de un sistema, incluyendo las posiciones de los átomos y la velocidad con que estos se mueven. La conversión de esta información en propiedades macroscópicas como la presión, la energía, la capacidad calorífica, etc, requiere del empleo de la ‘mecánica estadística’ (44).

Las propiedades macroscópicas se pueden dividir en:

- a) Propiedades de equilibrio (ej. constantes de unión, la energía potencial promedio)

- b) Propiedades dinámicas o de no-equilibrio (ej. viscosidad de un líquido, cinéticas de reacción, procesos de difusión)

La elección de la técnica para calcular las propiedades dinámicas de un sistema depende de la pregunta que se desea responder y de la capacidad del método para dar resultados útiles. Idealmente, la ecuación de Schrödinger describe las propiedades de sistemas moleculares con una gran certeza, pero cualquier cosa más compleja que un sistema en equilibrio con pocos átomos no puede ser manejado a nivel *ab-initio*. Por lo tanto, se requieren aproximaciones al problema. Mientras más complejo es el sistema y más largo el tiempo de integración que se quiere estudiar, se requieren aproximaciones más reduccionistas, y en algún momento la complejidad es tal que la aproximación *ab-initio* debe ser aumentada o remplazada por una parametrización empírica del modelo usado.

Las propiedades macroscópicas son promedios de un ensamble estadísticamente representativo de un sistema molecular (ya sea en equilibrio o no). Para el modelado molecular esto tiene dos consecuencias importantes (44):

- 1) Conocer una sola estructura, aunque ésta estructura corresponda al mínimo de energía global, no es suficiente para calcular propiedades macroscópicas, por lo que es necesario generar un ensamble representativo a una temperatura dada para poder calcular estas propiedades. Sin embargo, esto no es suficiente para calcular las propiedades termodinámicas en el equilibrio basadas en términos de energía libre, como son: equilibrios de fase, constantes de unión, solubilidades, estabilidades de conformaciones moleculares, entre otras. Para calcular las energías libres y los potenciales termodinámicos se requiere una extensión de las técnicas de simulación molecular.

- 2) Mientras que las simulaciones de dinámica molecular (en principio) proveen detalles atómicos de las estructuras y de sus movimientos, frecuentemente estos detalles no son relevantes para las propiedades macroscópicas de interés, de manera que se abre un camino para simplificar la descripción de las interacciones y realizar promedios sobre detalles relevantes. La mecánica estadística provee un andamio teórico para tales simplificaciones, y existen una gran variedad de métodos que van desde considerar grupos de átomos como una unidad, hasta dinámicas mesoscópicas que describen densidades en lugar de átomos; flujos que responden a gradientes termodinámicos en lugar de velocidades ó aceleraciones que responden a fuerzas.

Para generar un ensamble representativo del equilibrio existen al menos dos métodos:

- 1) Simulaciones de Monte-Carlo (MC)
- 2) Simulaciones de Dinámica Molecular (MD)

Mientras que para generar ensambles que no se encuentren al equilibrio sólo se pueden utilizar las dinámicas moleculares. Por otra parte, a pesar de que en las simulaciones de Monte-Carlo no se requiere el cálculo de fuerzas y, por lo tanto, son más sencillas que las MD (en términos de cálculos), en general con las simulaciones de Monte-Carlo no son capaces de obtener mejores resultados que con las MD en el mismo tiempo de cómputo.

Cuando se realizan simulaciones de MD frecuentemente se desea hacer una extrapolación a las propiedades macroscópicas del sistema. La mecánica estadística provee de relaciones matemáticas que relacionan las propiedades macroscópicas con la distribución y movimiento de los átomos de un sistema. Este sistema puede ser desde uno simple, como una colección de moléculas de solvente, hasta un sistema complejo como una o varias proteínas solvatadas. Con el objetivo de relacionar el comportamiento microscópico de un sistema con el macroscópico, se requiere hacer relaciones dependientes con el tiempo.

El estado termodinámico de un sistema se define usualmente por un número limitado de parámetros, por ejemplo: la temperatura (T), la presión (P) y el número de partículas (N).

El estado mecánico o estado microscópico se encuentra definido por las posiciones atómicas (p) y el momento (q). Estas dos “propiedades” también pueden ser consideradas como coordenadas en un espacio multidimensional llamado “espacio de fase”.

Las simulaciones de dinámica molecular se basan en la segunda ley de Newton (ecuación del movimiento): $F = ma$, donde “F” es la fuerza aplicada a una partícula, “m” es la masa y “a” es la aceleración. Conociendo la fuerza en cada átomo en el sistema es posible determinar la aceleración de cada átomo. La integración de la ecuación del movimiento da lugar a una trayectoria que describe la posición, velocidad y aceleración de las partículas del sistema a lo largo del tiempo. El método es determinista, es decir, una vez que se conocen las velocidades y posiciones de cada átomo, en principio, el estado del sistema se puede predecir en cualquier momento en el pasado o en el futuro, aunque puede ser muy costoso computacionalmente. Hoy en día se realizan simulaciones de dinámica molecular de proteínas solvatadas con tiempos de integración de nanosegundos, e inclusive existen casos en los que se han reportado tiempos de integración en la escala de los milisegundos.

Si tomamos la ecuación del movimiento de Newton:

$$F_i = m_i a_i = m \frac{d^2 x}{dt^2}$$

F es la fuerza aplicada a una partícula “i”, “m” es la masa de la partícula “i” y “a” es la aceleración de la partícula “i”. La fuerza también se puede expresar como el gradiente de la energía potencial:

$$F_i = -\nabla_i V = -\frac{dV}{dr_i}$$

combinando estas dos ecuaciones tenemos

$$-\frac{dV}{dr_i} = m_i \frac{d^2 r_i}{dt^2}$$

donde “V” es la energía potencial del sistema y “r” es la posición. Se observa que la ecuación del movimiento puede relacionar la derivada de la energía potencial con respecto a la posición, con los cambios en las posiciones de los átomos en función del tiempo.

$$F = ma = \frac{dV}{dr} = m \frac{d^2 x}{dt^2}$$

Si se toma un caso sencillo en el que la aceleración es constante:

$$a = \frac{dv}{dt}$$

Se obtiene una expresión para la velocidad tras la integración:

$$v = at + v_0$$

y dado que

$$v = \frac{dx}{dt}$$

se puede integrar nuevamente para obtener:

$$x = vt + x_0$$

Si se combina esta ecuación con la expresión para la velocidad obtenemos la siguiente relación que nos da el valor de “x” en el tiempo “t” como una función de la aceleración, la posición inicial “x₀” y la velocidad inicial “v₀”:

$$x = at^2 + v_0t + x_0$$

La aceleración está dada por la derivada de la energía potencial con respecto a la posición:

$$a = -\frac{1}{m} \frac{dV}{dr}$$

Por lo tanto, para calcular una trayectoria, sólo se necesita conocer: las posiciones iniciales de los átomos, la distribución inicial de las velocidades y la aceleración (esta última está determinada por el gradiente de la función de energía potencial). Como se mencionó, las ecuaciones del movimiento son deterministas, por lo que las velocidades y posiciones en el tiempo cero determinan las posiciones y velocidades en cualquier otro momento. Las posiciones iniciales pueden ser obtenidas de estructuras experimentales (ej. estructuras determinadas por cristalografía de rayos X o aquellas obtenidas mediante espectroscopía de resonancia magnética nuclear (NMR), por sus siglas en inglés de Nuclear Magnetic Resonance).

La distribución inicial de las velocidades de los átomos se determina usualmente de una distribución aleatoria de una magnitud “igual” a la temperatura requerida y corregida de manera que no exista momento en el sistema. Por ejemplo:

$$p = \sum_{i=1}^N m_i v_i = 0$$

En este tipo de estudios, las velocidades (v_i) frecuentemente se eligen aleatoriamente de una distribución de Maxwell-Boltzman o de una distribución Gaussiana a una temperatura dada, la

cual define la probabilidad de que el átomo “i” tenga la velocidad “v” en la dirección “x” a la temperatura “T”:

$$p(v_{ix}) = \left(\frac{m_i}{2\pi k_B T} \right)^{\frac{1}{2}} \exp \left[-\frac{1}{2} \frac{m_i v_{ix}^2}{k_B T} \right]$$

Por otra parte, la energía potencial es una función de las posiciones atómicas (en tres dimensiones “3N”) de todos los átomos en el sistema. No existe una solución analítica a las ecuaciones de movimiento, y por lo tanto, deben ser resueltas de manera numérica. Se han generado numerosos algoritmos para integrar las ecuaciones de movimiento (Verlet, Leap-frog, Velocity verlet, Beemans, etc), la elección del algoritmo de integración es una parte muy importante de una simulación de dinámica molecular e idealmente debería cumplir los siguientes criterios:

- El algoritmo debe conservar la energía y el momento
- Ser eficiente computacionalmente
- Permitir pasos largos para la integración

Se puede decir que las simulaciones de dinámica molecular resuelven la ecuación de Newton para el movimiento de un sistema de N átomos que interactúan:

$$m_i \frac{d^2 r_i}{dt^2} = F_i, i = 1 \dots N$$

Las fuerzas son las derivadas negativas de una función de potencial V (r₁, r₂, ..., r_N):

$$F_i = - \frac{dV}{dr_i}$$

Estas ecuaciones se resuelven simultáneamente para pasos temporales pequeños. El sistema se sigue por algún tiempo, teniendo cuidado de que la temperatura y la presión permanezcan en los

valores requeridos y las coordenadas resultantes se escriben en intervalos regulares a un archivo de salida. Estas coordenadas en función del tiempo, representan la trayectoria del sistema. Después de algunos cambios iniciales el sistema usualmente alcanzará un estado de equilibrio, y promediando sobre una trayectoria al equilibrio se pueden extraer propiedades macroscópicas del sistema.

Es importante considerar que las MD tienen limitaciones, por lo que siempre se deben revisar propiedades experimentales conocidas para valorar la precisión de la simulación (44). A continuación se mencionan algunas de las limitaciones de las MD:

- Las simulaciones utilizan mecánica clásica. Usar la ecuación de Newton implica usar mecánica clásica para describir los movimientos de los átomos. Esto generalmente no representa problema para la mayor parte de los átomos a temperaturas biológicas, pero existen excepciones, por ejemplo, algunas veces los movimientos de los protones son esencialmente cuánticos y esto no puede ser tratado correctamente por la mecánica clásica. Además de realizar simulaciones de dinámica cuántica, este problema se puede tratar de dos maneras: a) realizar simulaciones utilizando osciladores armónicos para los enlaces, ó b) tratar a los enlaces como restricciones en la ecuación de movimiento. La lógica detrás de esto es que un oscilador cuántico en su estado basal se parece más a un enlace restringido que a un oscilador clásico. Una razón práctica para utilizar estas restricciones es que, al remover las frecuencias más rápidas del sistema se pueden utilizar pasos de tiempo más largos para la integración.
- Los electrones se encuentran en el estado basal. En las dinámicas moleculares se utiliza un campo de fuerza que solamente es una función de las posiciones de los átomos. Esto significa que los movimientos electrónicos no se consideran, sino que se supone que “los

electrones ajustan su dinámica de manera instantánea cuando una posición atómica cambia” (Aproximación de Born-Oppenheimer) y permanecen en el estado basal. Por este motivo, en una MD no se consideran los estados electrónicos excitados ni los procesos de transferencia de electrones, y por lo tanto las reacciones químicas no se pueden simular.

- Los campos de fuerza son aproximaciones. Los campos de fuerza proveen las fuerzas y en realidad no son parte del método de la simulación, de manera que, de ser necesario, sus parámetros pueden ser modificados por el usuario.
- Las interacciones de largo-alcance tienen un punto de corte. Se utiliza un punto de corte para las interacciones de Lennard-Jones y algunas veces para las interacciones de Coulomb. Para sistemas grandes los errores generados por esto son pequeños. Sin embargo, para sistemas pequeños, con mucha correlación espacial interna y condiciones periódicas en los límites, los puntos de corte pueden provocar que aumente la correlación interna y por tanto el error.

Hipótesis

Las proteínas de la familia de las PBPs presentan un plegamiento consenso, sin embargo tienen una gran variabilidad en su secuencia primaria. Esto sugiere que el sitio de unión de las PBPs es “flexible” para permitir modificaciones, por lo cual proponemos que, mediante técnicas de modelado *in silico* se puede modificar el sitio de unión de LAO para unir específicamente ligandos distintos a los fisiológicos, conservando el mecanismo de acción de la proteína.

Objetivos

General

Generar, mediante técnicas *in silico*, mutantes de la proteína LAO del organismo *Salmonella typhimurium* que unan aminoácidos aromáticos, así como desarrollar una metodología computacional para evaluar estas mutantes.

Particulares

- Diseñar el sitio de unión de LAO mediante el programa ROSETTA, con el fin de generar mutantes que unan aminoácidos aromáticos.
- Establecer una metodología *in silico* para evaluar los diseños generados.

Metodología

Estructuras proteicas

Se usaron cuatro estructuras cristalinas de la proteína LAO de *Salmonella typhimurium*: en el estado abierto [PDB: 2LAO], y en el estado cerrado con los ligandos arginina, lisina e histidina [PDBs: 1LAF, 1LST y 1LAG respectivamente] (19). Todas las estructuras se encuentran disponibles en el RCSB Protein Data Bank (24). 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 (45).

Al no existir la estructura de la proteína LAO en el estado abierto con los ligandos arginina, lisina e histidina en el sitio de unión, se construyeron dichas estructuras “artificiales” a partir de la estructuras conocidas de la siguiente manera: ya que la estructura del esqueleto del lóbulo que contiene a los extremos amino y carboxilo terminal (lóbulo 2) es casi idéntica en los estados abierto y cerrado de la proteína, usando el programa PyMol (versión 0.99rc6) (46) se realizó un alineamiento de estos lóbulos 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-LIS y 2LAO-HIS, respectivamente); 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 dichos impedimentos estéricos de acuerdo con los radios de van der Waals (VDW). Para la estandarización del estado cerrado, en la estructura 1LAG se

reemplazó el ligando histidina por arginina o lisina, mediante un simple alineamiento de las estructuras correspondientes (1LAG vs 1LAF o 1LST).

En el caso de los cálculos electrostáticos, los átomos de hidrógeno se construyeron mediante el programa en línea PDB2PQR (47) utilizando el campo de fuerza de CHARMM. El pH de protonación al que fueron sometidas las estructuras fue de 7, mientras que, en el caso de las simulaciones por dinámica molecular y el diseño por ROSETTA, se permitió que cada uno de los programas protonara independientemente a la estructuras ignorando cualquier estado de protonación previo. En el caso de las simulaciones por dinámica molecular se obligó a que la única histidina presente (HIS-113) se encontrara en su forma protonada (carga = +1e).

Para los ligandos “no-fisiológicos”: tirosina, fenilalanina y triptofano, se utilizó la metodología de docking descrita a continuación, y para el estado abierto se usó un procedimiento análogo al de los ligandos fisiológicos (a partir de las estructuras obtenidas por docking).

Docking

Las coordenadas de LAO (1LAG) se tomaron de la base de datos del PDB. Para asegurarnos de la calidad de las estructuras (ausencia de residuos incompletos) todos los PDBs que se usaron fueron procesados en el servidor en línea de WHATIF (48). Las moléculas de agua y los hidrógenos se eliminaron con el programa PyMol. Para preparar los archivos para el docking se utilizó el ambiente gráfico llamado Autodock Tools (ADT). A la proteína se le agregaron protones polares. Las cargas para la proteína son las de Gasteiger mientras que para el ligando son las de Kollman. Al ligando se le permitió rotar todos los enlaces (químicamente permitidos). La rejilla de búsqueda se centró en el sitio de unión al ligando y se extendió al máximo permitido con una resolución de 0.35 Å (x 126 puntos en cada eje). Para la búsqueda se utilizó el programa AUTODOCK (versión 3.0) (49), utilizando un algoritmo genético lamarckiano con los siguientes

parámetros: población inicial de 300 individuos, 20 millones de evaluaciones de energía, 30 mil generaciones, elitismo de 1, y un total de 100 corridas.

Diseño computacional

En el diseño de la proteína LAO se utilizaron las estructuras cristalográficas (PDB) de la proteína sin ligando (2LAO) y con los ligandos arginina (1LAF), lisina (1LST) e histidina (1LAG). Además, se utilizaron las estructuras cristalográficas de la PBP que une leucina, isoleucina y valina (LivJ PDBs: 1Z17, 1Z16 y 1Z18) y la de la PBP que une maltosa (1ANF)

El diseño del sitio de unión y las MDs se realizaron a partir de la estructura cristalográfica 1LAG (LAO en el “estado cerrado” unida al ligando histidina).

Con el fin de diseñar el sitio de unión de LAO se empleó el programa AUTODOCK (versión 3.0), para unir a 1LAG los ligandos L-tirosina ó L-fenilalanina ó L-triptofano ó serotonina. Se seleccionaron los resultados en los que la interacción de los ligandos resultó más favorable energéticamente, y posteriormente se eligieron aquellas estructuras donde el ligando, además de encontrarse en la cavidad de unión de la proteína, tuviera una orientación semejante a la del aminoácido histidina (con excepción de la serotonina, en la que simplemente se eligieron las estructuras de menor energía)

Usando una versión modificada del programa ROSETTA (versión “en desarrollo” del programa ROSETTA, proveída por el Dr. David Baker, University of Washington, publicación pendiente), se re-empacaron las estructuras de docking, con el fin de tener como punto de partida aquellas estructuras que tuvieran los mejores parámetros de interacción proteína-ligando de acuerdo con el programa ROSETTA. Este proceso de re-empacamiento consiste esencialmente en que, sin modificar la estructura del esqueleto proteico, se trata de encontrar aquella combinación de rotámeros de la secuencia proteica en la cual la función de energía de ROSETTA obtiene los

valores de interacción más favorables (tanto para el lig_sum, como para el bk_tot). Es decir que el re-empacamiento es un proceso parecido a una minimización de energía, pero en este caso se encuentra en relación con funciones de evaluación de ROSETTA y su algoritmo de diseño (i.e. en lugar de realizarse átomo por átomo se realiza mediante combinaciones de rotámetros).

Las estructuras resultantes se diseñaron empleando el programa ROSETTA en su modalidad de “diseño de sitios de unión para ligandos” con una aproximación de esqueleto proteico fijo. Adicionalmente se empleó el parámetro de perturbación de la posición del ligando:”dok_pert” con valores: $n=0.5$, $p=0.5$ y $r=5$. Donde “n” es la perturbación normal en Å (en el plano de la línea de los centros), “p” es la perturbación paralela en Å (en el plano perpendicular a la línea de los centros) y “r” es la rotación expresada en grados en ambos planos. Finalmente, en algunos de los mejores diseños (según los parámetros internos bk_tot y lig_sum del programa ROSETTA) se evaluó la contribución electrostática de la unión del ligando.

Cálculos electrostáticos

Los potenciales electrostáticos se calcularon solucionando numéricamente la ecuación lineal de Poisson-Boltzmann mediante el programa APBS versión 0.3.2 (50, 51).

La constante se definió como 78 para el solvente y 2 para la proteína; el límite entre el solvente y la proteína se definió por los radios de VDW de los átomos de la superficie; la fuerza iónica del solvente se asignó como cero; la temperatura se fijó en 298.15 K.

Los potenciales electrostáticos se calcularon usando cubos rectangulares de 84.98 X 78.25 X 106.2 Å. Para cada proteína se calculó primero el potencial electrostático con una malla gruesa de 84.98 X 78.25 X 106.22 Å. Posteriormente, a partir de la malla gruesa se obtuvieron las condiciones del límite para una malla fina de 35 X 35 X 35 Å, y se calculó el potencial

electrostático de los diferentes archivos de estructura (enfoque electrostático). La malla gruesa siempre se centró en el complejo proteína ligando y la malla fina en el ligando.

El ΔG electrostático se calculó de la siguiente manera:

$$\Delta G^{\text{elec}} = G^{\text{elec}}_{\text{holoproteína}} - G^{\text{elec}}_{\text{apoproteína}} - G^{\text{elec}}_{\text{ligando}}$$

y el ΔG de Coulomb se calculó:

$$\Delta G^{\text{coul}} = G^{\text{coul}}_{\text{holoproteína}} - G^{\text{coul}}_{\text{apoproteína}} - G^{\text{coul}}_{\text{ligando}}$$

donde la G^{coul} es el resultado del cálculo de la energía electrostática de la proteína en el vacío (es decir, con una constante dieléctica del solvente igual a la de la proteína).

Simulaciones de dinámica molecular

Las MDs se realizaron con el programa GROMACS (ver. 3.3.1) (44, 52, 53). El procedimiento general para cada MD consistió en:

- Minimizar la estructura de la proteína LAO mediante un algoritmo de paso descendente, durante el número necesario de pasos para que el cambio máximo en la energía fuera menor a $100 \text{ kJ mol}^{-1} \text{ nm}^{-1}$.
- Solvatar la proteína en una caja dodecaédrica de un tamaño de 375.79 nm^3 que se llenó con moléculas de solvente preequilibradas (agua, modelo scp216), con un total de moléculas de solvente (HOH) ≈ 11300 . El tamaño de la caja es suficiente para que el límite se encuentre por lo menos a 1 nm de distancia de cualquier átomo de la proteína LAO.
- Minimizar nuevamente la estructura usando el mismo criterio que en la primera minimización.

- Realizar una “MD de posiciones restringidas” para relajar el solvente durante 250 ps, es decir, mantener fijo el esqueleto de la proteína y permitir el libre movimiento de las moléculas de agua durante la simulación.
- Finalmente se liberaron las restricciones del esqueleto de la proteína y se realizó la MD de producción con un tiempo de integración equivalente a 20 ns.

Las velocidades de los átomos de la proteína se generaron siempre antes de comenzar la MD de posiciones restringidas y la MD de producción. Las velocidades se generaron de acuerdo con una distribución de Maxwell con una temperatura de referencia $T_{ref} = 318$ K. En el caso de la “MD de posiciones restringidas” se usó la semilla 173529. Para las distintas repeticiones de las MDs de producción se utilizaron las semillas: 173529, 823139 y 511825.

Los parámetros de las MDs de “posiciones restringidas” y de producción fueron: campo de fuerza GROMOS_53a6; algoritmo “md”; baño térmico a 318 K; celdas periódicas; electrostática PME; presión constante de 1 atm; temperatura acoplada por un termostato de Berendsen; presión acoplada por un presostato de Berendsen de manera isotrópica; compresibilidad del sistema $4.5e^{-5}$ bar; enlaces restringidos por LINCS; sistema con carga total de cero; átomos de hidrógeno “dummies” (sitios virtuales de interacción); pasos de integración de 5 fs.

El análisis del estado de apertura y/o cierre de LAO se realizó mediante la medición de tres ángulos diedros distintos, definidos por los átomos correspondientes a los carbonos- α de los aminoácidos (**figura 7**). El primero de estos ángulos diedros (diedro-L1) está formado por los aminoácidos: [141, 90, 192, 13] y el segundo (diedro-L2) por los aminoácidos [53, 192, 90, 144]. Adicionalmente se definió un ángulo diedro para la bisagra, el cual refleja el movimiento lateral de un lóbulo con respecto a otro, y que se encuentra definido por los carbonos- α de los aminoácidos: [11, 88, 91, 142].

Visualización

La visualización de los archivos de estructura PDB y PQR, así como de los archivos correspondientes a los mapas de potencial electrostático, se realizó mediante el programa PyMol versión 0.99rc6.

La visualización de las trayectorias obtenidas de las simulaciones de dinámica molecular se realizó mediante el programa “Visual Molecular Dynamics” (VMD, versión 1.8.5) (54).

Recursos computacionales

Los cálculos del programa ROSETTA se realizaron con un procesador de dos núcleos “Intel core2 duo”, velocidad de 2.4 Ghz, bajo el sistema operativo “Fedora Core 3”, 2 Gb de RAM; y con otro sistema análogo, sólo que con procesador “Athlon 64 X2 4200+”, velocidad de 2.2 Ghz.

Los cálculos de dinámica molecular se realizaron inicialmente en un cluster híbrido ATHLON/INTEL 32/64bits de 8 procesadores, bajo los sistemas operativos Fedora Core 3 o 5 y un total de 3.5 Gb de RAM. Los procesos se ejecutaron de forma paralela, utilizando 2 procesadores por simulación.

La mayor parte de las simulaciones de producción por MD se realizaron en la supercomputadora KanBalam de la Universidad Nacional Autónoma de México (UNAM, DGSCA), que cuenta con procesadores Opteron a una velocidad de 2.6 Ghz, (4 procesadores y 4 Gb de RAM, por nodo).

En esta supercomputadora se paralelizaron los procesos utilizando entre 4 y 16 procesadores para cada simulación

Resultados y discusión

Evaluación de las interacciones PBP-ligando fisiológico por ROSETTA

En el pasado se han reportado los parámetros experimentales de interacción de varias PBP, así como de algunos diseños exitosos de estas proteínas. Con el fin de determinar si ROSETTA es capaz de evaluar correctamente la unión de ligandos fisiológicos y la estabilidad de las proteínas silvestres, se eligieron algunas estructuras cristalográficas de PBPs para ser evaluadas por este algoritmo.

La PBP que une leucina, isoleucina y valina, del organismo *Escherichia coli* (*LivJ*), se evaluó para sus tres ligandos de alta afinidad ($K_d = 200\text{nM}$) (55). Estos cálculos se realizaron a partir del re-empacamiento de la estructura original de la proteína. Todos los resultados evaluados en términos de *lig_sum* (parámetro de afinidad por ligandos del programa ROSETTA) y *bk_tot* (parámetro de estabilidad de la proteína) mostraron ser favorables (valores numéricos negativos) (Tabla 1).

	<i>lig_sum</i>	<i>bk_tot</i>
LivJ-isoleucina	-7.512	-1227.407
LivJ-leucina	-14.542	-1223.629
LivJ-valina	-14.662	-1163.257

Tabla 1. Valores de *lig_sum* y *bk_tot* para *LivJ*. Cada resultado en esta tabla, muestra el promedio de 10 estructuras de re-empacamiento por ROSETTA (PDBs: 1Z17, 1Z16 y 1Z18, respectivamente).

La PBP que une maltosa (MBP, PDB: 1ANF) del organismo *Escherichia coli*, ha sido caracterizada extensivamente por varios grupos de investigación ($K_d = 800\text{ nM}$) (56). Al someter esta proteína a la misma evaluación realizada para *LivJ*, contrariamente a lo esperado los valores obtenidos fueron más favorables que los encontrados para *LivJ*, promedio de 10 estructuras: (*lig_sum* = -19.75, *bk_tot* = -882.54).

Por otra parte, en un estudio hecho por el grupo de Hellinga, se propuso que la afinidad de la MBP puede ser modificada al favorecer el estado cerrado de la proteína (57), es decir que, si se desplaza el equilibrio de la proteína hacia el estado cerrado se obtendrá mayor afinidad por los ligandos. A este respecto, se han generado varias mutaciones de la Ile-329 (residuo que se encuentra en la bisagra) para tratar de favorecer el estado cerrado a través de “estorbar” el movimiento de la bisagra y de esta manera aumentar la afinidad de la proteína. Con el fin de evaluar la capacidad de ROSETTA para evaluar diferencias en la estabilidad (tanto proteica como de interacción con el ligando) de las proteínas correspondientes a las distintas mutaciones, se pidió al programa ROSETTA en su modalidad de diseño de ligandos, que mutara la posición Ile-329 por los residuos estudiados en el trabajo de Hellinga, y se analizó tanto el bk_{tot} como el lig_{sum} de estas mutaciones (**Tabla 2**).

Los resultados obtenidos para bk_{tot} no guardan ninguna correlación con las K_d 's experimentales, de igual manera sucede con el lig_{sum} . Esto podría significar que el papel de estabilización de estas mutaciones se encuentra más bien en un nivel dinámico, y por lo tanto no se puede evaluar su impacto en la afinidad estudiando sólo una estructura cristalográfica. Es posible que sea más apropiado estudiar el efecto de este tipo de mutaciones mediante el uso de dinámicas moleculares, a través de las cuales se podrían analizar parámetros como la velocidad de apertura y cierre de la proteína, los tiempos de residencia del ligando o inclusive la energética de la interacción.

Mut (Ile-329)	lig_sum	bk_tot	K_d(nM)
ILE (WT)	-19.78	-882.54	800
ALA	-20.08	-886.11	1800
ARG	-20.00	-887.52	840
CYS	-19.55	-883.14	660
PHE	-19.86	-882.89	180
TYR	-20.09	-886.94	103
TRP	-20.04	-886.37	35

Tabla 2. Valores de **lig_sum** y **bk_tot** para MBP wt y mutantes. Cada resultado en esta tabla, muestra el promedio de 10 estructuras de re-empacamiento por ROSETTA (PDBs: 1Z17, 1Z16 y 1Z18, respectivamente).

En otro estudio de la MBP, realizado en el grupo de Shilton, mediante la eliminación de ciertos residuos se logró aumentar la afinidad de la MBP ($K_d = 110$ nM, PDB:1N3W) (58). El análisis de esta estructura con ROSETTA muestra un **lig_sum** menos favorable que el de la MBP-WT (promedio de 10 estructuras: **lig_sum** = -11.91) pero un **bk_tot** más favorable (promedio de 10 estructuras: **bk_tot** = -1070.57). Estos resultados también son contradictorios con respecto a las constantes de afinidad y podrían estar relacionados con eventos dinámicos de la proteína.

Se probó con ROSETTA otro diseño de una MBP (PDB: 1ZMG) que se sabe une zinc ($K_d = 0.29\mu\text{M}$) (59). El resultado obtenido para esta proteína fue desfavorable (promedio de 10 estructuras: **lig_sum** = 1.34, **bk_tot** = -1026.05).

Finalmente, se estudiaron a través de ROSETTA las estructuras cristalográficas de LAO que corresponden a la proteína con sus ligandos fisiológicos arginina, lisina e histidina (PDBs: 1LAF, 1LST y 1LAG, respectivamente). Así como, reemplazando al ligando histidina por los ligandos arginina y lisina, en la estructura 1LAG (a partir del alineamiento de los $C\alpha$ de las estructuras cristalográficas). De esta manera se determinó que es factible reemplazar a los ligandos en una sola estructura cristalográfica del estado cerrado (1LAG) y obtener resultados comparables

(**Tabla 3**). Como se esperaba, el ligando arginina siempre obtiene los valores más favorables de lig_sum. Sin embargo los ligandos lisina e histidina muestran valores de lig_sum muy parecidos, lo cual no es congruente con las afinidades experimentales reportadas, ya que existe más de orden de magnitud de diferencia entre éstos. En cuanto al número de resultados de re-empaquetamiento que se pidieron a ROSETTA, se puede observar que para LAO, no existe una diferencia significativa entre pedir 10 o 100 estructuras.

Estructura de LAO	Ligando	Kd (nm)	lig_sum		bk_tot	
			10 repeticiones	100 repeticiones	10 repeticiones	100 repeticiones
1LAF	arg	14	-21.91 ± 0.28	-22.04 ± 0.63	-859.35 ± 0.58	-858.87 ± 1.05
1LAG		*	-21.62 ± 0.10	-21.65 ± 0.12	-842.29 ± 0.68	-842.34 ± 0.85
1LST	lys	15	-15.62 ± 0.31	-15.70 ± 0.27	-869.29 ± 0.48	-869.49 ± 0.36
1LAG		*	-15.14 ± 0.23	-15.17 ± 0.22	-838.31 ± 0.40	-837.82 ± 0.89
1LAG	hsa	500	-13.52 ± 0.04	-13.50 ± 0.03	-835.70 ± 0.59	-835.50 ± 0.55
1LAG	hsb		-15.28 ± 0.06	-15.27 ± 0.04	-836.93 ± 0.63	-837.17 ± 0.73
1LAG	hsh		-15.15 ± 0.04	-15.17 ± 0.06	-837.39 ± 0.42	-837.10 ± 0.64

Tabla 3. Valores de lig_sum y bk_tot para LAO con sus ligandos fisiológicos: arg, lys e his (en sus tres formas tautoméricas: hsa, hsb y hsh). Se muestran los resultados de cada ligando en su estructura cristalográfica correspondiente, e insertados artificialmente en la estructura correspondiente a 1LAG. Cada resultado en esta tabla, muestra el promedio de 10 o 100 estructuras de re-empaquetamiento por ROSETTA.

En resumen, los resultados obtenidos muestran que ROSETTA, en general, calcula valores lig_sum negativos para ligandos fisiológicos de algunas PBPs. Sin embargo, arroja valores desfavorables al evaluar una proteína diseñada para un ligando muy pequeño (Zinc). En cuanto a la estabilidad de las proteínas, ROSETTA es capaz de evaluar como “estables” a las estructuras silvestres de diversas PBPs.

Es importante señalar que aunque los valores de lig_sum son favorables, no guardan ninguna relación con la afinidad de la proteína. Esto puede deberse a una combinación de la parametrización de la función de energía que evalúa las interacciones en ROSETTA y a que en las PBPs el carácter dinámico de la unión del ligando podría jugar un papel importante en la afinidad y con ROSETTA somos incapaces de evaluar el impacto de la dinámica de una proteína.

De estas observaciones concluimos que el parámetro *lig_sum*, en términos generales, si bien habla de la interacción entre la proteína y el ligando, sólo puede ser utilizado como parámetro interno de referencia para el diseño de una proteína, y no como una correlación con la afinidad. El *bk_tot* para las PBP estudiadas parece oscilar entre valores de “-800 y -1300”, que se pueden utilizar como valores de referencia para la estabilidad de otras PBP’s.

Diseño de LAO para la unión de aminoácidos aromáticos

Usando la estructura proteica correspondiente a LAO unida al ligando histidina (1LAG) (pero sin el ligando), se utilizó el programa AUTODOCK para unir, por separado, los ligandos no fisiológicos: fenilalanina, tirosina, triptofano y serotonina. En general los complejos que mostraron las energías más favorables correspondieron al ligando dentro de la cavidad de unión de la proteína. La orientación del carboxilo y del nitrógeno del carbono- α (extremo amino-carboxilo), así como del anillo de los ligandos, no siempre correspondió con la orientación de estos elementos en el ligando original de 1LAG (histidina).

En cada caso se eligieron al menos dos estructuras provenientes del docking para comenzar un diseño con ROSETTA: el complejo de menor energía de interacción, y la estructura en que el extremo amino-carboxilo y el anillo aromático correspondieran a la misma orientación en que se encuentran en el ligando histidina en la estructura 1LAG. Tales estructuras siempre se encontraron dentro del grupo de los primeros seis mejores resultados del primer grupo de resultados de AUTODOCK.

A partir de estas estructuras se generaron los distintos diseños del sitio de unión de LAO para los ligandos deseados mediante el programa ROSETTA.

Diseño LAO-Fenilalanina

En el caso de la fenilalanina se escogieron tres estructuras provenientes del proceso de docking (dock1-PHE, dock2-PHE y dock3-PHE), y a partir de estas estructuras se re-empacó la molécula mediante ROSETTA para obtener como punto de partida la estructura de menor energía (mayor estabilidad) de interacción con el ligando (**Tabla 4**).

	lig_sum	bk_tot
PHE- R1 -repack	-11.92	-830.23
PHE- R2 -repack	21.29	-799.03
PHE- R3 -repack	-3.29	-821.43

Tabla 4. Valores de **lig_sum** y **bk_tot** después del re-empacamiento para los distintos docks del ligando fenilalanina a LAO. Cada resultado en esta tabla muestra el promedio de 10 estructuras de re-empacamiento por ROSETTA. Los interfijos R1, R2 y R3 corresponden a los diseños realizados a partir de las estructuras dock1-PHE, dock2-PHE y dock3-PHE, respectivamente.

En dos casos los resultados del re-empacamiento mostraron una interacción favorable (1LAG-PHE-R1-repack y PHE-R3-repack), mientras que el otro caso (PHE-R2-repack) mostró un resultado desfavorable.

Con el fin de obtener diseños de LAO que unan al ligando fenilalanina, se sometió a las estructuras del re-empacamiento a un proceso de “diseño del sitio de unión” mediante ROSETTA. En los resultados se buscó que los diseños obtuvieran mejores valores de **lig_sum** que las estructuras iniciales. A partir de este proceso se obtuvieron 100 estructuras resultantes. En la **tabla 5** se sintetizan los 5 diseños con los mejores parámetros de interacción para cada grupo de diseño.

# Res Res Ori # Mut	14 TYR	17 PHE	70 SER	72 SER	117 LEU	120 SER	122 GLN	124 ALA	143 GLN	160 GLN	161 ASP	164 ALA	lig_sum	bk_tot	Id Estructura Silvestre
5	ASN	***	***	***	VAL	***	ASP	***	***	THR	***	SER	-20.19	-849.93	PHE- R1-78
6	ASN	***	***	ASP	VAL	***	GLU	***	***	THR	***	SER	-20.08	-849.45	PHE- R1-85
5	ASN	***	***	***	VAL	***	ASP	***	***	THR	***	SER	-20.06	-849.31	PHE- R1-84
6	ASN	HIS	***	***	ASN	***	***	***	THR	THR	***	GLN	-20.05	-852.1	PHE- R1-91
5	ASN	***	***	***	ASN	***	ASP	***	THR	THR	***	***	-19.82	-849.64	PHE- R1-81
6	ASN	***	GLY	***	ILE	***	***	GLU	GLU	THR	***	***	-17.86	-843.85	PHE- R2-55
6	ASN	***	GLY	***	ILE	***	***	GLU	GLU	THR	***	***	-17.71	-843.97	PHE- R2-92
8	ASN	***	***	***	ILE	GLY	ASP	GLU	GLU	THR	***	SER	-17.67	-844.42	PHE- R2-32
6	ASN	***	GLY	***	ILE	***	***	GLU	GLU	THR	***	***	-17.63	-843.7	PHE- R2-24
6	ASN	***	GLY	***	ILE	***	***	GLU	GLU	THR	***	***	-17.58	-843.65	PHE- R2-14
8	ASN	***	GLY	ASP	ILE	***	GLU	***	GLU	THR	THR	***	-19.88	-841.77	PHE- R3-74
8	ASN	***	GLY	ASP	ILE	***	GLU	***	GLU	THR	THR	***	-19.49	-841.65	PHE- R3-52
7	ASN	***	GLY	ASP	ILE	***	GLU	***	GLU	THR	***	***	-19.33	-840.64	PHE- R3-95
7	ASN	***	GLY	ASP	ILE	***	GLU	***	GLU	THR	***	***	-19.07	-841.5	PHE- R3-93
8	ASN	***	GLY	ASP	ILE	***	GLU	***	GLU	THR	THR	***	-19.03	-842.28	PHE- R3-82

Tabla 5. Valores de lig_sum y bk_tot del diseño de LAO para unir el ligando fenilalanina. Se muestran los cinco mejores diseños de cada grupo.

En todos los casos ROSETTA fue capaz de generar diseños en los que disminuyeron tanto el lig_sum como el bk_tot (de cada grupo). Es notable que algunos diseños correspondan a las mismas combinaciones de mutaciones pero arrojan valores ligeramente distintos de lig_sum. Esto se debe a que, con el objetivo de encontrar los mejores parámetros de interacción de cada mutante, se empleó en el diseño una opción que rota y mueve “ligeramente” el ligando dentro del sitio de unión (dock_pert; ver métodos), ya que se sabe que perturbar ligeramente la posición del ligando permite algunas veces encontrar mejores valores de interacción en ROSETTA. Por este motivo, en las diferentes estructuras resultantes el ligando se puede encontrar ligeramente movido y presentar diferentes valores de lig_sum con la misma combinación de mutaciones.

En trabajos anteriores (60) hemos determinado que la contribución electrostática al proceso de unión y difusión de ligandos en LAO podría jugar un papel determinante en la afinidad. Por esto, para determinar las contribuciones electrostáticas de las distintas modificaciones a la unión del ligando, los diseños se sometieron a un análisis con el programa APBS. Para realizar este análisis

se tomaron las primeras 4 estructuras de menor energía en cada diseño (con excepción de las estructuras correspondientes al dock1 en el que sólo se escogieron 3 estructuras), cuidando que cada una de estas estructuras tuviera distintas mutaciones. Además, se analizaron como parámetros de referencia las estructuras provenientes del proceso de docking antes y después del re-empacamiento (**Tabla 6**).

Estos resultados muestran dos aspectos importantes:

- 1) Que las estructuras provenientes del “docking” no se encuentran optimizadas electrostáticamente para unir este ligando, ya que tras el re-empacamiento las energías electrostáticas de interacción mejoran. Lo cual evidencia que el re-empacamiento generado por ROSETTA es favorable para las contribuciones electrostáticas.
- 2) Que en general los diseños obtenidos pueden llegar a tener un lig_sum mejorado y energías electrostáticas favorecidas.

estructura	ΔG^{elec} (kcal/mol)	$\Delta G^{\text{coulomb}}$ (kcal/mol)	lig_sum
PHE- R1	0.11	-65.73	***
PHE- R1- repack	-6.04	-73.66	-11.92
PHE- R1- 78	-3.70	-60.56	-20.19
PHE- R1- 84	-3.86	-62.03	-20.06
PHE- R1- 85	-13.55	-89.34	-20.08
PHE- R2	0.11	-69.16	***
PHE- R2- repack	-7.73	-76.43	21.29
PHE- R2- 32	-0.46	-72.04	-17.67
PHE- R2- 34	-6.60	-63.40	-17.49
PHE- R2- 55	-7.83	-78.45	-17.86
PHE- R2- 91	-7.76	-77.65	-17.4
PHE- R3	1.29	-62.94	***
PHE- R3- repack	-3.01	-67.62	-3.29
PHE- R3- 73	-3.65	-63.38	-18.93
PHE- R3- 74	-10.17	-86.22	-19.88
PHE- R3- 77	-9.81	-83.91	-18.84
PHE- R3- 95	-5.27	-80.14	-19.33

Tabla 6. Valores de ΔG^{elec} , $\Delta G^{\text{coulomb}}$ y lig_sum de algunos diseños de LAO para unir el ligando fenilalanina. Como referencia se muestran los valores para las estructuras de cada grupo, antes y después del re-empacamiento.

Diseño LAO-Tirosina

Con el fin de diseñar a LAO para unir el ligando tirosina se siguió la misma estrategia empleada con la fenilalanina, con la diferencia de que únicamente se tomaron dos estructuras provenientes del docking (dock1-TYR y dock2-TYR). Posteriormente estas estructuras se re-empacaron con ROSETTA para obtener las estructuras iniciales (**Tabla 7**).

	lig_sum	bk_tot
TYR -R1-repack	-8.83	-826.13
TYR -R2-repack	-11	-826.63

Tabla 7. Valores de lig_sum y bk_tot después del re-empacamiento para los distintos docks del ligando tirosina a LAO. Cada resultado en esta tabla muestra el promedio de 10 estructuras de re-empacamiento por ROSETTA. Los interfijos R1 y R2 corresponden a los re-empacamientos realizados a partir de las estructuras dock1-TYR y dock2-TYR, respectivamente.

Los resultados mostraron un re-empacamiento favorable para la tirosina en ambos casos. A partir de estas estructuras se generaron 100 diseños del sitio de unión de LAO para unir a la tirosina y, al igual que en el diseño de la fenilalanina, se buscaron los diseños de menor energía en los parámetros lig_sum y bk_tot (**Tabla 8**).

# Res	14	52	70	72	90	117	121	122	143	160	161	164	lig_sum	bk_tot	id Estructura
Res	TYR	PHE	SER	SER	ALA	LEU	THR	GLN	GLN	GLN	ASP	ALA			Silvestre
# Mut															
5	ASN	***	***	***	***	ASP	***	GLU	ASP	THR	***	***	-26.35	-847.56	TYR-R1-62
6	ASN	***	***	ASP	***	ASP	***	GLU	ASN	THR	***	***	-25.67	-847.75	TYR-R1-89
7	ASN	THR	***	ASP	***	ASP	***	GLU	GLU	THR	***	***	-25.46	-848.37	TYR-R1-83
6	THR	***	***	ASP	***	ASP	***	GLU	ASP	THR	***	***	-25.38	-845.36	TYR-R1-72
6	THR	***	***	ASP	***	ASP	***	GLU	ASP	THR	***	***	-25.33	-845.11	TYR-R1-99
6	ASP	GLU	GLY	***	***	ILE	***	***	***	THR	***	SER	-22.88	-843.56	TYR-R2-84
5	CYS	***	GLY	***	ASP	***	***	***	***	GLY	PRO	***	-22.46	-840.43	TYR-R2-95
6	CYS	***	GLY	***	ASP	***	***	THR	***	GLY	PRO	***	-22.23	-839.47	TYR-R2-41
6	THR	***	GLY	***	***	ILE	***	ASP	***	THR	***	SER	-22	-839.61	TYR-R2-60
8	CYS	***	GLY	***	ASP	***	***	GLU	ARG	GLY	PRO	ASP	-21.86	-839.41	TYR-R2-96

Tabla 8. Valores de lig_sum y bk_tot después del diseño de LAO para unir el ligando tirosina. Se muestran los cinco mejores diseños de cada grupo.

De manera similar al ligando fenilalanina, los valores de lig_sum y bk_tot mejoraron notablemente para la tirosina después del diseño. De estos resultados se tomaron las cuatro

estructuras correspondientes a los valores más favorables de cada diseño (sin repetición de combinaciones mutaciones) y se analizaron electrostáticamente (**Tabla 9**).

Nuevamente los cálculos electrostáticos muestran que la estructura original no se encuentra optimizada electrostáticamente para unir a este ligando. El simple re-empacamiento mejora considerablemente la energía electrostática, pero en el caso del dock1 no es capaz de generar una interacción con energía electrostática favorable.

estructura	ΔG^{elec} (kcal/mol)	$\Delta G^{\text{coulomb}}$ (kcal/mol)	lig_sum
TYR -R1	8.32	-57.22	***
TYR -R1-repack	0.11	-64.61	-8.83
TYR -R1-11	-7.09	-79.42	-24.79
TYR -R1-62	-3.02	-69.80	-26.35
TYR -R1-83	-8.01	-81.30	-25.46
TYR -R1-89	-7.54	-81.13	-25.67
TYR -R2	3.03	-63.39	***
TYR -R2-repack	-7.82	-77.38	-11
TYR -R2-41	-6.27	-76.57	-22.23
TYR -R2-60	-4.43	-58.68	-22.00
TYR -R2-84	-7.54	-78.46	-22.00
TYR -R2-95	-5.70	-79.31	-22.46

Tabla 9. Valores de ΔG^{elec} , $\Delta G^{\text{coulomb}}$ y lig_sum de algunos diseños de LAO para unir el ligando tirosina. Como referencia se muestran los valores para las estructuras de cada grupo antes y después del re-empacamiento.

Diseño LAO-triptofano

Mediante la misma estrategia descrita para la fenilalanina y la tirosina, se diseñó la proteína LAO para unir el ligando triptofano. La **tabla 10** muestra los valores de referencia obtenidos tras el re-empacamiento de las estructuras del docking (dock1-TRP y dock6-TRP).

	lig_sum	bk_tot
TRP-R1-repack	-3.44	-816.48
TRP-R6-repack	10.71	-806.05

Tabla 10. Valores de lig_sum y bk_tot después del re-empacamiento para los distintos docks del ligando triptofano a LAO. Cada resultado en esta tabla muestra el promedio de 10 estructuras de re-empacamiento por ROSETTA. Los interfijos R1 y R6 corresponden a los re-empacamientos realizados a partir de las estructuras dock1-TRP y dock6-TRP, respectivamente.

Tras el re-empacamiento una de las estructuras de “docking” (dock6-TRP) no logró alcanzar un valor favorable de lig_sum. Con estas estructuras como punto de partida se realizó el diseño del sitio de unión de LAO para unir al ligando triptofano. En el caso del diseño correspondiente a TRP-R1-repack se analizaron 100 estructuras de diseño, mientras que para TRP-R6-repack sólo se analizaron 63. Los datos de las 5 estructuras con los mejores parámetros de interacción obtenidos se muestran en la **tabla 11**.

# Res	14	52	70	71	72	90	117	120	122	124	143	160	161	164	lig_sum	bk_tot	Estructura
Res Original	TYR	PHE	SER	LEU	SER	ALA	LEU	SER	GLN	ALA	GLN	GLN	ASP	ALA			Silvestre
# Mut																	
5	***	***	***	***	ASP	***	GLY	GLY	ASP	GLU	***	***	***	***	-21.51	-837.2	TRP-R1-64
7	ASN	***	GLY	***	***	***	ASP	***	GLU	GLU	GLU	THR	***	***	-20.93	-841.95	TRP-R1-55
7	ASN	***	***	***	***	***	THR	***	THR	GLU	GLU	THR	***	SER	-20.73	-848.63	TRP-R1-99
6	ASN	***	***	***	***	***	THR	***	THR	GLU	GLU	THR	***	***	-20.45	-847.02	TRP-R1-90
7	ASN	***	***	***	***	***	THR	***	ASP	GLU	GLU	THR	***	SER	-20.4	-849.18	TRP-R1-26
8	ILE	HIS	***	***	***	ASP	ASP	***	GLU	***	***	THR	GLY	SER	-17.5	-838.75	TRP-R6-62
7	ILE	HIS	***	***	***	***	ASP	***	ASP	***	***	THR	THR	SER	-17.48	-839.69	TRP-R6-41
8	ILE	HIS	***	***	***	SER	ASP	***	THR	***	***	THR	SER	SER	-17.2	-840.41	TRP-R6-35
8	ASN	HIS	***	ASP	***	***	ASP	***	ASP	***	GLU	THR	***	SER	-16.97	-837.45	TRP-R6-50
8	ILE	HIS	***	***	***	SER	ASP	***	ASP	***	***	THR	SER	SER	-16.81	-840.56	TRP-R6-52

Tabla 11. Valores de lig_sum y bk_tot después del diseño de LAO para unir el ligando triptofano. Se muestran los cinco mejores diseños de cada grupo.

Los resultados muestran que los valores de lig_sum y bk_tot mejoraron para el diseño del ligando triptofano. Al igual que con los diseños anteriores se tomaron las cuatro estructuras correspondientes a los valores de lig_sum más favorables de cada diseño y se analizaron electrostáticamente (**Tabla 12**).

Todos los diseños correspondientes al grupo TRP-R1 mostraron mejoría en sus valores de energía electrostática después del diseño. En el caso del grupo TRP-R6 sólo uno de los diseños alcanzó un valor negativo de ΔG^{elec} .

estructura	ΔG^{elec} (kcal/mol)	$\Delta G^{\text{coulomb}}$ (kcal/mol)	lig_sum
TRP-R1	-1.66	-70.21	***
TRP-R1- repack	1.90	-51.48	-3.44
TRP-R1- 55	-4.70	-75.62	-20.93
TRP-R1- 64	-7.26	-66.51	-21.51
TRP-R1- 90	-4.51	-71.47	-20.45
TRP-R1- 99	-4.48	-71.92	-20.73
TRP-R6	22.01	-31.19	***
TRP-R6- repack	7.075	-47.07	10.71
TRP-R6- 35	17.09	-19.80	-17.2
TRP-R6- 41	14.67	-31.41	-17.48
TRP-R6- 50	17.90	-29.96	-16.97
TRP-R6- 62	-1.33	-51.02	-17.5

Tabla 12. Valores de ΔG^{elec} , $\Delta G^{\text{coulomb}}$ y lig_sum de algunos diseños de LAO para unir el ligando triptofano. Como referencia se muestran los valores para las estructuras de cada grupo antes y después del re-empacamiento.

Diseño LAO-serotonina

En esta serie de experimentos se intentó diseñar un sitio de unión para la serotonina, utilizando la misma estrategia que para los ligandos anteriores. Ya que para el análisis electrostático se requiere un campo de fuerza del cual se puedan tomar las cargas y los radios de los átomos, los diseños generados no fueron evaluados electrostáticamente. Estos campos de fuerza se encuentran disponibles (públicamente) para todos los aminoácidos, pero en el caso de otros compuestos es necesario usar técnicas de mecánica cuántica para estimar las cargas parciales de los átomos.

El re-empacamiento de los dockings correspondientes a LAO unida a serotonina fue desfavorable en los dos casos analizados, y especialmente en el caso de “SERO-R1-repack” arroja un valor muy alto (lig_sum = 41.51) (**Tabla 13**). Pensamos que esto se puede deber principalmente a que este ligando no presenta ninguna analogía con el extremo amino-carboxilo de los ligandos fisiológicos de la proteína, a lo que se suma el hecho de que la serotonina es muy grande

comparada con los ligandos fisiológicos y por lo tanto es baja la probabilidad de que presente interacciones favorables con la proteína silvestre.

	lig_sum	bk_tot
SERO_R1_repack	41.51	-775.69
SERO_R2_repack	2.87	-808.66

Tabla 13. Valores de lig_sum y bk_tot después del re-empacamiento para los distintos docks de la serotonina con LAO. Cada resultado en esta tabla muestra el promedio de 10 estructuras de re-empacamiento por ROSETTA. Los interfijos R1 y R2 corresponden a los re-empacamientos realizados a partir de las estructuras dock1-SERO y dock2-SERO, respectivamente.

Tras el re-empacamiento se realizó la rutina de diseño de LAO por ROSETTA para generar un sitio de unión para la serotonina. En el caso del grupo SERO_R1 se generaron 100 diseños, y en el caso del grupo SERO_R2 sólo se generaron 50 diseños (**Tabla 14**).

Mediante ROSETTA fue posible mejorar notablemente los valores de lig_sum y de bk_tot de las dos estructuras iniciales. Es interesante observar que la estructura del grupo SERO_R1 que originalmente era muy desfavorable logró obtener valores de lig_sum favorables con tan sólo 6 mutaciones.

# Res	14	52	117	120	121	122	143	160	161	164	191	lig_sum	bk_tot	Estructura Silvestre
Res Original	TYR	PHE	LEU	SER	THR	GLN	GLN	GLN	ASP	ALA	PHE			
# Mut														
6	ASP	***	***	***	ASP	THR	***	THR	ILE	SER	***	-17.09	-840.27	SERO-R1-27
6	SER	TYR	***	***	***	THR	***	THR	ILE	SER	***	-17	-843.82	SERO-R1-84
9	GLU	SER	ASN	GLY	ASP	VAL	ASP	ASN	ILE	***	***	-16.94	-840.8	SERO-R1-42
6	SER	TYR	***	***	***	THR	***	THR	ILE	SER	***	-16.92	-843.18	SERO-R1-100
9	SER	TYR	VAL	GLY	***	ASN	***	THR	ILE	SER	ASN	-16.89	-840.68	SERO-R1-40
8	GLU	SER	ASP	GLY	ASP	VAL	***	THR	ILE	***	***	-20.09	-840.95	SERO-R2-31
9	GLU	HIS	VAL	GLY	ASP	ASP	***	THR	ILE	SER	***	-19.91	-842.51	SERO-R2-20
8	GLU	SER	VAL	***	ASP	THR	GLU	THR	ILE	***	***	-19.69	-842.36	SERO-R2-39
9	GLU	HIS	VAL	GLY	ASP	ASP	GLU	THR	ILE	***	***	-19.36	-841.98	SERO-R2-22
8	GLU	SER	THR	***	ASP	THR	GLU	THR	ILE	***	***	-19.16	-840.91	SERO-R2-50

Tabla 14. Valores de lig_sum y bk_tot después del diseño de LAO para unir el ligando serotonina. Se muestran los cinco mejores diseños de cada grupo.

Correlación energía electrostática y lig_sum

Al realizar una gráfica de la energía electrostática o de Coulomb y el lig_sum obtenido en los diseños analizados, se puede observar que no existe una relación lineal entre estos parámetros

(Figura 4), pues en algunos casos se pueden encontrar diseños con lig_sum no favorables y energías electrostáticas favorables y viceversa. Sin embargo, se observa que en algunos diseños el ΔG^{elec} o el $\Delta G^{\text{coulomb}}$ mejoraron después del diseño. Se puede concluir que en general los diseños de ROSETTA presentan una energía electrostática mejorada, pero no es posible predecir, a partir del lig_sum, cuales diseños presentan esta mejora. Este resultado no es contradictorio, pues como se discutió en la introducción, ROSETTA evalúa mediante parametrizaciones internas varios términos en su función de energía, y por lo tanto los pares iónicos, puentes de hidrógeno y energías de Coulomb (contribuciones electrostáticas) sólo cuentan para una parte de la energía de interacción calculada en los diseños. Además, ROSETTA cuenta con una función para evaluar interacciones de Coulomb con un algoritmo de punto de corte que se encuentra optimizada para una alta velocidad de cálculo, por lo que se espera que su estimación electrostática sea limitada.

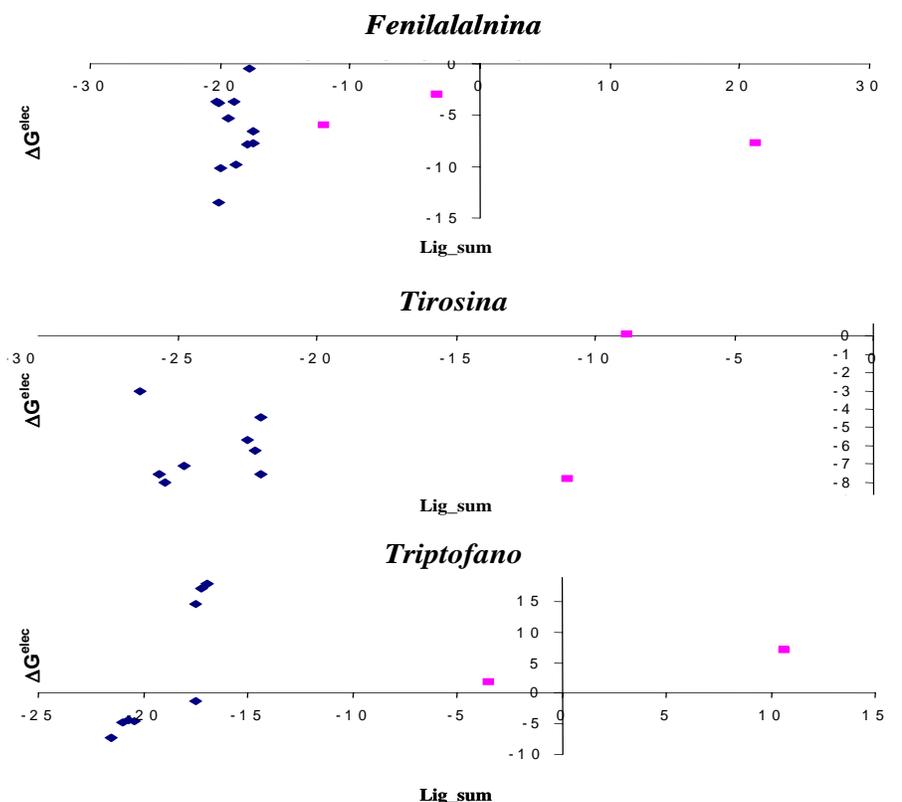


Figura 4. Gráficas de dispersión del lig_sum vs ΔG^{elec} de los diseños para los ligandos: phe, tyr, trp. En azul se muestran los diseños y en rosa se muestran los re-empaquetamientos.

Dinámica molecular de LAO como un método para analizar diseños

Con el objetivo de contar con otro método de evaluación para los diseños generados por ROSETTA, exploramos la posibilidad de utilizar simulaciones de dinámica molecular (MD) para observar las propiedades dinámicas de la interacción de LAO con ligandos fisiológicos y no-fisiológicos.

Posteriormente utilizamos estos resultados para analizar algunos de los diseños generados por ROSETTA.

Estabilidad de la estructura secundaria de LAO a diferentes temperaturas

Se sabe que el aumento de la temperatura en una simulación de dinámica molecular (MD) es equivalente a acelerar los procesos dinámicos del sistema. Con este fin y para disminuir la barrera energética que pudiera mantener unido de manera inespecífica a un ligando no-fisiológico, se buscó la temperatura más alta a la cual se pudieran realizar las MD manteniendo la estructura secundaria de la proteína en el tiempo de la simulación.

Tomando en cuenta un experimento de desnaturalización de la proteína LAO que realizamos (**Figura 5**), se analizaron distintas dinámicas en un intervalo de temperaturas (300, 310, 318 y 326 K). Se eligió una temperatura de 318 K para realizar la caracterización dinámica de LAO, ya que experimentalmente esta temperatura se encuentra antes de la transición de la desnaturalización y en dinámicas moleculares de 20 ns se conserva la estructura secundaria (**Figura 6**).

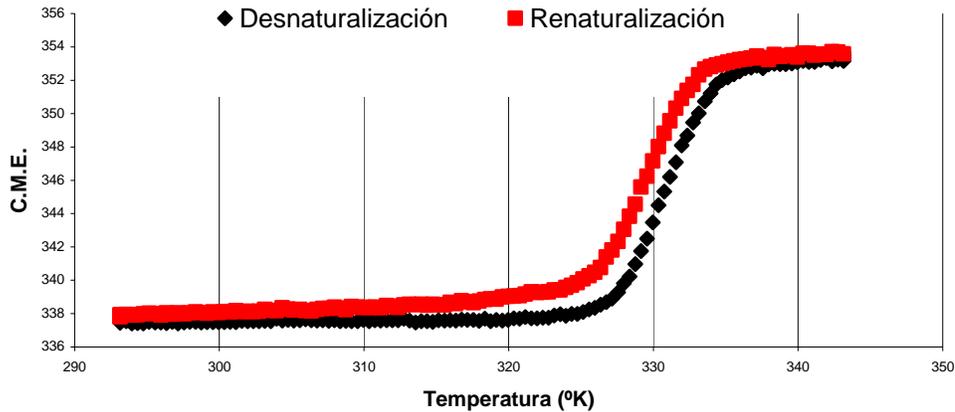


Figura 5. Gráfica de la desnaturalización y renaturalización de LAO por temperatura (cambio en el centro de masa espectral (CME)).

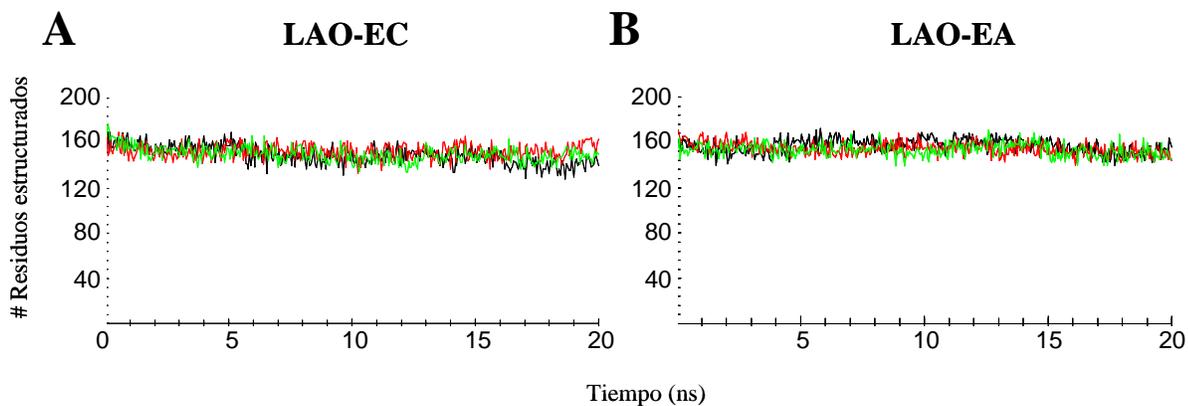


Figura 6. Gráficas del número de residuos de LAO participando en estructuras secundarias a lo largo de MDs de 20 ns a 318 °K. Partiendo de: **A)** el estado cerrado y **B)** el estado abierto de la proteína. En cada gráfica se presentan tres repeticiones de cada MD. Las estructuras secundarias sumadas para la gráfica son las correspondientes a: lamina- β , puente- β , vuelta y α -hélice (calculadas por el programa DSSP (61)).

Ángulos diedros para determinar el grado de apertura de LAO

Una vez que el parámetro de temperatura se fijó, se realizó un estudio sobre la apertura y cierre de la proteína en dinámicas de 20 ns a partir del estado cerrado (EC) y del estado abierto de la proteína (EA). Para realizar este estudio se definieron tres ángulos diedros denominados “diedro-H, diedro-L1 y diedro-L2”. El ángulo “diedro-H” se encuentra definido por los carbonos- α de los aminoácidos: [11, 88, 91, 142]; este ángulo, al encontrarse localizado en el plano de la bisagra,

nos habla del movimiento lateral de un lóbulo de la proteína con respecto a otro (**Figura 7:E-F**). Los ángulos “diedro-L1” y “diedro-L2”, definidos por los aminoácidos: [141, 90, 192, 13] y [53, 192, 90, 144] respectivamente, nos hablan del grado de apertura y cierre de la proteína. Aunque los diedros L1 y L2 son *quasi*-equivalentes (**Figura 7:A-D**), en diferentes simulaciones alguno de estos ángulos presenta más ruido oscilatorio que el otro, y por lo tanto la información que aportan es complementaria. Si estos ángulos son medidos en las estructuras cristalográficas de LAO en el estado abierto y cerrado (PDBs: 2LAO y 1LAG, respectivamente), se obtienen los siguientes valores:

- a) Para el estado abierto: $H1 = -21.94^\circ$, $L1 = 63.30^\circ$ y $L2 = 64.08^\circ$.
- b) Para el estado cerrado: $H1 = -3.74^\circ$, $L1 = 16.21^\circ$ y $L2 = 18.49^\circ$.

Por lo que se puede determinar que el cambio entre el EA y EC de LAO está dado por dos movimientos:

- 1) $\approx 17^\circ$ para el desplazamiento lateral de los lóbulos ($|\Delta\text{diedro-H}|=18.17^\circ$)
- 2) $\approx 46^\circ$ para el movimiento de apertura entre los lóbulos ($|\Delta\text{diedro-L1}|=47.09^\circ$, $|\Delta\text{diedro-L2}|=45.59^\circ$).

Estos parámetros se midieron en todas las dinámicas moleculares que se realizaron. Además se realizó un análisis visual de cada una de las trayectorias para determinar la permanencia de los ligandos en el sitio de unión de la proteína. Los estados iniciales que se utilizaron para realizar la caracterización dinámica fueron:

- a) El estado abierto de la proteína (EA-LAO): sin ligando; con los ligandos fisiológicos: arginina lisina e histidina; y con los ligandos no-fisiológicos: fenilalanina, triptofano y tirosina.
- b) El estado cerrado de la proteína (EC-LAO) en las mismas combinaciones que en “a”).

Para analizar la reproducibilidad de los fenómenos se realizaron tres repeticiones de MD para cada una de las condiciones. En cada repetición, para generar las velocidades iniciales de los átomos se utilizaron diferentes semillas, con el fin de generar diferentes trayectorias del ensamble dinámico.

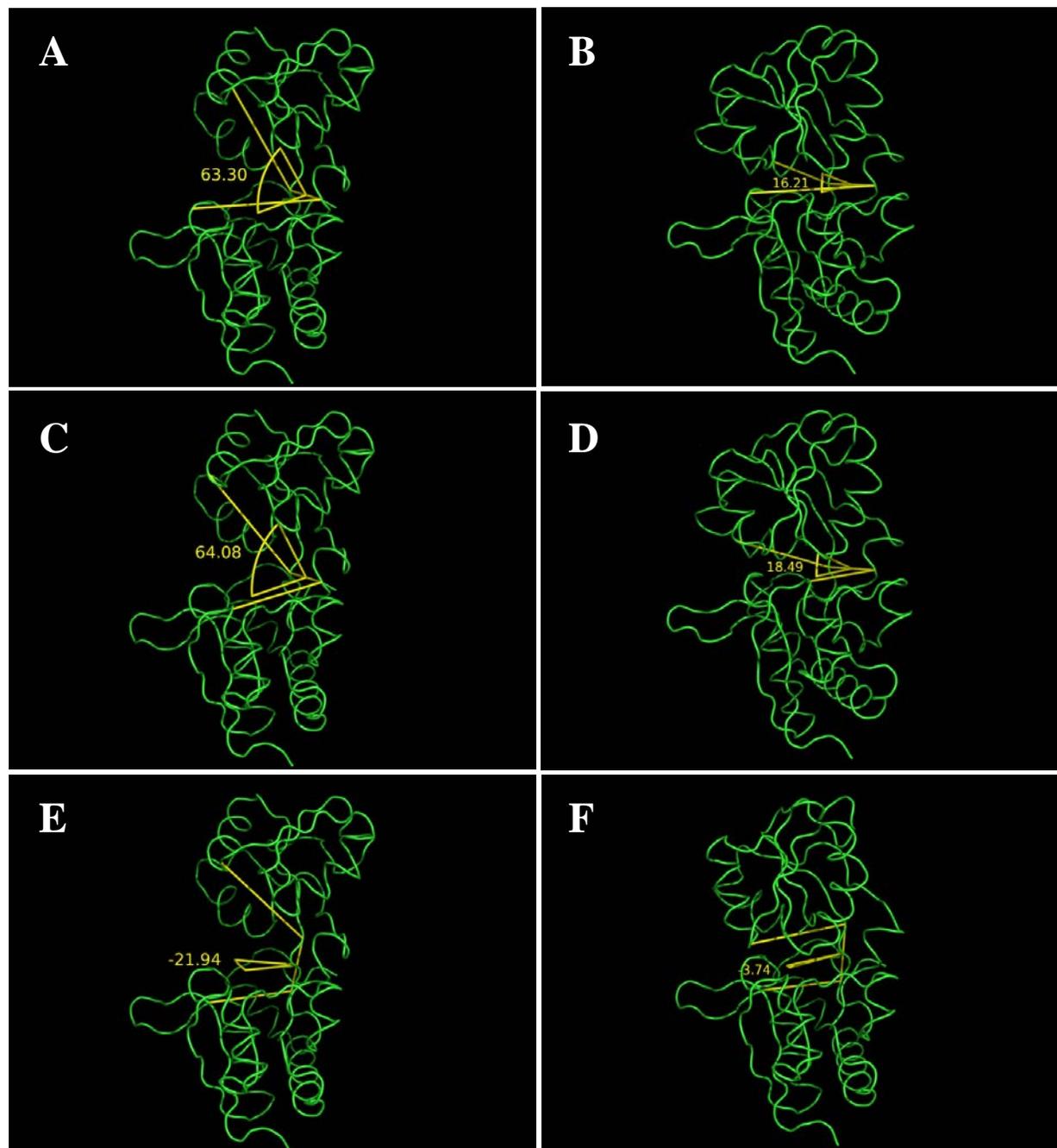


Figura 7. Ángulos diedros que fueron medidos en LAO en el estado abierto (lado izquierdo) y cerrado (lado derecho). **A-B)** ángulo diedro L1; **C-D)** ángulo diedro L2; **E-F)** ángulo diedro H1.

MD de LAO sin ligando

Las MD de LAO comenzando en el EA no mostraron una tendencia a cambiar de estado conformacional (ej. EA→EC) (**Figura 8**). Por otra parte, las dinámicas que comenzaron en el EC muestran claramente un cambio en el ángulo de cierre en los primeros 3 ns de simulación, pasando de $\approx 20^\circ$ iniciales a $\approx 50^\circ$ (diedros L1 y L2), es decir, cambiaron de conformación a un estado semi-abierto.

Estos resultados son congruentes con el mecanismo de acción propuesto para la proteína LAO, en el que se propone un “estado abierto libre de ligando” y un “estado cerrado estable cuando la proteína interactúa con el ligando” (**15**). En el caso del desplazamiento lateral de los lóbulos (diedro-H), se puede observar que el estado abierto muestra un cambio de $\approx 15^\circ$ en el primer nanosegundo de simulación. El estado cerrado también muestra un cambio significativo en los primeros 3 ns equivalente a $\approx 35^\circ$.

Se puede apreciar que en las dinámicas libres de ligando, independientemente del estado inicial la proteína tiende a visitar un espacio conformacional en común, tanto para los ángulos diedros L1 y L2 ($\approx 50^\circ$) como para el diedro-H ($\approx 30^\circ$).

En conclusión, los datos señalan que LAO, en MDs de 20 ns, a una temperatura de 318 K y en ausencia de ligando, tiende a una conformación “semi-abierta”. Esta conformación “semi-abierta” es visitada tanto por las simulaciones que comienzan en el EA como por aquellas que comienzan en el EC. Por lo tanto, este espacio conformacional podría representar un ensamble dinámico del estado nativo para la forma libre de ligando o un intermediario en el mecanismo de unión.

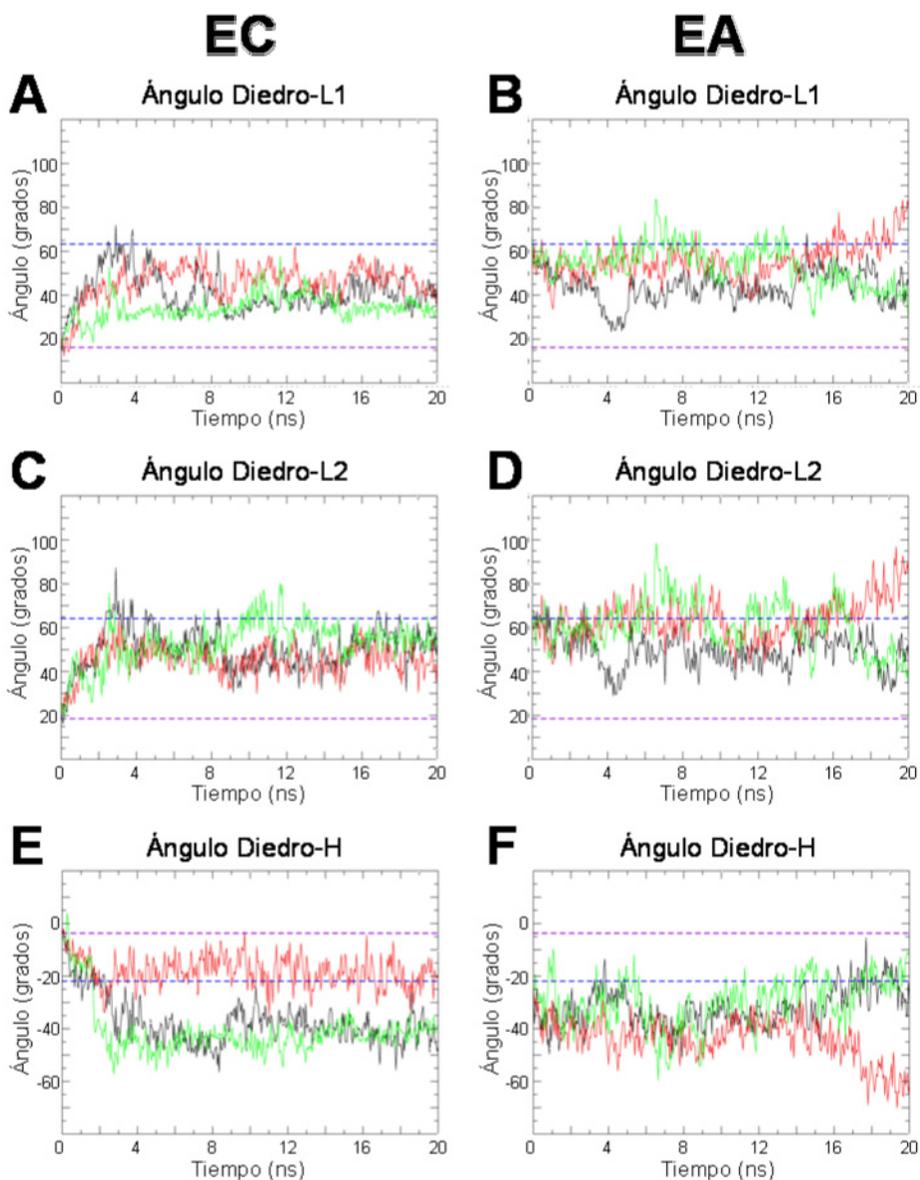


Figura 8. Gráficas de los ángulos diedros que fueron medidos lo largo de tres MDs independientes de LAO sin ligando, partiendo del EC (izquierda) y del EA (derecha). A-B) ángulo diedro L1; C-D) ángulo diedro L2; E-F) ángulo diedro H1. La línea (discontinua) magenta representa el valor correspondiente al EC cristalográfico, mientras que la línea (discontinua) azul representa el valor correspondiente EA cristalográfico.

MD de LAO con ligandos fisiológicos

Las MDs de LAO realizadas a partir del EC en presencia de los ligandos fisiológicos (arginina, lisina o histidina) muestran en todos los casos que la conformación cerrada se mantiene a lo largo de toda la dinámica. Es notable el hecho de que el aminoácido lisina provoca oscilaciones del

ángulo de apertura (diedros-L1 y L2) y un desplazamiento lateral de los lóbulos (diedro-H) más grandes que las que se observan con los otros ligandos.

Si tomamos como punto de partida el EA con los distintos ligandos fisiológicos, observamos diferentes comportamientos según el ligando que se usa en la MD. Los resultados para cada ligando se discuten a continuación.

MD LAO-arginina

En las simulaciones del ligando arginina, partiendo del EC, encontramos una conformación cerrada estable a lo largo de toda la simulación con muy baja oscilación en los ángulos medidos, así como una alta estabilidad del ligando en el sitio de unión (**Figura 9**). En el caso de las simulaciones del EA pudimos observar que en una de las simulaciones, la conformación de la proteína alcanza una conformación cerrada estable (\approx a los 8 ns de simulación) y en los otros dos casos se alcanza un estado semi-cerrado. Estos resultados concuerdan con lo esperado, pues las repeticiones de las dinámicas se realizan con diferentes semillas para asignar las velocidades iniciales de los átomos y por lo tanto representan diferentes trayectorias de un ensamble dinámico. Sin embargo, cabe mencionar que de las nueve simulaciones que realizamos con ligandos fisiológicos partiendo del EA, solo en tres casos observamos que LAO adquiere una conformación cerrada estable. Esto nos habla de que el evento de asociación del ligando a la proteína e inducción del EC presenta cierta barrera energética que no es fácil de vencer. Este hecho, en conjunto con una alta temperatura (318 °K), supone que los ligandos y la proteína se pueden solvatar con mayor facilidad y por lo tanto que sea más difícil alcanzar el EC de manera estable que a una temperatura mas baja. En el caso del diedro-H podemos observar que, a diferencia de las simulaciones sin ligando, el ángulo se mantiene estable a lo largo de toda la simulación para el EA y presenta un cambio muy pequeño para el EC.

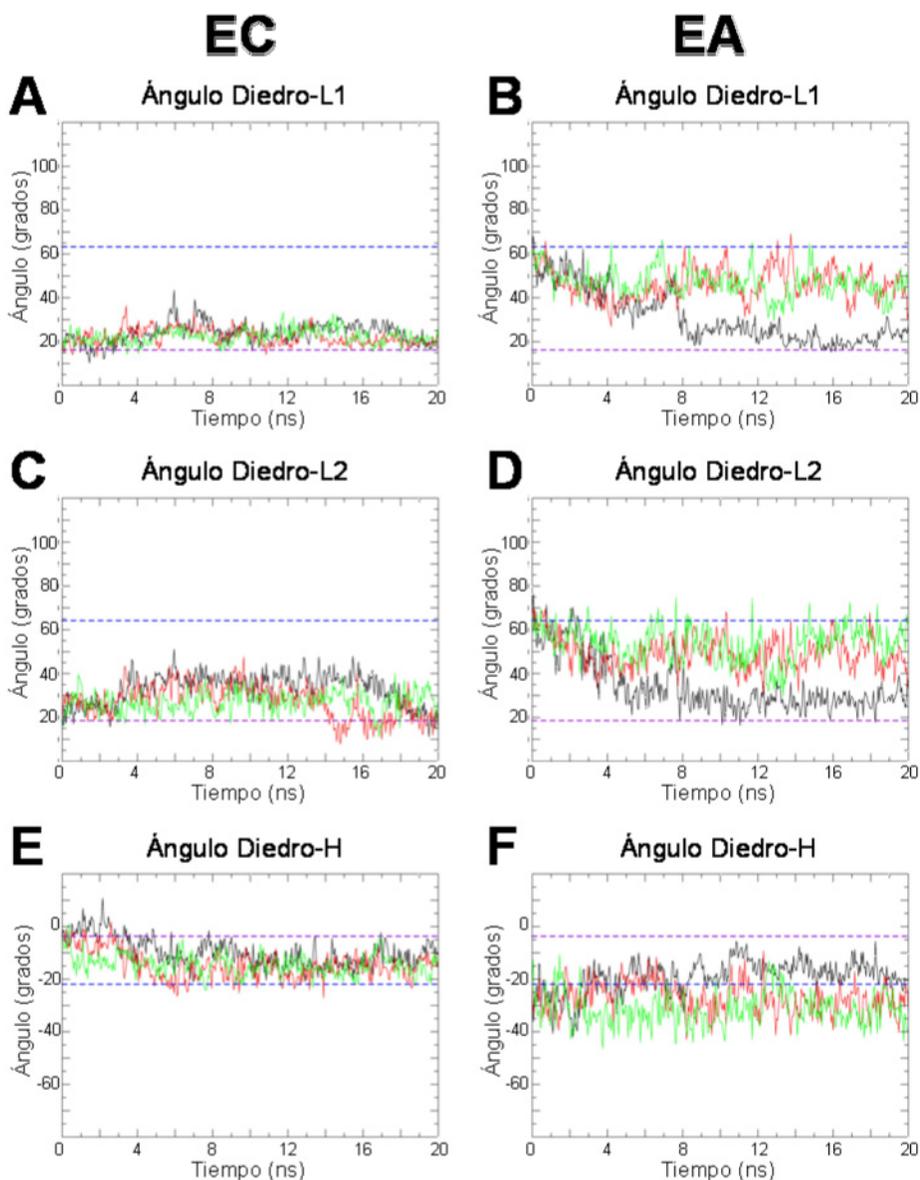


Figura 9. Gráficas de los ángulos diedros que fueron medidos lo largo de tres MDs independientes de LAO con el ligando arginina, partiendo del EC (izquierda) y del EA (derecha). **A-B** ángulo diedro L1; **C-D** ángulo diedro L2; **E-F** ángulo diedro H1. Los distintos colores de los trazos representan simulaciones independientes y son análogos en las distintas Gráfica. La línea (discontinua) magenta representa el valor correspondiente al EC cristalográfico, mientras que la línea (discontinua) azul representa el valor correspondiente EA cristalográfico.

Si utilizamos como parámetro para medir la estabilidad del ligando en el sitio de unión la distancia entre el centro de masa (CM) de la Ala-90 y el CM del ligando (**Figura 10**) observamos que, partiendo tanto del EC como del EA, el ligando arginina permanece estable en el sitio de unión con pequeñas oscilaciones en la distancia (que se mantienen entre 0.75 y 1.5 nm) (**Figura**

11). Es importante notar que se presentan oscilaciones más grandes al inicio de las dinámicas que parten del EA. Este evento no es raro, pues el ligando tiene más espacio para moverse en la cavidad de unión del EA que en la cavidad del EC (donde se supone que el ligando ha sido estabilizado). Además, las oscilaciones en la distancia que se presentan en las MDs que parten del EA podrían representar reacomodos tanto del ligando como de las cadenas laterales del sitio de unión para generar una interacción favorable con el ligando. Esta inferencia se ve apoyada por el hecho de que después de 10 ns de simulación las oscilaciones disminuyen notablemente en amplitud.

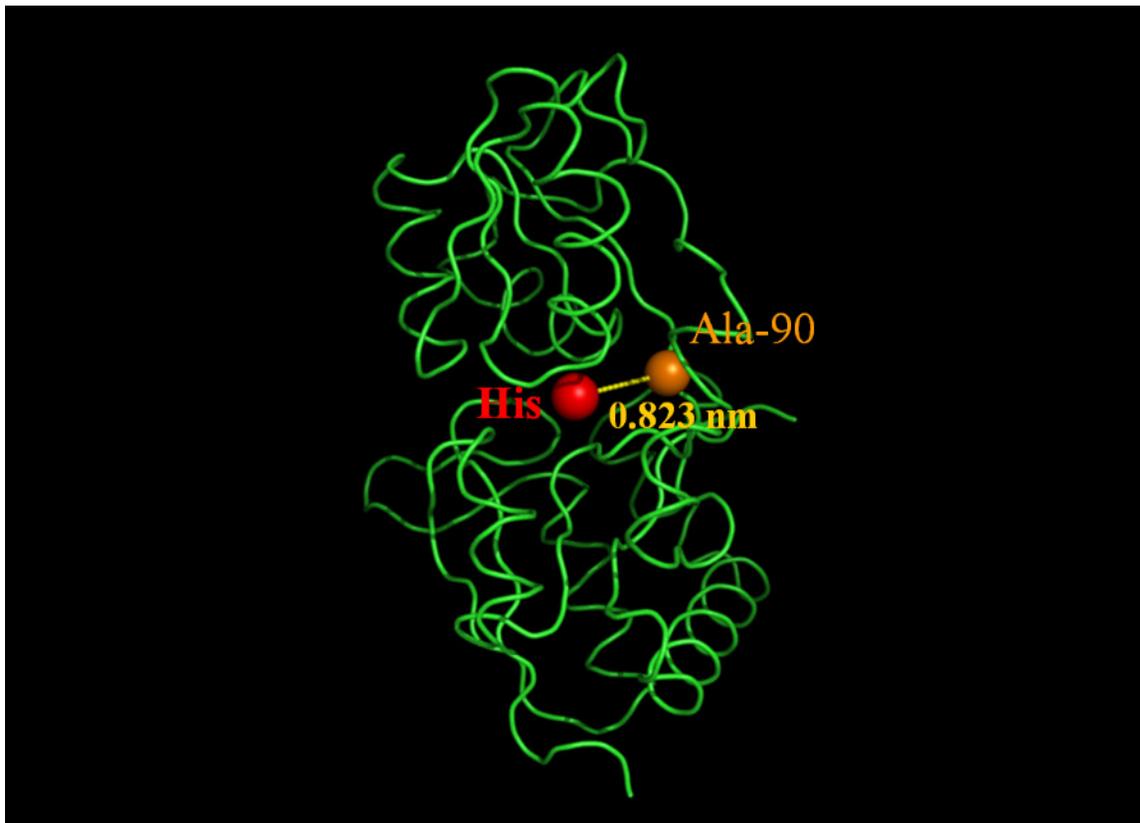


Figura 10. Distancia del centro de masa de 1 residuo Ala90 (que se encuentra en la bisagra) al centro de masa del ligando histidina.

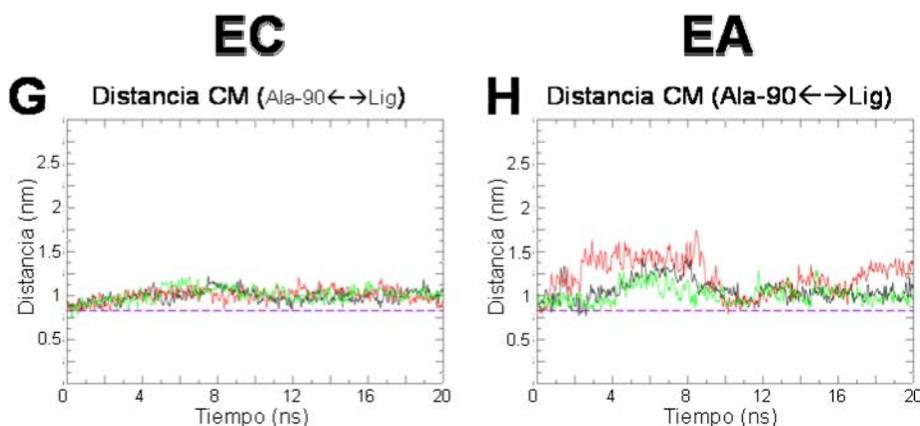


Figura 11. Gráficas de la distancia del “CM de la Ala-90 al CM ligando arginina”, a lo largo de tres MDs independientes de LAO, partiendo del EC (izquierda) y del EA (derecha). Los distintos colores de los trazos representan simulaciones independientes y son análogos en las distintas Gráfica. La línea (discontinua) magenta representa el valor correspondiente EC cristalográfico.

MD LAO-histidina

La simulación del ligando histidina partiendo del EC muestra la conformación cerrada más estable de las simulaciones realizadas con ligandos fisiológicos. Esto se ve reflejado en todos los ángulos diedros medidos (**Figura 12**). En las simulaciones del EA de LAO-His tenemos resultados heterogéneos: en un caso LAO alcanzó una conformación atípicamente abierta ($\approx 90^\circ$) y un ángulo diedro-H1 demasiado grande, pero el ligando se mantuvo estable en el sitio de unión; en otra simulación sólo se alcanzó un estado semi-cerrado, pero se perdió la interacción del ligando (distancia Ala-90 ≈ 3 nm); en una tercera simulación se alcanzó una conformación cerrada estable con el ligando unido. Contrariamente a lo que se podría suponer, mientras que la dinámica que alcanzó una conformación atípica de apertura de LAO mantuvo estable al ligando en el sitio de unión, la simulación que mostró un comportamiento “típico” del estado abierto perdió la interacción con el ligando.

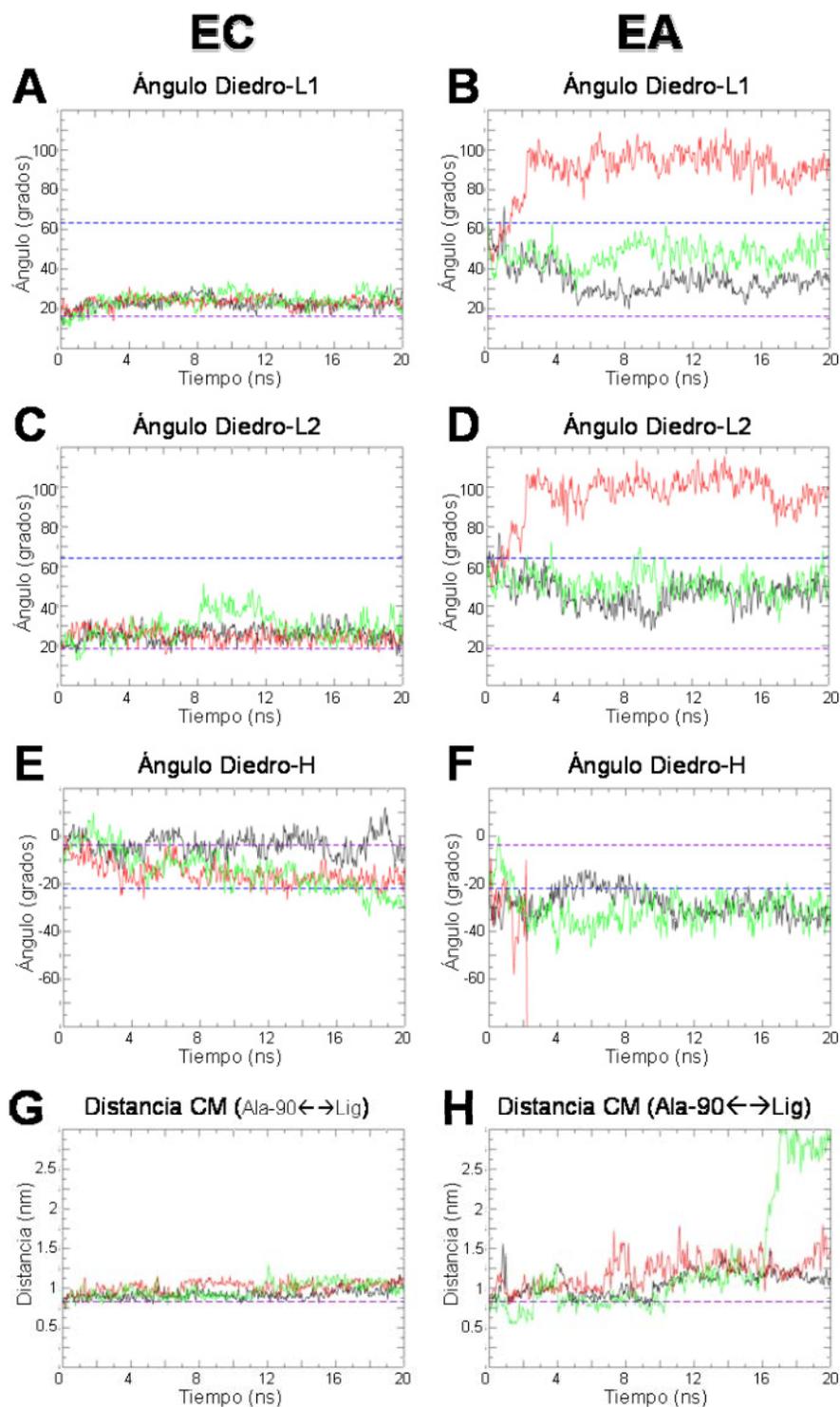


Figura 12. Gráficas de los ángulos diedros y distancias del ligando con respecto a la Ala-90, que fueron medidos lo largo de tres MDs independientes de LAO con el ligando histidina, partiendo del EC (izquierda) y del EA (derecha). **A-B** ángulo diedro L1; **C-D** ángulo diedro L2; **E-F** ángulo diedro H1; **G-H** distancia con respecto a la Ala-90. Los distintos colores de los trazos representan simulaciones independientes y son análogos en las distintas Gráfica. La línea (discontinua) magenta representa el valor correspondiente al EC cristalográfico, mientras que la línea (discontinua) azul representa el valor correspondiente EA cristalográfico.

El hecho de que los resultados de las simulaciones con el ligando histidina no muestren resultados tan homogéneos como los que se encuentran con la arginina podría deberse, entre otras cosas, a que LAO es más afín a la arginina por más de un orden de magnitud con respecto a la histidina (K_d His = 500 nM vs Arg = 14 nM). Además, el EA con el ligando ha sido creado de manera artificial y podría suceder que la posición en la que se encuentra inicialmente el ligando histidina no sea tan favorable como aquella que tiene la arginina. En cualquier caso, se requerirá una mayor cantidad de estudios para aclarar este punto.

MD LAO-lisina

En las MDs con el ligando lisina, partiendo del EC, encontramos una conformación cerrada estable pero con mayores oscilaciones (tanto en amplitud como en frecuencia) que para los otros ligandos fisiológicos (**Figura 13**). Una explicación posible sería que para todas las simulaciones del EC se utilizó la estructura cristalográfica correspondiente a LAO unida a histidina (ver métodos), y aunque se ha descrito que prácticamente todas las cadenas laterales del sitio de unión se encuentran en la misma disposición para todos los ligandos (ver introducción), podría ser que pequeñas diferencias que se requieren para unir el ligando lisina provocaran oscilaciones en el ángulo de apertura de proteína para corregir dichas posiciones. Esta conclusión se ve apoyada por dos hechos:

- 1) En las simulaciones que comienzan del EC, a pesar de las oscilaciones en el ángulo de apertura, la distancia de la Ala90 con respecto al ligando lisina permanece estable a lo largo del tiempo, es decir, que el ligando se mantiene estable en el sitio de unión.
- 2) En las simulaciones del EC con el ligando histidina se encuentran las oscilaciones más bajas del estado cerrado, lo cual podría ser el efecto de realizar una simulación entre un ligando y una estructura cristalográfica totalmente complementaria.

LAO presentó un comportamiento heterogéneo en las simulaciones partiendo del EA de LAO-Lis. En una simulación la proteína permaneció en el estado abierto a lo largo de toda la trayectoria, y el ligando aunque permaneció dentro de la cavidad de unión, modificó su posición y se acercó atípicamente a la Ala-90 (dicho de otra manera: “el ligando se desacomodó”). En otro caso, la proteína migró al estado cerrado desde el inicio de la simulación, pero perdió al ligando aproximadamente a los 6 ns. En la tercera simulación la proteína adquirió la conformación cerrada desde el inicio de la simulación, al mismo tiempo que logró estabilizar al ligando en el sitio de unión durante el resto de la simulación. Nuevamente, el hecho de que los resultados de las MDs partiendo del EA con el ligando lisina no muestren resultados tan homogéneos como los que se encuentran con la arginina podría deberse, entre otras cosas, a que el estado abierto con el ligando ha sido creado de manera artificial y podría no ser tan favorable como su contraparte con arginina (al igual que como se discutió para el ligando histidina). Sin embargo, es importante notar que en una de las simulaciones que partió del EA se logró obtener un estado cerrado estable con el ligando. Sería necesario realizar un mayor número de simulaciones para poder determinar con qué frecuencia se pueden observar los diversos eventos de unión estable de los ligandos fisiológicos. En cualquier caso, no cabe duda que estos eventos de unión del ligando se pueden observar en las simulaciones realizadas (con mayor o menor frecuencia), y por lo tanto las MD se pueden utilizar para observar diferencias entre el comportamiento dinámico de LAO en presencia de ligandos fisiológicos y no-fisiológicos en su sitio de unión.

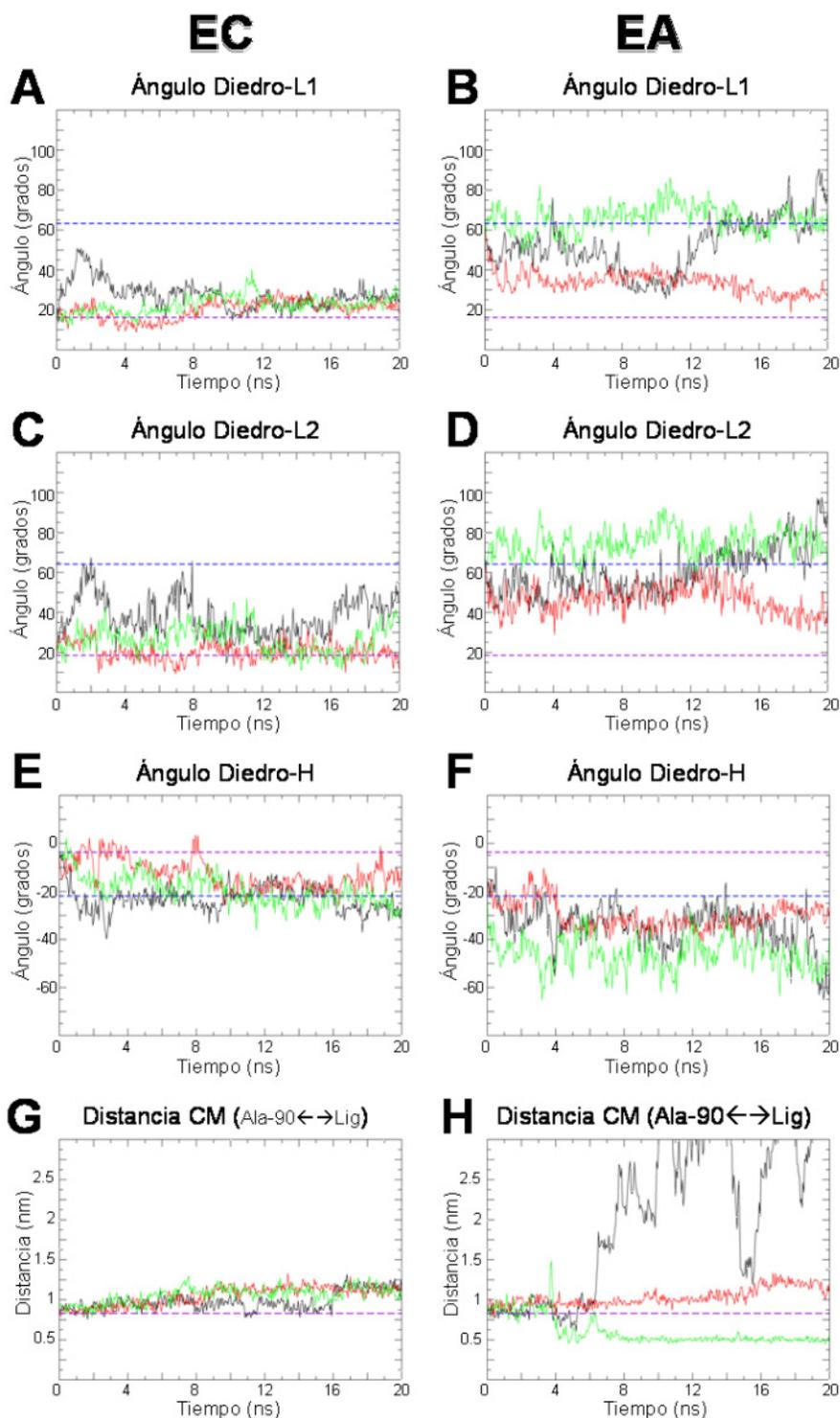


Figura 13. Gráficas de los ángulos diedros y distancias del ligando con respecto a la Ala-90, que fueron medidos lo largo de tres MDs independientes de LAO con el ligando lisina, partiendo del EC (izquierda) y del EA (derecha). A-B) ángulo diedro L1; C-D) ángulo diedro L2; E-F) ángulo diedro H1; G-H) distancia con respecto a la Ala-90. Los colores de los trazos son los mismos para cada repetición en todas las Gráfica. La línea (discontinua) magenta representa el valor correspondiente al EC cristalográfico, mientras que la línea (discontinua) azul representa el valor correspondiente EA cristalográfico.

LAO, los ligandos fisiológicos y sus propiedades dinámicas

Los resultados de las MDs de LAO que parten tanto del EA como del EC muestran, con reproducibilidad, propiedades dinámicas de la proteína LAO al interactuar con sus ligandos fisiológicos. Estas propiedades se listan a continuación:

- Partiendo del EC, un EC conformacionalmente estable.
- Partiendo del EC, un EC que mantiene al ligando en el sitio de unión.
- Partiendo del EA, un EA que mantiene al ligando en el sitio de unión.
- Partiendo del EA, un EA que tiende a un estado cerrado o semi-cerrado en los casos de estabilización exitosa del ligando.

Al comparar estas propiedades con las obtenidas para las dinámicas de la interacción de ligandos no-fisiológicos con la proteína silvestres podemos observar diferencias importantes.

MD de LAO con ligandos no-fisiológicos

En el caso de las MDs de LAO con ligandos no-fisiológicos en el sitio de unión (tirosina, triptofano y fenilalanina) podemos encontrar diferencias con respecto al comportamiento de la proteína en presencia de ligandos fisiológicos. Las diferencias esenciales son: una desestabilización de la conformación cerrada (con excepción de las MD con el ligando triptofano) y una desestabilización del ligando en el sitio de unión de las MDs que parten del EA. A continuación se explicarán los resultados para cada ligando no-fisiológico.

MD LAO-tirosina

A diferencia de las MDs realizadas con ligandos fisiológicos, las simulaciones de LAO-Tyr partiendo del EC, muestran una desestabilización de la conformación cerrada. Esta desestabilización está acompañada de grandes oscilaciones en los ángulos diedros L1 y L2 y una ligera desestabilización del ángulo diedro-H (**Figura 14**). La distancia del ligando con respecto a

la Ala-90 no presenta una diferencia notable con respecto a las MDs con ligandos fisiológicos. La inspección visual de las trayectorias revela que sólo en una ocasión el ligando tiene grandes rotaciones de la cadena lateral y desplazamientos dentro del sitio de unión.

Al observar a LAO en las dinámicas que parten del EA se observa un comportamiento que se asemeja al obtenido con los ligandos fisiológicos, alcanzando siempre conformaciones semi-cerradas. Sin embargo, en las dinámicas que parten del EA se puede notar que el ligando es inestable, ya que las oscilaciones en la distancia tienden a alejar cada vez más el ligando del sitio de unión. La inspección visual de las dinámicas partiendo del EA revela que los ligandos pierden la interacción con el sitio de unión.

El hecho de que el ligando se mantenga estable durante las simulaciones del estado cerrado, podría estar dado por una combinación de dos eventos:

- 1) Que la proteína primero tiene que alcanzar un estado lo suficientemente abierto para poder liberar al ligando, por lo que se requerirían simulaciones más largas para permitir la apertura de la proteína, lo que posteriormente permitiría la salida del ligando.
- 2) Que los ligandos no-fisiológicos que se usaron en las simulaciones comparten la característica de ser aminoácidos. Por lo tanto, pueden existir algunas interacciones en común que estabilizan al ligando en el sitio de unión, por lo que el proceso de discriminación se daría en tiempos más largos que si no existieran dichas interacciones.

En cualquier caso se requerirán más estudios para precisar este punto.

De los resultados de las MDs se pueden resaltar dos diferencias esenciales entre la presencia del ligando no-fisiológico tirosina y los ligandos fisiológicos. Estas diferencias son:

- Desestabilización de la conformación del EC.
- Desestabilización del ligando en el EA.

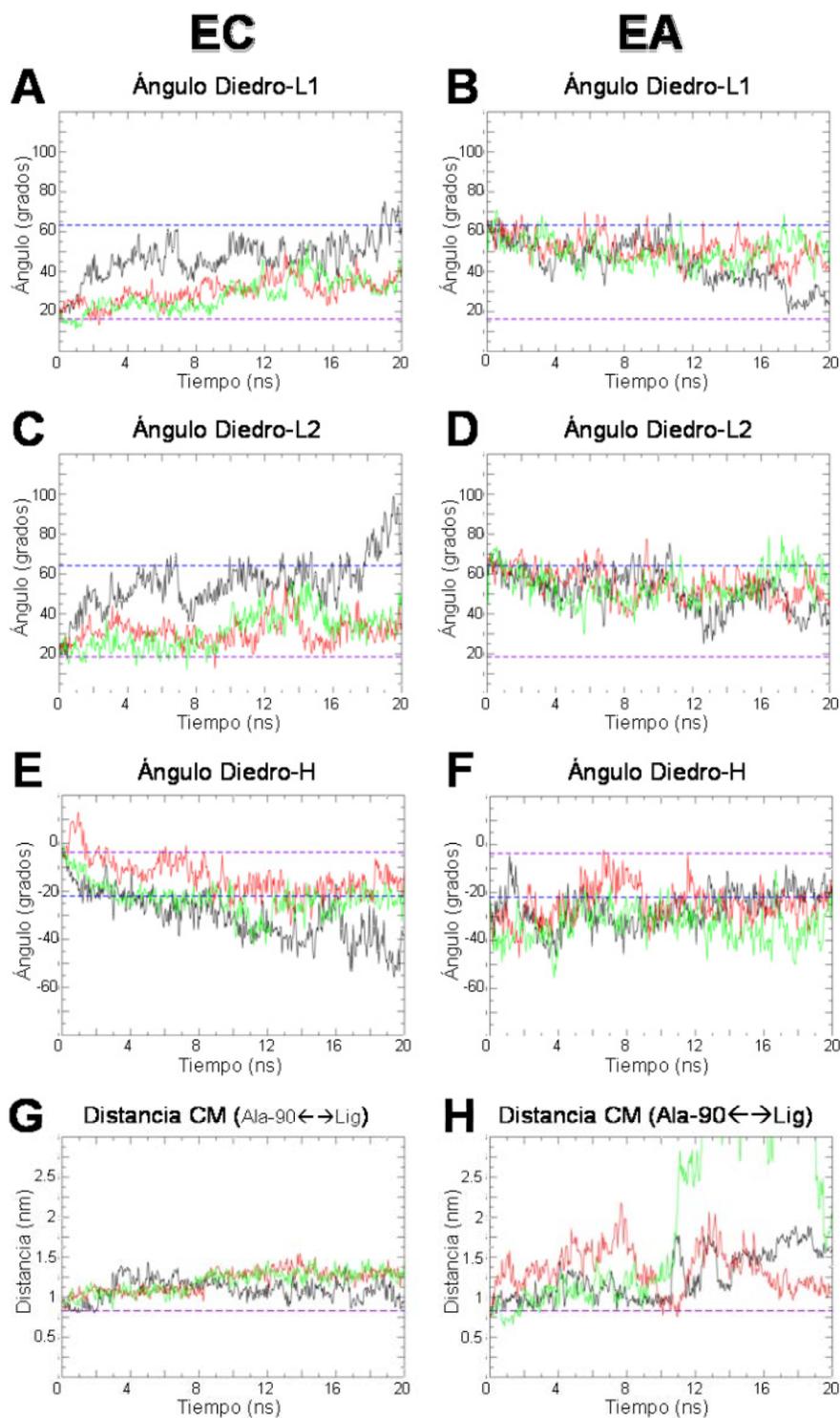


Figura 14. Gráficas de los ángulos diedros y distancias del ligando con respecto a la Ala-90, que fueron medidos lo largo de tres MDs independientes de LAO con el ligando tirosina, partiendo del EC (izquierda) y del EA (derecha). A-B) ángulo diedro L1; C-D) ángulo diedro L2; E-F) ángulo diedro H1; G-H) distancia con respecto a la Ala-90. Los colores de los trazos son los mismos para cada repetición en todas las Gráfica. La línea (discontinua) magenta representa el valor correspondiente al EC cristalográfico, mientras que la línea (discontinua) azul representa el valor correspondiente EA cristalográfico.

MD LAO-fenilalanina

En las MDs de LAO-Phe, partiendo del EC, sólo en un caso se observa una desestabilización de la conformación cerrada. En el ángulo diedro-H no se observa ninguna diferencia notable con los ligandos fisiológicos (**Figura 15**), y en el caso del ligando tampoco se observa ninguna diferencia notable con respecto a las dinámicas de los ligandos fisiológicos.

Al analizar las dinámicas que parten del EA se observa en LAO un comportamiento que se asemeja al observado con los ligandos fisiológicos, alcanzando conformaciones semi-cerradas. Al igual que con el ligando tirosina para las dinámicas que parten del EA, la diferencia más importante con respecto a los ligandos fisiológicos se observa en cuanto a la distancia del ligando con respecto a la Ala-90, ya que el ligando no permanece en el sitio de unión y tiende a alejarse cada vez más durante la dinámica. Las diferencias encontradas no son tan robustas como con el ligando tirosina, lo cual podría deberse a que la fenilalanina es más hidrofóbica, por lo que resulta más difícil solvatarlo y por lo tanto es más difícil que abandone la cavidad de unión (con respecto al ligando tirosina).

La inspección visual revela que en las dinámicas que parten del EA los ligandos pierden la interacción con el sitio de unión y se “desacomodan”. El hecho de que el ligando se mantenga en el sitio de unión durante las simulaciones del EC podría estar dado por la misma razón explicada para la tirosina.

Dados los resultados se distinguen dos diferencias entre la presencia del ligando no-fisiológico fenilalanina y los ligandos fisiológicos. Estas diferencias son:

- Ligera desestabilización de la conformación del EC.
- Desestabilización del ligando en el EA.

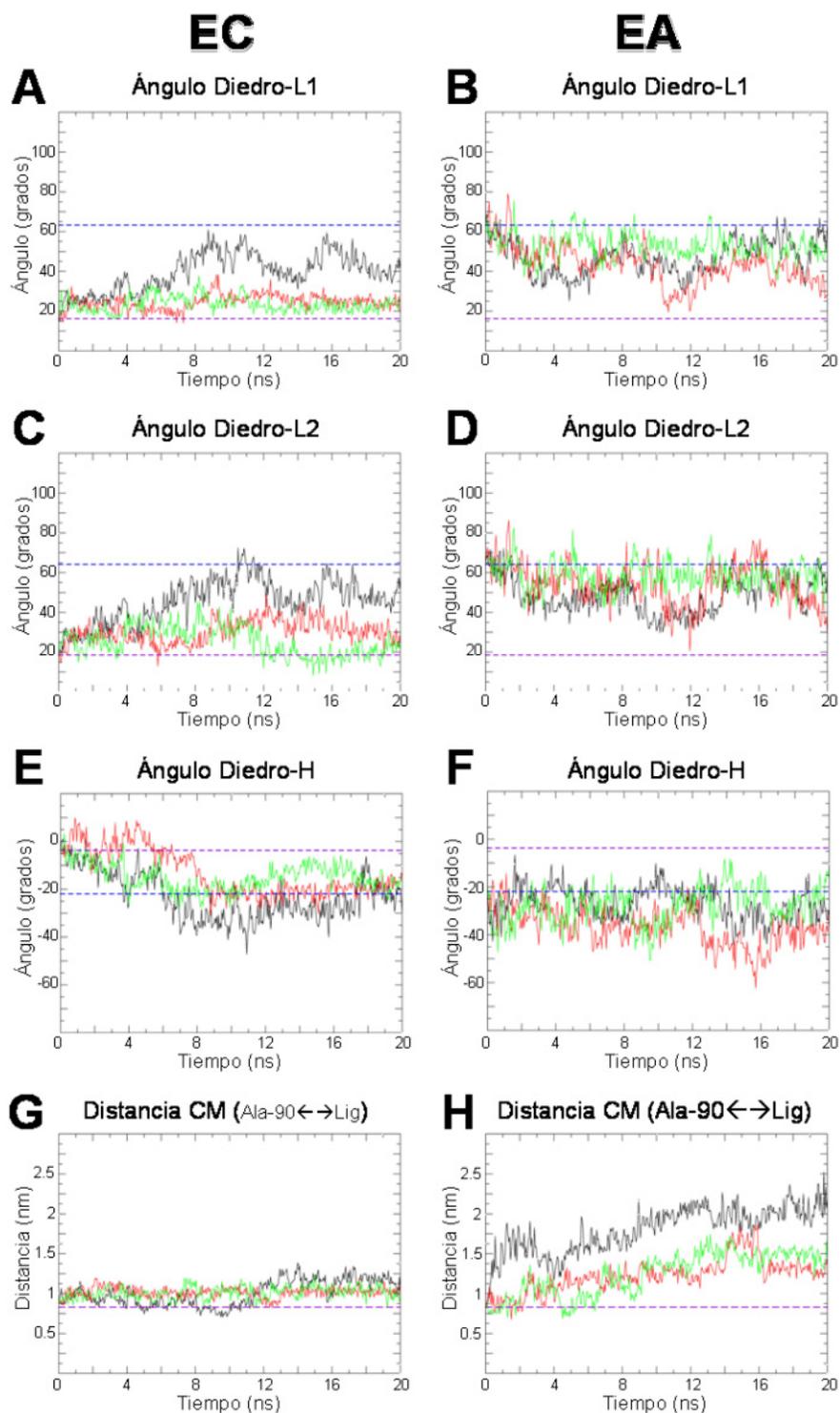


Figura 15. Gráficas de los ángulos diedros y distancias del ligando con respecto a la Ala-90, que fueron medidos lo largo de tres MDs independientes de LAO con el ligando fenilalanina, partiendo del EC (izquierda) y del EA (derecha). **A-B** ángulo diedro L1; **C-D** ángulo diedro L2; **E-F** ángulo diedro H1; **G-H** distancia con respecto a la Ala-90. Los colores de los trazos son los mismos para cada repetición en todas las Gráfica. La línea (discontinua) magenta representa el valor correspondiente al EC cristalográfico, mientras que la línea (discontinua) azul representa el valor correspondiente EA cristalográfico.

MD LAO-triptofano

Las simulaciones de dinámica molecular de LAO-Trp partiendo del EC son las que presentan menos diferencias con respecto a los ligandos fisiológicos (en lo que a cambios conformacionales se refiere). En las simulaciones que parten del EC no se observa ninguna diferencia con respecto al comportamiento de LAO en presencia de los ligandos fisiológicos, es decir, que la proteína se mantiene establemente cerrada y, de igual manera, el ligando se mantiene estable en el sitio de unión (**Figura 16**). Este comportamiento podría deberse a que el ligando triptofano es el ligando más hidrofóbico probado, y al insertarlo de manera artificial en una conformación cerrada de la proteína podría dar como resultado una barrera energética grande (debido al costo energético de solvatar un ligando muy hidrofóbico) para que la proteína pase a la conformación abierta.

Con el EA se observan dos casos que oscilan alrededor de la conformación abierta, y en uno de los casos se observa una tendencia a un estado semi-cerrado. Al observar la distancia del ligando con respecto a la Ala-90, inmediatamente se pueden ver diferencias con respecto a los ligandos fisiológicos, pues en todos los caso se pierde la interacción con el ligando, dando a lugar a su desacomodo en el sitio de unión. Por lo que para el ligando triptofano se puede resaltar sólo una diferencia esencial con respecto a los ligandos fisiológicos:

- Desestabilización del ligando en el EA.

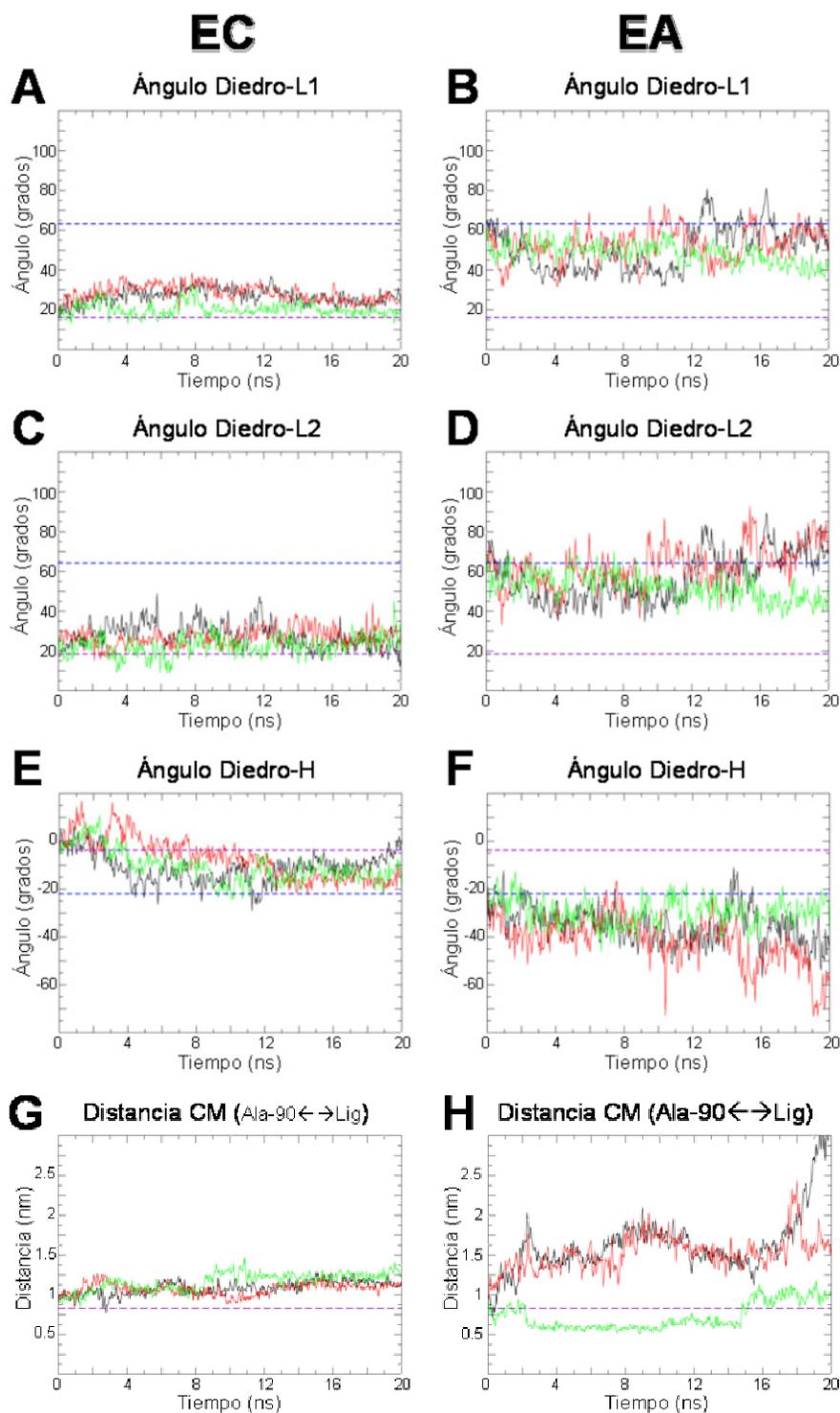


Figura 16. Gráficas de los ángulos diedros y distancias del ligando con respecto a la Ala-90, que fueron medidos lo largo de tres MDs independientes de LAO con el ligando triptofano, partiendo del EC (izquierda) y del EA (derecha). **A-B** ángulo diedro L1; **C-D** ángulo diedro L2; **E-F** ángulo diedro H1; **G-H** distancia con respecto a la Ala-90. Los colores de los trazos son los mismos para cada repetición en todas las Gráfica. La línea (discontinua) magenta representa el valor correspondiente al EC cristalográfico, mientras que la línea (discontinua) azul representa el valor correspondiente EA cristalográfico.

Propiedades dinámicas de los ligandos no-fisiológicos vs fisiológicos

Los resultados revelan diferencias clave entre las características observadas en las MDs de LAO en presencia de ligandos fisiológicos y de ligandos no-fisiológicos. Probablemente la diferencia más importante sea que, en general, los ligandos no-fisiológicos no se mantienen estables en el sitio de unión de las dinámicas que parten del EA, mientras que los ligandos fisiológicos tienden a permanecer en el sitio de unión. También es importante el hecho de que las simulaciones que parten del EC con los ligandos fisiológicos siempre presentan una conformación cerrada estable, mientras que las dinámicas con los ligandos no-fisiológicos presentan generalmente una desestabilización del estado cerrado (con excepción, como ya se mencionó, del ligando triptofano).

Por otra parte, en las MDs que parten del EA, tanto para el caso de ligandos fisiológicos como para el de los ligandos no-fisiológicos, se observan movimientos muy similares de la proteína, es decir, que LAO visita conformaciones semi-cerradas. Esto refuerza la hipótesis de que esta(s) conformación(es) semi-cerrada(s) pueden pertenecer a un intermediario en el mecanismo de unión del ligando, y por lo tanto, esta conformación puede ser importante para reconocer al ligando y para discriminar a los ligandos incorrectos.

Finalmente, en las MDs de los ligandos fisiológicos se encuentran algunos casos en los que la proteína pasa de la conformación abierta a la cerrada de manera estable, y al mismo tiempo mantiene al ligando en el sitio de unión. Mientras que los ligandos no-fisiológicos nunca inducen un estado cerrado estable.

En la **tabla 15** se pondera la estabilidad del EC y el EA así como la estabilidad del ligando en ambos estados, tanto para ligandos fisiológicos como no-fisiológicos. La ponderación está realizada en una escala de cero a 1, donde cero significa que el evento no ocurre y 1 que el evento se da por completo.

Ligando	Repetición	Mantiene el EC	Ligando estable	Inspección visual del ligando	Cambia al EC	Ligando estable	Inspección visual del ligando	Suma	Promedio
Vacía	1	0	*	*	0	*	*	*	*
	2	0	*	*	0	*	*	*	
	3	0	*	*	0	*	*	*	
Arginina	1	1	1	1	1	1	1	6	5.33
	2	1	1	1	0	1	1	5	
	3	1	1	1	0	1	1	5	
Lisina	1	1	1	1	0	0	0	3	4.16
	2	1	1	1	1	1	1	6	
	3	1	1	1	0	0	0.5	3.5	
Histidina	1	1	1	1	1	1	1	6	4.5
	2	1	1	1	0	1	0.5	4.5	
	3	1	1	1	0	0	0	3	
Tirosina	1	0	1	0.75	0	0	0	1.75	1.92
	2	0	1	1	0	0	0	2	
	3	0	1	1	0	0	0	2	
Fenilalanina	1	0	1	0.75	0	0	0.25	2	2.75
	2	1	1	1	0	0	0.25	3.25	
	3	1	1	1	0	0	0	3	
Triptofano	1	1	1	1	0	0	0	3	3.25
	2	1	1	1	0	0	0.25	3.25	
	3	1	1	1	0	0	0.5	3.5	

Tabla 15. Ponderación de diversos factores de la estabilidad conformacional y del ligando de MDs de LAO en presencia de ligandos fisiológicos o no-fisiológicos, partiendo tanto del EC (en azul, columnas 3-5) como del EA (en azul, columnas 6-8).

Los resultados en la **tabla 15** presentan algunos elementos interesantes:

- Los ligandos fisiológicos tienen promedios más altos que los ligandos no-fisiológicos.
- El ligando arginina, que es el ligando de mayor afinidad ($K_d = 14$ nM), obtiene la mejor calificación, mientras que la lisina y la histidina obtienen calificaciones muy parecidas, lo cual estaría en desacuerdo con las afinidades reportadas para estos ligandos ($K_d = 15$ nM y 500 nM, respectivamente). No obstante, en un estudio previo que realizamos (**60**), se ha sugerido la posibilidad de que la histidina en su forma protonada (la forma usada en este estudio) podría ser unida por LAO tan bien como el ligando lisina y que la diferencia de afinidades entre estos ligandos estaría dada por los estados tautoméricos de la histidina no protonada.

- Los ligandos no-fisiológicos, además de tener “promedios ponderados” más bajos que los ligandos fisiológicos, presentan un orden de calificaciones que es proporcional a su hidrofobicidad. Este resultado podría indicar que los ligandos aromáticos más hidrofóbicos obtienen mejores calificaciones no porque presenten mejores interacciones específicas con LAO, sino porque el costo energético de ser expulsados del sitio de unión es más alto. Esta hipótesis se ve apoyada por el análisis visual de las dinámicas, que muestra que cuando los ligandos no-fisiológicos no son expulsados del sitio de unión, la región del ligando que corresponde a los oxígenos del carboxilo y el nitrógeno del amino se expone constantemente al solvente y el ligando se queda unido a la proteína por el anillo aromático. Esto sugiere que existe una barrera energética relacionada con la solvatación para la salida del ligando no-fisiológico del sitio de unión.

En conjunto, los resultados proporcionan un marco teórico (con correlación a estructuras cristalográficas experimentales) que nos permite analizar algunos de los diseños generados por ROSETTA, para determinar, a través de MD, si las mutaciones sugeridas proporcionan a LAO propiedades dinámicas que permitan una interacción correcta con los ligandos no-fisiológicos para los cuales fueron diseñadas.

Para considerar exitoso a un diseño de LAO, las propiedades “nativas” buscadas en una simulación por MD serán:

- Estabilización del estado cerrado.
- Retención del ligando en el sitio de unión, tanto en el estado abierto como en el cerrado.
- A partir del EA, inducción de un estado cerrado o semi-cerrado estable.

Análisis de los diseños mediante MD

Se analizaron, mediante simulaciones de MD, algunos de los diseños de LAO generados por ROSETTA en este trabajo, partiendo tanto del EC como del EA, con el objetivo de definir los diseños que son potencialmente exitosos.

Se escogieron cuatro diseños para cada ligando no-fisiológico (phe, trp, tyr), considerando, para cada uno la mejor estructura con respecto al lig_sum y la mejor con respecto al ΔG^{elec} (cuando estas estructuras resultaron ser la misma combinación de mutaciones, se eligió la segunda mejor estructura según el lig_sum, siempre y cuando esta segunda estructura tuviera diferentes mutaciones). Los códigos correspondientes a las tablas de diseño (ver resultados de diseño) que se sometieron a análisis por MD son: TYR-R1-62, TYR-R1-83, TYR-R2-84, TYR-R2-95, PHE-R1-78, PHE-R1-85, PHE-R3-74, PHE-R3-95, TRP-R1-55, TRP-R1-64, TRP-R6-41 y TRP-R6-62.

Los diseños se analizaron por MD en presencia del ligando no-fisiológico para el que fueron diseñadas, partiendo tanto del EA como del EC. Se utilizó el programa ROSETTA en su modo de “diseño de proteínas” para generar las mutaciones correspondientes en el estado abierto (forzando al programa a que generara las mutaciones deseadas en cada caso). De esta manera se generaron diez estructuras del EA para cada diseño, y se eligió aquella que tuviera el mejor bk_tot. Posteriormente se insertó el ligando no-fisiológicos correspondiente en el sitio de unión y se realizaron las MD bajo las mismas condiciones que se emplearon para la caracterización dinámica de LAO.

Los resultados de los ángulos de apertura, rotación y distancia del ligando con respecto a la Ala-90 para los diseños analizados por MD se muestran en las **Figuras 17-22**.

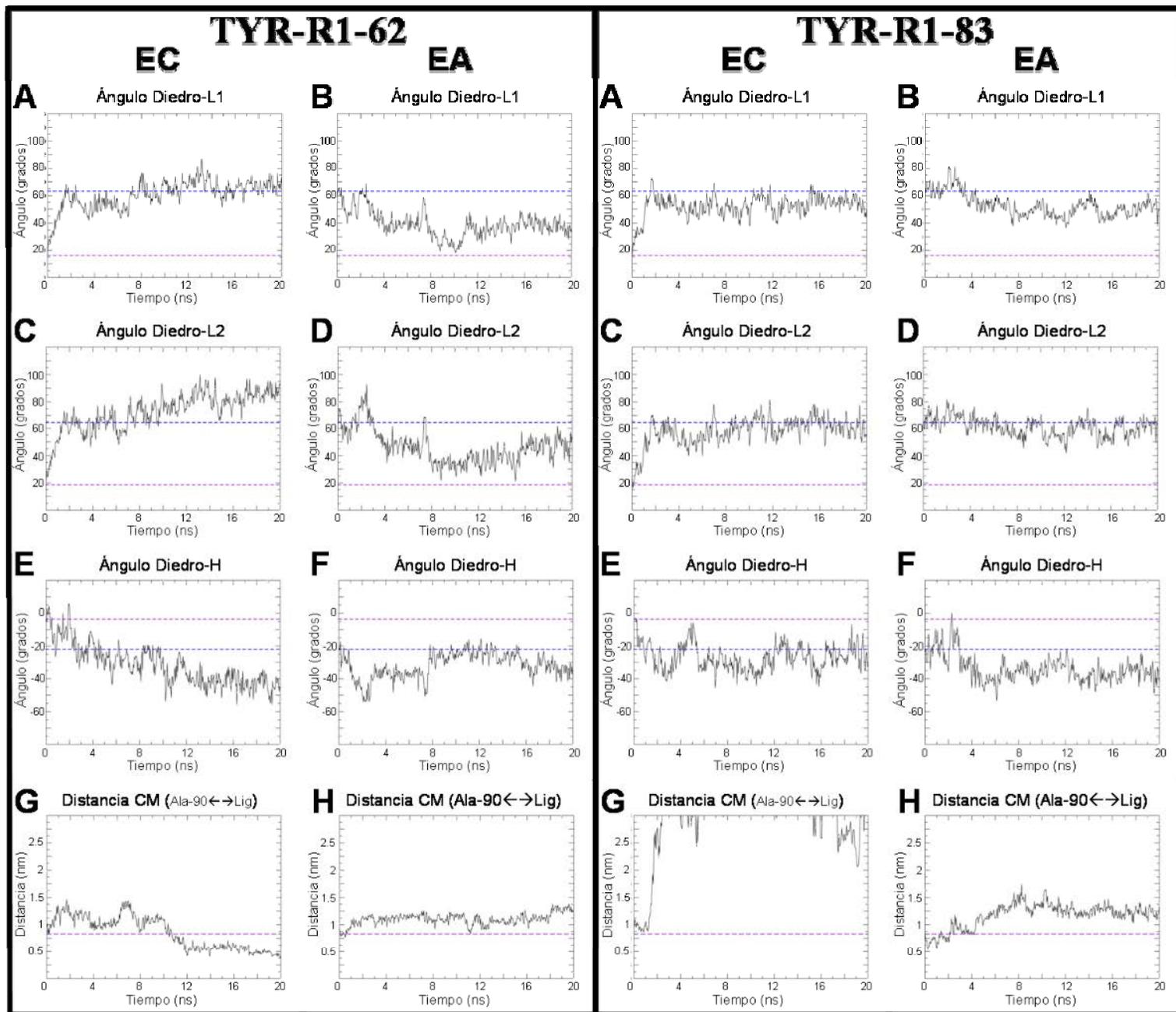


Figura 17. Gráficas de los ángulos diedros y distancias del ligando con respecto a la Ala-90 que fueron medidos lo largo de MDs para dos diseños independientes de LAO generados para unir el ligando “no-fisiológico” tirosina. **Panel izquierdo:** diseño TYR-R1-62 partiendo del EC (izquierda) y del EA (derecha); **A-B)** ángulo diedro L1; **C-D)** ángulo diedro L2; **E-F)** ángulo diedro H1; **G-H)** distancia con respecto a la Ala-90. Los colores de los trazos son los mismos para cada repetición en todas las Gráfica. **Panel derecho:** diseño TYR-R1-83 partiendo del EC (izquierda) y del EA (derecha); **A-B)** ángulo diedro L1; **C-D)** ángulo diedro L2; **E-F)** ángulo diedro H1; **G-H)** distancia con respecto a la Ala-90. Los colores de los trazos son los mismos para cada repetición en todas las Gráfica. La línea (discontinua) magenta representa el valor correspondiente al EC cristalográfico, mientras que la línea (discontinua) azul representa el valor correspondiente EA cristalográfico.

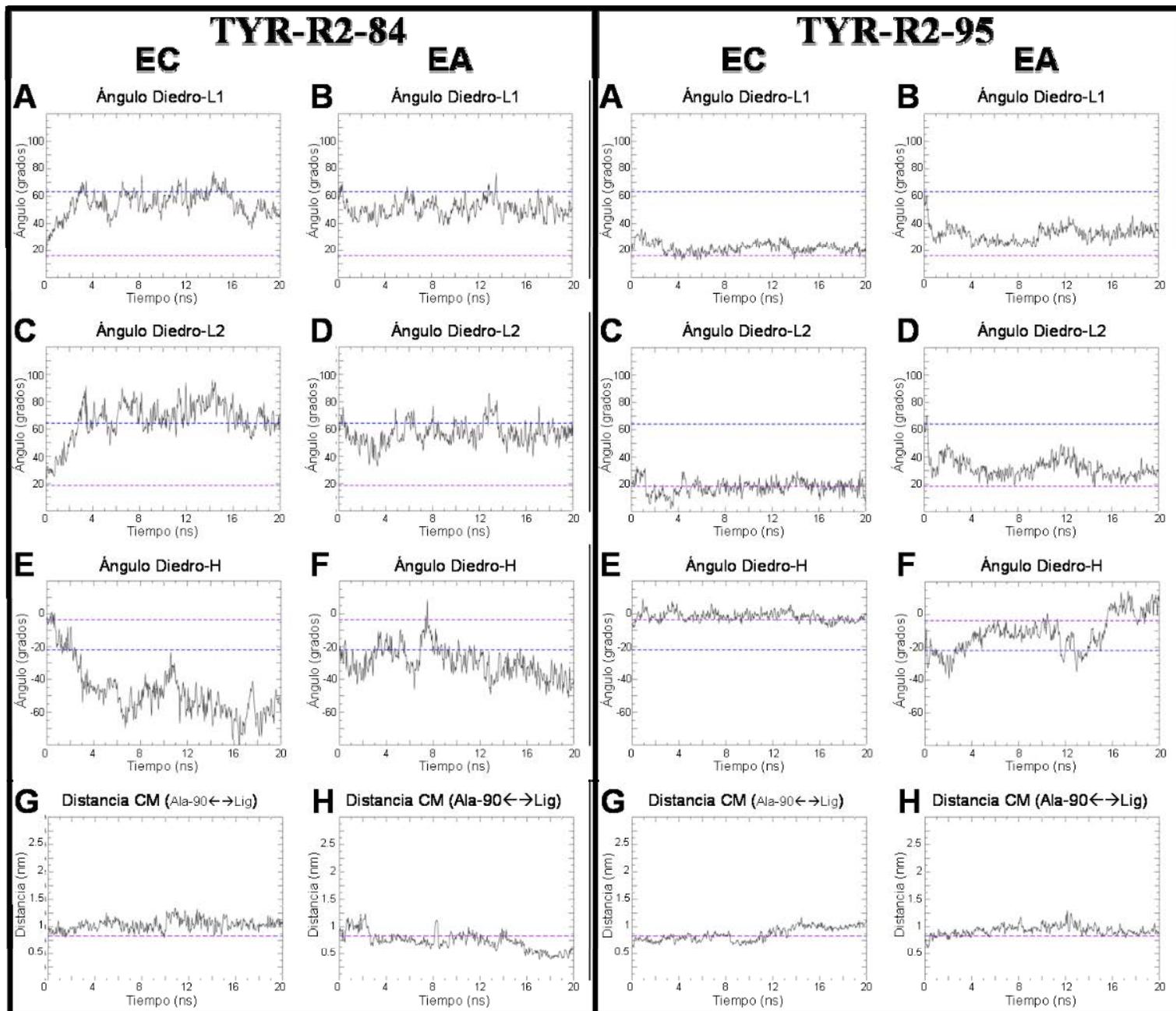


Figura 18. Gráficas de los ángulos diedros y distancias del ligando con respecto a la Ala-90 que fueron medidos lo largo de MDs para dos diseños independientes de LAO generados para unir el ligando “no-fisiológico” tirosina. **Panel izquierdo:** diseño TYR-R2-84 partiendo del EC (izquierda) y del EA (derecha); **A-B)** ángulo diedro L1; **C-D)** ángulo diedro L2; **E-F)** ángulo diedro H1; **G-H)** distancia con respecto a la Ala-90. Los colores de los trazos son los mismos para cada repetición en todas las Gráfica. **Panel derecho:** diseño TYR-R2-95 partiendo del EC (izquierda) y del EA (derecha); **A-B)** ángulo diedro L1; **C-D)** ángulo diedro L2; **E-F)** ángulo diedro H1; **G-H)** distancia con respecto a la Ala-90. Los colores de los trazos son los mismos para cada repetición en todas las Gráfica. La línea (discontinua) magenta representa el valor correspondiente al EC cristalográfico, mientras que la línea (discontinua) azul representa el valor correspondiente EA cristalográfico.

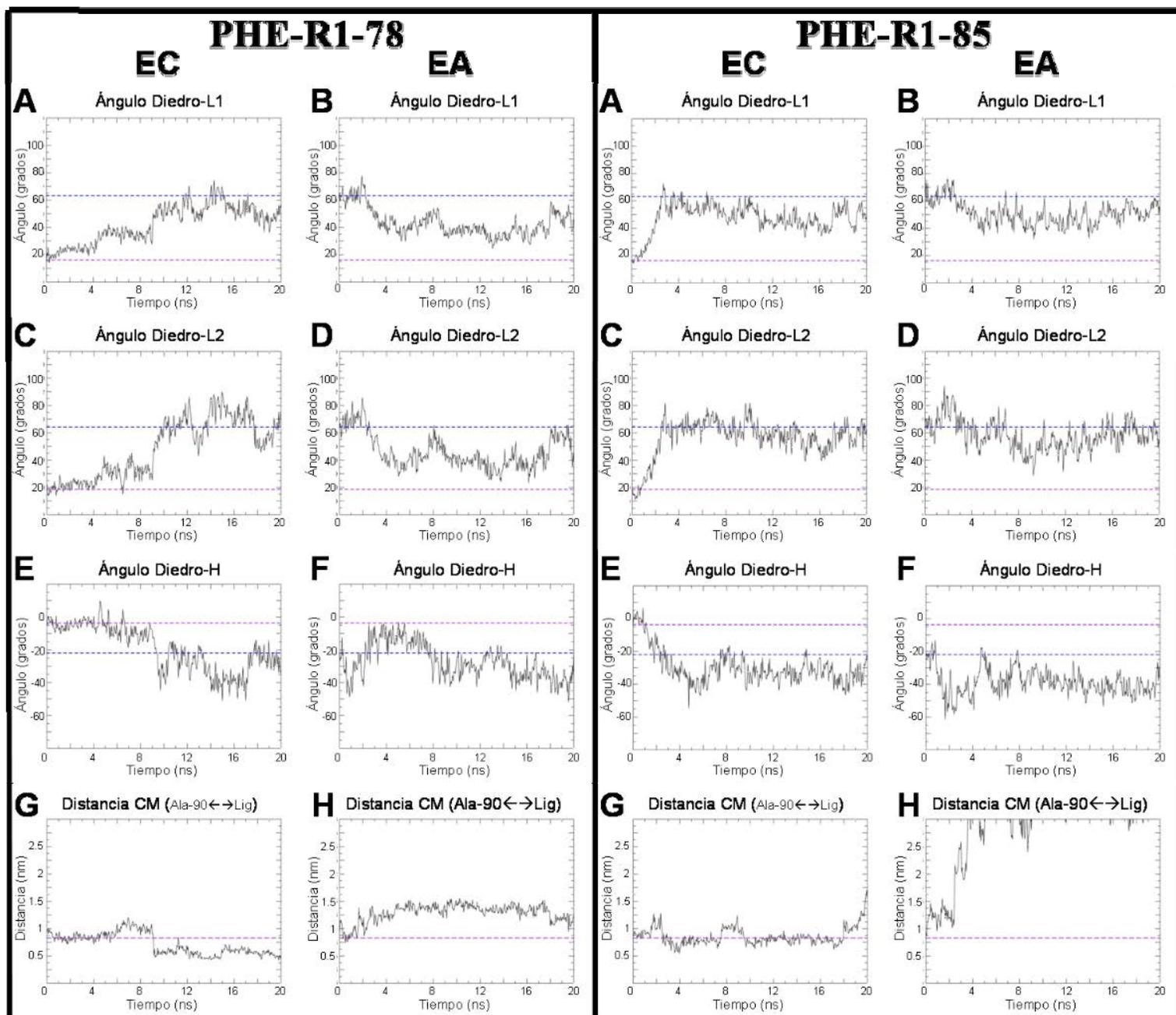


Figura 19. Gráficas de los ángulos diedros y distancias del ligando con respecto a la Ala-90 que fueron medidos lo largo de MDs para dos diseños independientes de LAO generados para unir el ligando “no-fisiológico” fenilalanina. **Panel izquierdo:** diseño PHE-R1-78 partiendo del EC (izquierda) y del EA (derecha); **A-B)** ángulo diedro L1; **C-D)** ángulo diedro L2; **E-F)** ángulo diedro H1; **G-H)** distancia con respecto a la Ala-90. Los colores de los trazos son los mismos para cada repetición en todas las Gráfica. **Panel derecho:** diseño PHE-R1-85 partiendo del EC (izquierda) y del EA (derecha); **A-B)** ángulo diedro L1; **C-D)** ángulo diedro L2; **E-F)** ángulo diedro H1; **G-H)** distancia con respecto a la Ala-90. Los colores de los trazos son los mismos para cada repetición en todas las Gráfica. La línea (discontinua) magenta representa el valor correspondiente al EC cristalográfico, mientras que la línea (discontinua) azul representa el valor correspondiente EA cristalográfico.

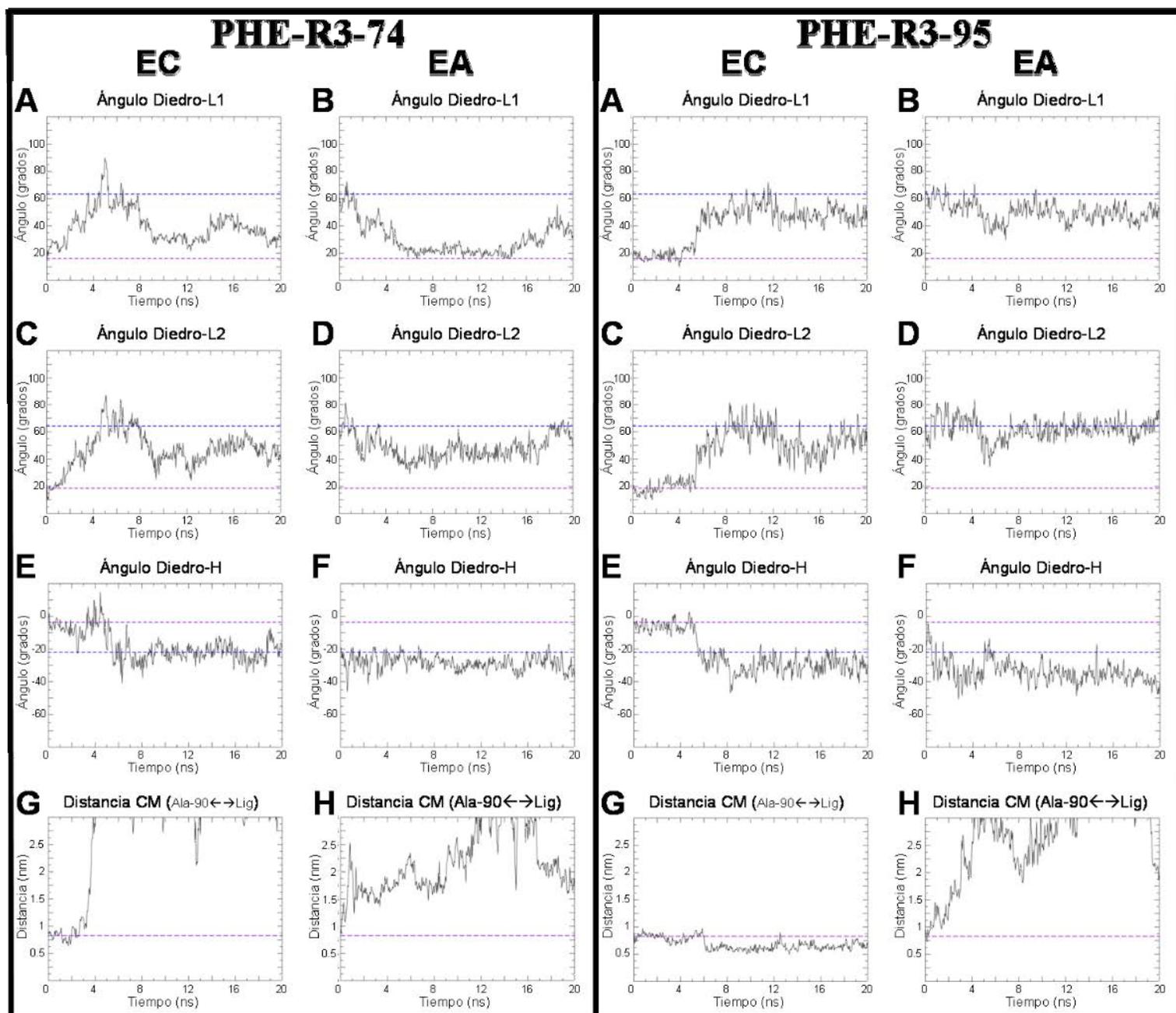


Figura 20. Gráficas de los ángulos diedros y distancias del ligando con respecto a la Ala-90 que fueron medidos lo largo de MDs para dos diseños independientes de LAO generados para unir el ligando “no-fisiológico” fenilalanina. **Panel izquierdo:** diseño PHE-R3-74 partiendo del EC (izquierda) y del EA (derecha); **A-B)** ángulo diedro L1; **C-D)** ángulo diedro L2; **E-F)** ángulo diedro H1; **G-H)** distancia con respecto a la Ala-90. Los colores de los trazos son los mismos para cada repetición en todas las Gráfica. **Panel derecho:** diseño PHE-R3-95 partiendo del EC (izquierda) y del EA (derecha); **A-B)** ángulo diedro L1; **C-D)** ángulo diedro L2; **E-F)** ángulo diedro H1; **G-H)** distancia con respecto a la Ala-90. Los colores de los trazos son los mismos para cada repetición en todas las Gráfica. La línea (discontinua) magenta representa el valor correspondiente al EC cristalográfico, mientras que la línea (discontinua) azul representa el valor correspondiente EA cristalográfico.

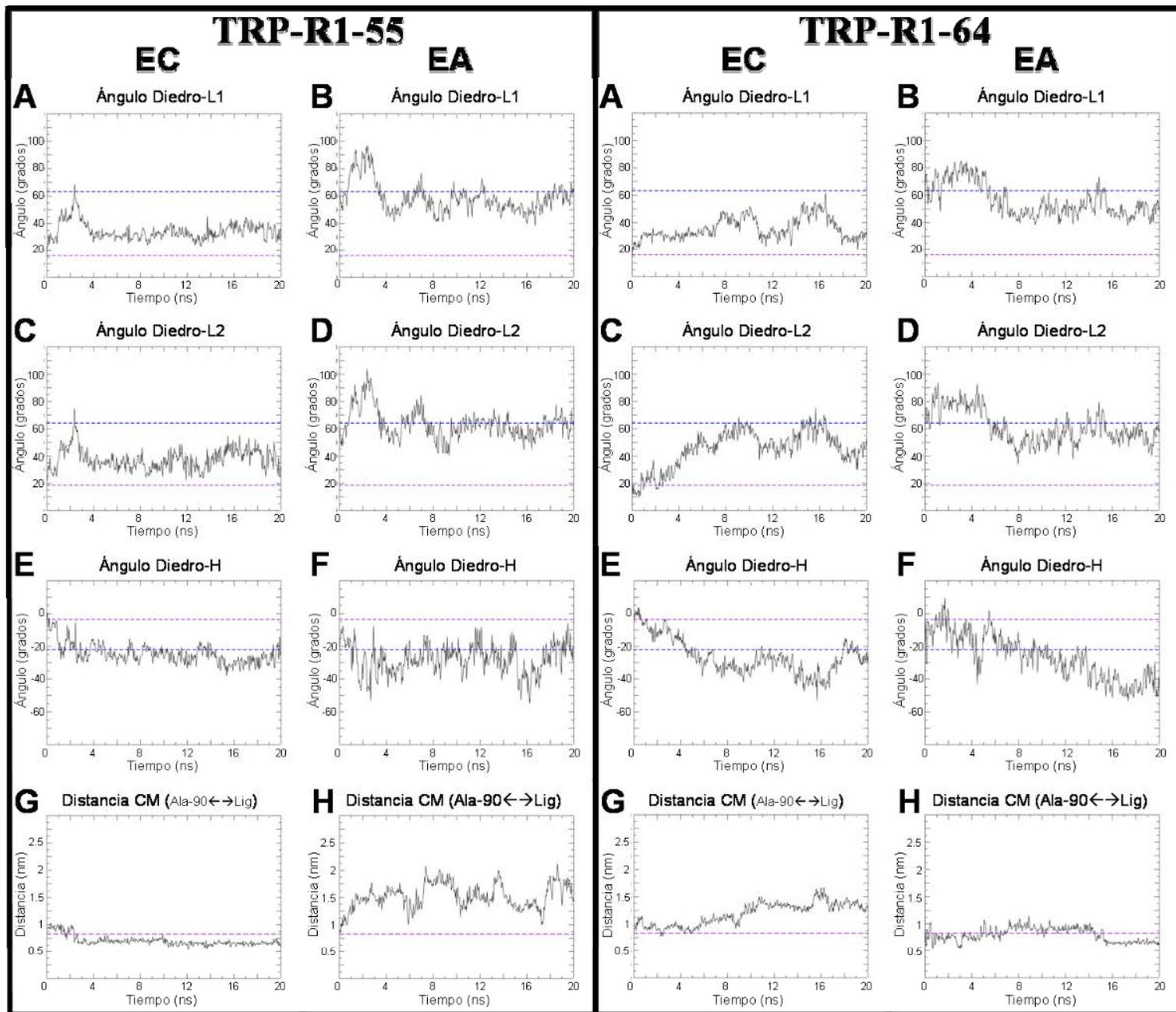


Figura 21. Gráficas de los ángulos diedros y distancias del ligando con respecto a la Ala-90 que fueron medidos lo largo de MDs para dos diseños independientes de LAO generados para unir el ligando “no fisiológico” triptofano. **Panel izquierdo:** diseño TRP-R1-55 partiendo del EC (izquierda) y del EA (derecha); **A-B)** ángulo diedro L1; **C-D)** ángulo diedro L2; **E-F)** ángulo diedro H1; **G-H)** distancia con respecto a la Ala-90. Los colores de los trazos son los mismos para cada repetición en todas las Gráfica. **Panel derecho:** diseño TRP-R1-64 partiendo del EC (izquierda) y del EA (derecha); **A-B)** ángulo diedro L1; **C-D)** ángulo diedro L2; **E-F)** ángulo diedro H1; **G-H)** distancia con respecto a la Ala-90. Los colores de los trazos son los mismos para cada repetición en todas las Gráfica. La línea (discontinua) magenta representa el valor correspondiente al EC cristalográfico, mientras que la línea (discontinua) azul representa el valor correspondiente EA cristalográfico.

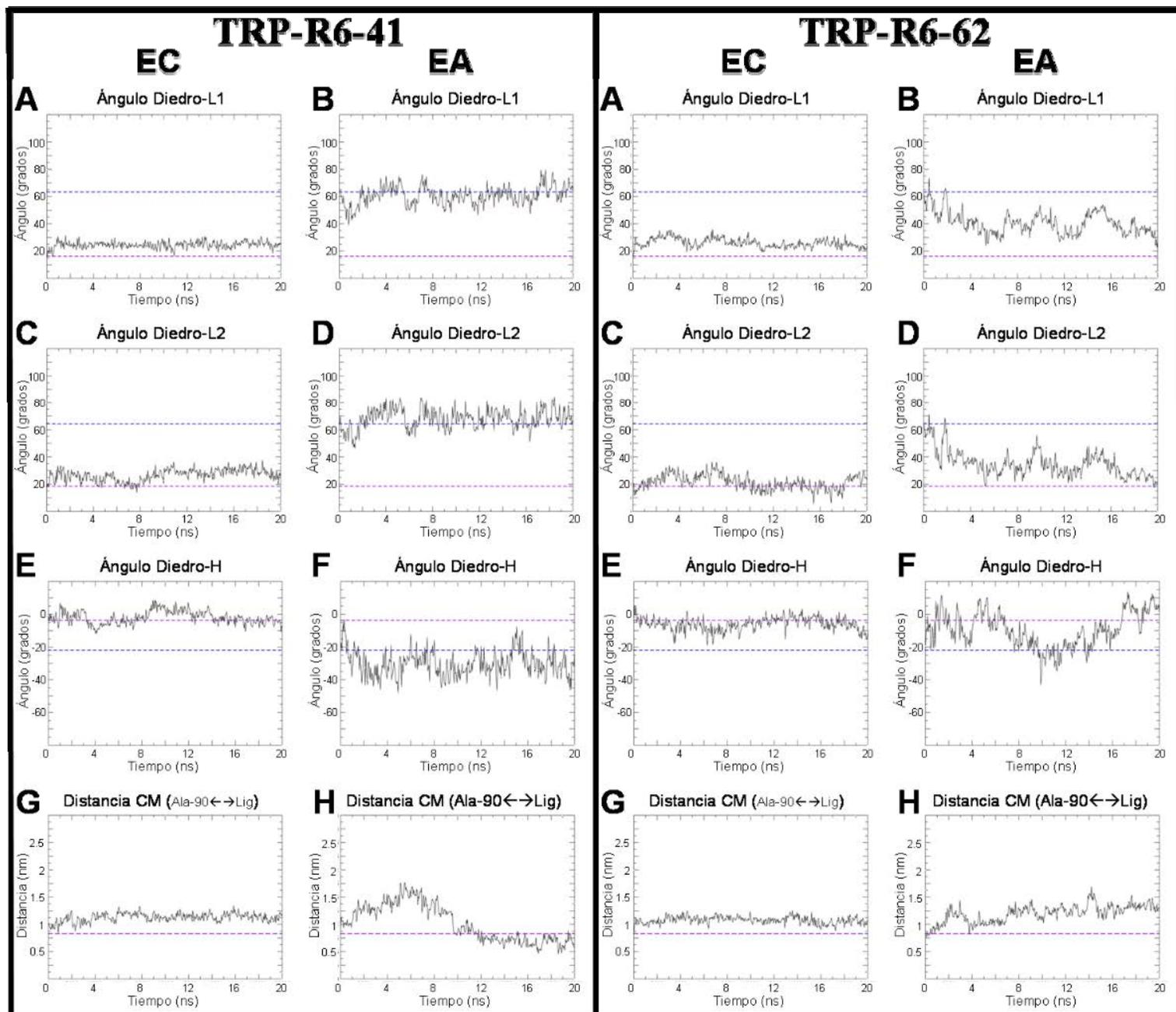


Figura 22. Gráficas de los ángulos diedros y distancias del ligando con respecto a la Ala-90 que fueron medidos lo largo de MDs para dos diseños independientes de LAO generados para unir el ligando “no-fisiológico” triptofano. **Panel izquierdo:** diseño TRP-R6-41 partiendo del EC (izquierda) y del EA (derecha); **A-B)** ángulo diedro L1; **C-D)** ángulo diedro L2; **E-F)** ángulo diedro H1; **G-H)** distancia con respecto a la Ala-90. Los colores de los trazos son los mismos para cada repetición en todas las Gráfica. **Panel derecho:** diseño TRP-R6-62 partiendo del EC (izquierda) y del EA (derecha); **A-B)** ángulo diedro L1; **C-D)** ángulo diedro L2; **E-F)** ángulo diedro H1; **G-H)** distancia con respecto a la Ala-90. Los colores de los trazos son los mismos para cada repetición en todas las Gráfica. La línea (discontinua) magenta representa el valor correspondiente al EC cristalográfico, mientras que la línea (discontinua) azul representa el valor correspondiente EA cristalográfico.

Los resultados obtenidos son heterogéneos. En muchos casos muestran que las mutaciones afectaron diferentes combinaciones de las propiedades dinámicas de la proteína. Por ejemplo, la proteína silvestre presenta generalmente una conformación cerrada “poco estable” con los ligandos no-fisiológicos (con excepción del ligando triptofano), mientras que varios de los diseños de ROSETTA muestran un EC “muy inestable”, que se refleja en una apertura de la proteína en los primeros nanosegundos de simulación de las MD que parten del EC (TYR-R1-62, TYR-R1-83, TYR-R2-84, PHE-R1-78, PHE-R1-85, PHE-R3-74, PHE-R3-95, y TRP-R1-64). Además, en general, las MD que parten del EC de la proteína silvestre no pierden al ligando (inclusive los ligandos no-fisiológicos), mientras que en los diseños observamos varios casos en los cuales se pierde al ligando del sitio de unión.

Hay cuatro diseños que muestran un EC estable (TYR-R2-95, TRP-R1-55, TRP-R6-41 y TRP-R6-62). Tres de estos diseños presentan la capacidad de retener al ligando en el EA (TRP-R6-41, TRP-R6-62 y TYR-R2-95), pero sólo dos de estos últimos adquirieron la propiedad más importante: pasar del EA a la conformación cerrada con el ligando establemente unido en el sitio de unión. Podemos concluir que:

“Dos de los diseños analizados muestran todas las propiedades que se buscaban para considerar exitoso a un diseño (TYR-R2-95 y TRP-R6-62)”.

En la **tabla 16** se muestran las principales propiedades analizadas en los diseños y su ponderación.

Diseño	Mantiene el EC	Ligando estable	Inspección visual del ligando	Cambia al EC	Ligando estable	Inspección visual del ligando	Suma	Valor de referencia
PHE-R1-78	0	0.5	0.5	0.5	0.5	0.5	2.5	2.66
PHE-R1-85	0	0.5	0.25	0	0	0	0.75	
PHE-R3-74	0.5	0	0	0.5	0	0	1	
PHE-R3-95	0	1	1	0	0	0	2	
TRP-R1-55	1	1	1	0	0	0.25	3.25	3.25
TRP-R1-64	0	0.5	0.25	0	1	0.5	2.25	
TRP-R6-41	1	1	1	0	0.75	0.75	4.5	
TRP-R6-62	1	1	1	1	0.75	0.75	5.5	
TYR-R1-62	0	0.5	0.25	0.75	1	0.75	3.25	1.5
TYR-R1-83	0	0	0	0	0.5	0.25	0.75	
TYR-R2-84	0	1	1	0	1	0.25	3.25	
TYR-R2-95	1	1	1	1	1	0.75	5.75	

Tabla 16. Ponderación de diversos factores de la estabilidad conformacional y de estabilidad del ligando de MDs realizadas a distintos diseños de LAO en presencia de los ligandos “no-fisiológicos” para los cuales fueron diseñadas. (en azul, columnas 2-4) como del EA (en azul, columnas 5-7).

Conclusiones

El diseño racional de proteínas por computadora es una alternativa (*in silico*) a diversos métodos experimentales con el fin de generar secuencias proteicas que presenten características deseadas. En el caso de LAO, se sabe, por estudios de cristalografía de rayos-X, que existe un gran cambio conformacional asociado con la unión del ligando, cambio que hemos sido capaces de reproducir mediante simulaciones de dinámica molecular.

En este trabajo hemos caracterizado el comportamiento de una interacción exitosa “LAO \leftrightarrow Ligando”. Apoyándonos en estos resultados hemos buscado las mismas propiedades en 12 diseños que generamos previamente con la ayuda del programa ROSETTA y de técnicas de docking (usando AUTODOCK), con la finalidad de modificar la afinidad fisiológica de LAO para unir diferentes aminoácidos aromáticos (tyr, trp, phe). Estos diseños fueron inicialmente analizados mediante el uso de métodos de cálculo de “energía electrostática de interacción”. No obstante, dado el carácter dinámico de LAO, esta información no es suficiente para determinar cuales de estos diseños son exitosos.

Las dinámicas moleculares mostraron diferencias claras en el comportamiento de los diseños con respecto a la proteína silvestre. La mayor parte de estos diseños falló en el sentido de presentar las propiedades buscadas, es decir:

- Un estado cerrado estable con el ligando.
- Un estado abierto estable con el ligando que tendiera a un estado cerrado o semi-cerrado de la proteína.

Sin embargo, dos diseños presentaron las propiedades buscadas (TYR-R2-95 y TRP-R6-62), y por lo tanto, estos diseños son nuestros candidatos para ser probados experimentalmente. Los diseños “exitosos” corresponden a una mutante de LAO que, en teoría, une el aminoácido

triptofano, y a otra que une tirosina. Generar estas mutantes a partir de la proteína silvestre LAO representaría 8 mutaciones para TRP-R6-62 y 5 mutaciones para TYR-R2-95 (**Tabla 17**). En el caso del aminoácido fenilalanina ninguno de los diseños analizados mostró las propiedades adecuadas para ser considerado exitoso. Es posible que un muestreo más amplio por MD de los diseños, o que diseños que sean generados con otros parámetros de búsqueda, sean capaces de arrojar diseños que cumplan con las características adecuadas para el aminoácido fenilalanina.

# Res	14	52	70	90	117	122	160	161	164	lig_sum	bk_tot	ΔG^{elec}	$\Delta G^{coulomb}$	Estructura
Res Ori Ori	TYR	PHE	SER	ALA	LEU	GLN	GLN	ASP	ALA					Silvestre
# Mut														
5	CYS	***	GLY	ASP	***	***	GLY	PRO	***	-22.46	-840.43	-5.7	-79.31	TYR-R2-95
8	ILE	HIS	***	ASP	ASP	GLU	THR	GLY	SER	-17.5	-838.75	-1.33	-51.05	TRP-R6-62

Tabla 17. Mutaciones en LAO correspondientes a los diseños: TYR-R2-95 y TRP-R6-62, únicos diseños que se calificaron como exitosos después del análisis por MD.

Sólo dos de los diseños analizados mostraron las propiedades necesarias para considerarlos “diseños exitosos”. Esto no es desalentador, ya que en los últimos años el diseño racional de proteínas ha comenzado a arrojar resultados cada vez mejores, pero aún se encuentra en un momento en el cual la relación “resultados_exitosos/resultados_fallidos” está inclinada hacia el lado de los fracasos, por lo que se requiere analizar, y muchas veces probar, varios diseños para encontrar uno que posea las características deseadas. Mediante el refinamiento de los métodos de diseño existentes y el desarrollo de nuevas metodologías será posible invertir esta relación en el futuro.

También hemos realizado una caracterización extensiva del comportamiento dinámico de la unión de tres ligandos fisiológicos (arg, lys, his) en el sitio de unión de LAO, así como de la discriminación de tres ligandos no-fisiológicos (phe, tyr, trp). Esta caracterización representa el primer paso hacia el entendimiento del mecanismo dinámico de unión de ligandos múltiples en LAO, y sienta las bases para futuros estudios al respecto.

Finalmente, agradecemos de manera especial a la DGSCA, ya que, gracias a su apoyo en recursos computacionales, ha sido posible este trabajo. Cada dinámica de producción realizada a LAO equivale a un tiempo de integración de 20 ns, y esto significa ≈ 66 horas de cómputo en 4 procesadores de la supercomputadora KanBalam, y por lo tanto a un tiempo de pared (tiempo en un solo procesador) de ≈ 264 horas. En total para este trabajo hemos realizado 78 dinámicas (mismas que se presentan en los resultados), que equivalen a un tiempo de pared de 2.35 años de cómputo continuo (sólo para las MDs de producción, sin tomar en cuenta las MDs de “posiciones restringidas”, ni el tiempo de análisis de los datos). Este esfuerzo computacional ha sido posible gracias al uso de la supercomputadora KanBalam (HP Cluster Plataform 4000), perteneciente a la UNAM y localizada en la Dirección General de Servicios de Computo Académico (DGSCA).

Perspectivas

El estudio mediante dinámicas moleculares de la proteína LAO expone una gran variedad de experimentos que se pueden realizar para estudiar las bases de la unión de ligandos a la proteína, así como para estudiar los factores importantes en la transición $EA \leftrightarrow EC$ y, por supuesto, analizar el efecto de diversos diseños de la proteína. Dentro de las perspectivas generales para el desarrollo del proyecto se encuentran:

- Generar las mutantes correspondientes a los diseños TRP-R6-62 y TYR-R2-95, con el fin de correlacionar los resultados teóricos con los experimentales y refinar la estrategia de diseño.
- Generar topologías para el ligando serotonina y analizar los diseños generados para este ligando mediante “cálculos de energía electrostática” y MD.
- Estudiar mediante MD el efecto de la presencia de ligandos con carga total negativa (ej. aa glutámico y aspártico), ya que se espera que dado el ambiente electrostático negativo de la cavidad de unión de LAO, los ligandos con carga negativa sean expulsados fácilmente del sitio de unión.
- Determinar la importancia de los distintos aminoácidos que forman el “sitio de unión” en la afinidad, a través de la sustitución independiente por alaninas (exploración por alaninas).
- Calcular la energética de unión (ΔG) de los ligandos fisiológicos a la proteína mediante dinámicas de perturbación de energía libre.
- Aunque en la literatura se ha sugerido que la concentración de NaCl no tiene efecto en la unión de los ligandos (5), cálculos de energía

electrostática que hemos realizado sugieren que la electrostática juega un papel importante en la unión del ligando (**60**), por lo que sería importante corroborar tanto experimentalmente como mediante MD, el efecto de las sales en la unión de los ligandos fisiológicos.

Bibliografía

1. Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L., and Baker, D. (2003) Design of a novel globular protein fold with atomic-level accuracy, *Science (New York, N.Y)* 302, 1364-1368.
2. van den Burg, B., and Eijssink, V. G. (2002) Selection of mutations for increased protein stability, *Current opinion in biotechnology* 13, 333-337.
3. Kim, D. E., Gu, H., and Baker, D. (1998) The sequences of small proteins are not extensively optimized for rapid folding by natural selection, *Proceedings of the National Academy of Sciences of the United States of America* 95, 4982-4986.
4. Demetrius, L. (2002) Thermodynamics and kinetics of protein folding: an evolutionary perspective, *Journal of theoretical biology* 217, 397-411.
5. Nikaido, K., and Ames, G. F. (1992) Purification and characterization of the periplasmic lysine-, arginine-, ornithine-binding protein (LAO) from *Salmonella typhimurium*, *The Journal of biological chemistry* 267, 20706-20712.
6. Ames, G. F. (1986) Bacterial periplasmic transport systems: structure, mechanism, and evolution, *Annual review of biochemistry* 55, 397-425.
7. Ames, G. F. (1988) Structure and mechanism of bacterial periplasmic transport systems, *Journal of bioenergetics and biomembranes* 20, 1-18.
8. Locher, K. P. (2004) Structure and mechanism of ABC transporters, *Current opinion in structural biology* 14, 426-431.
9. Kustu, S. G., and Ames, G. F. (1974) The histidine-binding protein J, a histidine transport component, has two different functional sites, *The Journal of biological chemistry* 249, 6976-6983.
10. 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, *The Journal of biological chemistry* 266, 23893-23899.
11. Liu, C. E., Liu, P. Q., Wolf, A., Lin, E., and Ames, G. F. (1999) Both lobes of the soluble receptor of the periplasmic histidine permease, an ABC transporter (traffic ATPase), interact with the membrane-bound complex. Effect of different ligands and consequences for the mechanism of action, *The Journal of biological chemistry* 274, 739-747.
12. Adams, M. D., and Oxender, D. L. (1989) Bacterial periplasmic binding protein tertiary structures, *The Journal of biological chemistry* 264, 15739-15742.
13. Quijcho, F. A. (1990) Atomic structures of periplasmic binding proteins and the high-affinity active transport systems in bacteria, *Philosophical transactions of the Royal Society of London* 326, 341-351; discussion 351-342.
14. 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, *The Journal of biological chemistry* 269, 4135-4143.
15. 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, *The Journal of biological chemistry* 268, 11348-11355.

16. Fukami-Kobayashi, K., Tateno, Y., and Nishikawa, K. (1999) Domain dislocation: a change of core structure in periplasmic binding proteins in their evolutionary history, *Journal of molecular biology* 286, 279-290.
17. Dwyer, M. A., and Hellinga, H. W. (2004) Periplasmic binding proteins: a versatile superfamily for protein engineering, *Current opinion in structural biology* 14, 495-504.
18. Dwyer, M. A., Looger, L. L., and Hellinga, H. W. (2004) Computational design of a biologically active enzyme, *Science (New York, N.Y)* 304, 1967-1971.
19. 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.
20. Hellinga, H. W. (1997) Rational protein design: combining theory and experiment, *Proceedings of the National Academy of Sciences of the United States of America* 94, 10015-10017.
21. Kuhlman, B., and Baker, D. (2000) Native protein sequences are close to optimal for their structures, *Proceedings of the National Academy of Sciences of the United States of America* 97, 10383-10388.
22. Bonneau, R., Tsai, J., Ruczinski, I., Chivian, D., Rohl, C., Strauss, C. E., and Baker, D. (2001) Rosetta in CASP4: progress in ab initio protein structure prediction, *Proteins Suppl* 5, 119-126.
23. Rohl, C. A., Strauss, C. E., Misura, K. M., and Baker, D. (2004) Protein structure prediction using Rosetta, *Methods in enzymology* 383, 66-93.
24. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank, *Nucleic acids research* 28, 235-242.
25. Dantas, G., Kuhlman, B., Callender, D., Wong, M., and Baker, D. (2003) A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins, *Journal of molecular biology* 332, 449-460.
26. Lazaridis, T., and Karplus, M. (1999) Effective energy function for proteins in solution, *Proteins* 35, 133-152.
27. Dunbrack, R. L., Jr., and Cohen, F. E. (1997) Bayesian statistical analysis of protein side-chain rotamer preferences, *Protein Sci* 6, 1661-1681.
28. Meiler, J., and Baker, D. (2006) ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility, *Proteins* 65, 538-548.
29. Halperin, I., Ma, B., Wolfson, H., and Nussinov, R. (2002) Principles of docking: An overview of search algorithms and a guide to scoring functions, *Proteins* 47, 409-443.
30. Sousa, S. F., Fernandes, P. A., and Ramos, M. J. (2006) Protein-ligand docking: current status and future challenges, *Proteins* 65, 15-26.
31. Wade, R. C., Gabdouliline, R. R., Ludemann, S. K., and Lounnas, V. (1998) Electrostatic steering and ionic tethering in enzyme-ligand binding: insights from simulations, *Proceedings of the National Academy of Sciences of the United States of America* 95, 5942-5949.
32. Gordon, D. B., Marshall, S. A., and Mayo, S. L. (1999) Energy functions for protein design, *Current opinion in structural biology* 9, 509-513.
33. Kono, H., and Saven, J. G. (2001) Statistical theory for protein combinatorial libraries. Packing interactions, backbone flexibility, and the sequence variability of a main-chain structure, *Journal of molecular biology* 306, 607-628.

34. Liang, S., and Grishin, N. V. (2004) Effective scoring function for protein sequence design, *Proteins* 54, 271-281.
35. Marshall, S. A., Morgan, C. S., and Mayo, S. L. (2002) Electrostatics significantly affect the stability of designed homeodomain variants, *Journal of molecular biology* 316, 189-199.
36. Sheinerman, F. B., and Honig, B. (2002) On the role of electrostatic interactions in the design of protein-protein interfaces, *Journal of molecular biology* 318, 161-177.
37. Lee, L. P., and Tidor, B. (2001) Barstar is electrostatically optimized for tight binding to barnase, *Nature structural biology* 8, 73-76.
38. Bourne, P. E., and Weissig, H. (2003) *Structural bioinformatics*, Wiley-LISS, Hoboken, NJ.
39. Alder, B. J., and Wainwright, T. E. (1957) Phase Transition for a Hard Sphere System, *The Journal of Chemical Physics* 27, 1208-1209.
40. Alder, B. J., and Wainwright, T. E. (1959) Studies in Molecular Dynamics. I. General Method, *The Journal of Chemical Physics* 31, 459-466.
41. Rahman, A. (1964) Correlations in the Motion of Atoms in Liquid Argon, *Physical Review* 136, A405.
42. Rahman, A., and Stillinger, F. H. (1974) Propagation of sound in water. A molecular-dynamics study, *Physical Review A* 10, 368.
43. McCammon, J. A., Gelin, B. R., and Karplus, M. (1977) Dynamics of folded proteins, *Nature* 267, 585-590.
44. 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.
45. Guex, N., and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling, *Electrophoresis* 18, 2714-2723.
46. DeLano, W. L. (2002) The PyMOL Molecular Graphics System, en WWW <http://www.pymol.org>
47. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations, *Nucleic acids research* 32, W665-667.
48. Vriend, G. (1990) WHAT IF: a molecular modeling and drug design program, *J Mol Graph* 8, 52-56, 29.
49. Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., and Olson, A. J. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, *Journal of computational chemistry* 19, 1639-1662.
50. M. Holst, N. B. F. W. (2000) Adaptive multilevel finite element solution of the Poisson-Boltzmann equation I. Algorithms and examples, *Journal of computational chemistry* 21, 1319-1342.
51. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome, *Proceedings of the National Academy of Sciences of the United States of America* 98, 10037-10041.
52. 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.

53. 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.
54. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics, *Journal of Molecular Graphics* 14, 33-38.
55. Trakhanov, S., Vyas, N. K., Luecke, H., Kristensen, D. M., Ma, J., and Quioco, F. A. (2005) Ligand-free and -bound structures of the binding protein (LivJ) of the Escherichia coli ABC leucine/isoleucine/valine transport system: trajectory and dynamics of the interdomain rotation and ligand specificity, *Biochemistry* 44, 6597-6608.
56. Quioco, F. A., Spurlino, J. C., and Rodseth, L. E. (1997) Extensive features of tight oligosaccharide binding revealed in high-resolution structures of the maltodextrin transport/chemosensory receptor, *Structure* 5, 997-1015.
57. Marvin, J. S., and Hellinga, H. W. (2001) Manipulation of ligand binding affinity by exploitation of conformational coupling, *Nature structural biology* 8, 795-798.
58. Telmer, P. G., and Shilton, B. H. (2003) Insights into the conformational equilibria of maltose-binding protein by analysis of high affinity mutants, *The Journal of biological chemistry* 278, 34555-34567.
59. Telmer, P. G., and Shilton, B. H. (2005) Structural studies of an engineered zinc biosensor reveal an unanticipated mode of zinc binding, *Journal of molecular biology* 354, 829-840.
60. Silva Manzano, D. A. (2006) Estudio in silico de los potenciales electrostáticos en la proteína periplásmica de Salmonella typhimurium que une lisina, arginina y ornitina, in *Facultad de Medicina*, p 63, UNAM, México D.F.
61. Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers* 22, 2577-2637.